

Plan Overview

A Data Management Plan created using DMPTool

Title: Field Spectroscopy Database

Creator: William Philpot - **ORCID:** [0000-0002-5283-4774](https://orcid.org/0000-0002-5283-4774)

Affiliation: Cornell University (cornell.edu)

Principal Investigator: William Philpot

Data Manager: William Philpot

Funder: National Science Foundation (nsf.gov)

Funding opportunity number: NSF 15-516

Template: NSF-EAR: Earth Sciences

Last modified: 06-09-2015

Copyright information:

The above plan creator(s) have agreed that others may use as much of the text of this plan as they would like in their own plans, and customize it as necessary. You do not need to credit the creator(s) as the source of the language used, but using any of the plan's text does not imply that the creator(s) endorse, or have any relationship to, your project or proposal

Field Spectroscopy Database

Types of data

Preservation of all data, samples, physical collections and other supporting materials needed for long-term earth science research and education is required of all EAR-supported researchers.

Spectroradiometer data will consist of spectral radiance of illumination sources (sun, sky, clouds, & artificial lamps), radiance reflected from natural surfaces, and reflectance from natural surfaces. The basic data will consist of the wavelengths, the appropriate radiometric parameter at each wavelength (radiance, reflectance) and a detailed description of the target. Also noted for each data set will be the calibration parameters, viewing conditions, e.g., illumination and viewing angles, sensor field-of-view (FOV).

Data collected in support of the spectral measurements will vary considerably with the application and setting. For example, field observations of reflectance will include information about the target (e.g., soil) and its characteristics (e.g., location, roughness, moisture content, particle size distribution, etc.), time of day, illumination conditions, any pertinent instrument characteristics (viewing angle, field of view, integration time, etc.), and record images of the study site. In contrast, a laboratory experiment designed to monitor reflectance of a soil sample throughout a drying event would require a record of the time and sample weight each time a spectrum is collected, and characteristics of the sample and sample holder, and possibly microscope images of the sample as well as record images of the experimental setup.

Data and metadata standards

Data archives must include easily accessible information about data holdings, including quality assessments, supporting ancillary information, and guidance and aids for locating and obtaining data.

Individual spectra are initially saved in a proprietary format, *.asd. Complete data sets will be saved together in a single directory along with a preliminary metadata file in *.txt format identifying the individual spectra and associating the set with a specific experiment or task. When a data set is selected for further processing, the *.asd files will be converted to *.csv format and a more complete metadata file will be created, with information about the experiment, a description of the target, and including data collected in support of the experiment (or lists of associated files). Complete experiments could include as much 5 MB of spectral data. The metadata could be as little as 10 kB, but could reach 100 MB if there are a substantial number of images collected as part of the

experiment. Record images collected of the study site or instruments will generally be saved in *.jpg format, while microscope data, or other data from which measurements may be required, will be saved in *.jp2 (lossless) or *.tiff format.

At a minimum, metadata will include descriptions of the sample, the study site and conditions, and the instrument configuration. Metadata may also include text files as well as iamges of handwritten field-notes. The specific content and extent of the metadata will depend entirely on the application.

Policies for access and sharing

It is the responsibility of researchers and organizations to make results, data, derived data products, and collections available to the research community in a timely manner and at a reasonable cost. In the interest of full and open access, data should be provided at the lowest possible cost to researchers and educators. This cost should, as a first principle, be no more than the marginal cost of filling a specific user request. Data may be made available for secondary use through submission to a national data center, publication in a widely available scientific journal, book or website, through the institutional archives that are standard for a particular discipline (e.g. IRIS for seismological data, UNAVCO for GP data), or through other EAR-specified repositories. Data inventories should be published or entered into a public database periodically and when there is a significant change in type, location or frequency of such observations. Principal Investigators working in coordinated programs may establish (in consultation with other funding agencies and NSF) more stringent data submission procedures.

Since there will be multiple research efforts using the field spectroradiometer, with spectra collected from a wide variety of materials under a wide range of conditions, in the field and in the laboratory, protocols and procedures will vary substantially from project to project. *The overall goal is to make the spectral data and essential, associated metadata generally available with no restrictions*, but the details of what is included and the timing of release may vary substantially from project to project.

Policies and provisions for re-use, re-distribution

For those programs in which selected principle investigators have initial periods of exclusive data use, data should be made openly available as soon as possible, but no later than two (2) years after the data were collected. This period may be extended under exceptional circumstances, but only by agreement between the Principal Investigator and the National Science Foundation. For continuing observations or for long-term (multi-year) projects, data are to be made public annually.

The timing of making the data available publicly will also vary with the project. In some cases, e.g., soil characterization, a goal is to create a spectral database, and virtually all of the data that has been processed and deemed acceptable will be made a part of the database. In others, e.g., the soil moisture study, It is likely that only a portion of the data collected will be selected for analysis. All data selected for analysis and fully processed will be made available via archiving at the conclusion of the project, if not before.

Plans for archiving and preservation of access

Remember - Data may be made available for secondary use through submission to a national data center, publication in a widely available scientific journal, book or website, through the institutional archives that are standard for a particular discipline (e.g. IRIS for seismological data, UNAVCO for GP data), or through other EAR-specified repositories.

Data will be archived using the eCommons@Cornell service. Spectral data and metadata in tabular form will be stored as *.csv format. Record images collected of the study site or instruments will generally be saved in *.jpg format, while microscope data, or other data from which measurements may be required, will be saved in *.jp2 (lossless) or *.tiff format.
