National Science Foundation (nsf.gov): Arctic Data Center: NSF Polar Programs

Roles and responsibilities

1. What parties and individuals will be involved with data management in this project?
2. What will be the roles and responsibilities of each party and or individual with respect to management of the data?
3. Who will be the lead or primary person responsible for ultimately ensuring compliance with the Data Management Plan?

Note: if you plan to submit data to the Arctic Data Center please refer to the guidance in the panel on the right.

Example Answer:

The project's principal investigator, Jane Doe, will ultimately be responsible for all of the data management. It is Doe's responsibility to make sure all of the project team members are taught the proper data management skills and uphold the data management requirements. Doe will delegate data management duties to the laboratory project data manager, Bonnie, along with graduate student, Clyde, working in the field. The NSF Arctic Data Center will provide data archival, preservation, access and metadata authoring services for the project.

Guidance:

Arctic Data Center Identification Policy: The Arctic Data Center utilizes ORCiDs (https://orcid.org/) to identify individuals associated with each dataset. An ORCiD will be required for the primary contact of each dataset. ORCiDs are not required for all associated parties but are encouraged so that proper identification and attribution can be given. Please plan on creating (when necessary) and recording ORCiDs for each individual involved with your project before submitting to the Arctic Data Center. The purpose of this section is to ensure that the individuals ultimately responsible for ensuring compliance with the data management plan are both aware and agree to their roles.

Types of data produced

What types of data, samples, collections, software, materials, etc. will be produced during your project?

Example Answer:

During this project, soil cores will be collected from three sites representing varying degrees of snow accumulation. Temperature, moisture and active layer thaw depth will be collected to record soil physical properties. Soil samples from each core will be analyzed for carbon concentration, nitrogen concentration and pH.

Aerial photos of each of the three sites will also be taken twice annually to visualize the snow cover. Additional data products that will be made available include data analysis codes in R and high school-level educational materials regarding soil properties in changing climatic conditions in the Arctic.

Guidance:

This is the most detailed section of the data management plan. Describe the categories of data being collected and how they tie into the data associated with the methods used to collect that data. Expect this section to be the most detailed section, taking up a large portion of your data management plan document.

What will be the approximate number and size of data files that will be produced during your project?

Example Answer:

In this project, there will be six experimental treatments applied to the *Saxifraga cespitosa* plants, with four replicates for each treatment collected three times during the year. Therefore, there will be approximately 72 data files each year (6 treatments x 4 replicates x 3 collections per year). The 72 data files will be approximately 720 MB in size.

Guidance:

This section is driven by sampling design, and is based on what you predict you will be collecting. Realistic estimates are sufficient. This section is important because reviewers need to understand what data you plan on collecting and how you plan on managing that data.

What type of metadata (information others might need to use your data) will be collected during your project?

Note: if you plan to submit data to the Arctic Data Center please refer to the guidance in the panel on the right.

Example Answer:

The project will collect and record a description of the time of soil core collection in the field, along with the time of soil core processing in the laboratory. The exact locations and conditions of the field sites will be described as well as the laboratory conditions and location of soil sample analysis. A detailed description of the soil core sample collection and processing methods in the field, as well as the laboratory soil carbon and nitrogen concentration collection and processing methods will be included. Units will be recorded for all samples. Soil carbon and nitrogen will be reported in %C and %N, respectively, along with the C:N ratio and the %delta 13C and %delta 15N. Soil depth will be recorded in centimeters.

Quality control procedures will be followed in both the field and the laboratory during sample collection and processing, and the details of these procedures will be included. The decisions for the inclusion of each component of the project, including the specific sampling methods, units, and procedures, will be provided. The hardware and software used will be provided. R Studio will be used for coding and data analysis, with the code shared in a Github repository.

Guidance:

[Arctic Data Center Metadata Policy](#): The Arctic Data Center stores metadata using the Ecological Metadata Language (EML). The Arctic Data Center does not require users to submit metadata in an EML format. When submitting to the Arctic Data Center through the

website, submitters can submit plain text descriptions of their data which will be automattically transformed into an EML format for archiving. That being said, submitters should have complete plain text records of the metadata associated with their project. Complete metadata records should contain the following information (among other information) when applicable.

1. descriptions of field and laboratory sampling times and locations
2. descriptions of field and laboratory sample collection methods
3. descriptions of field and laboratory sample processing methods
4. descriptions of any hardware and software used (including make, model, and version where applicable)
5. sampling units
6. quality control procedures
7. explanations for why the particular components detailed above were chosen for this project

Data and metadata formats

What format(s) will data and metadata be collected, processed, and stored in?

Note: if you plan to submit data to the Arctic Data Center please refer to the guidance in the panel on the right.

Example Answer:

The local and traditional knowledge data will be collected during interviews using tape recording devices. The survey answers will then be transcribed and exported to CSV files for storage. Metadata will be documented in CSV files. The soil core data will be collected by hand in the field and then entered into a CSV file for storage. The soil core metadata will be entered in CSV files along with the metadata. Imagery metadata from the core sites will be encoded in NetCDF files. All metadata will be transformed from text into EML files by the Arctic Data Center online submission tool when submitting to the Arctic Data Center.

Guidance:

Arctic Data Center Data Format Policy: The Arctic Data Center primarily supports the upload of open-source, ubiquitous, and easy-to-read data formats. Examples of such formats are Comma Separated Values (CSV) files, text (TXT) files, PNG, JPEG or TIFF image files, and NetCDF files among many others. If you plan to submit to the Arctic Data Center, include your planned methods to create open-source, ubiquitous, and easy-to-read data. If you plan to work with any proprietary data formats such as Excel workbooks or MATLAB files, please include a plan to transform all data stored in these formats into an open-source format before submission to the Arctic Data Center. If you anticipate any data will not be able to be transformed into an open-source format, please provide your reasoning. Additionally, please provide ORCiD identification for all individuals collecting and analyzing the data for proper citation and credit.

Policies for access and sharing

How will data be accessed and shared during the course of the project?

Example Answer:

During the course of the project, data files will be stored in a laboratory Github repository. The data will initially be stored in Excel files and transferred to CSV files. At the end of each year of data collection, the data files will deposited to a shared file system on the Github repository where all team members can access the files.

Guidance:

Arctic Data Center Publication Policy:The Arctic Data Center provides a long-lived and publicly accessible system that is free for users to obtain data and metadata files. Complete submissions to the Arctic Data Center meet the requirements set by the The Office of Polar Programs that require that metadata files, full data sets, and derived data products be deposited in a long-lived and publicly accessible archive. The Arctic Data Center publically releases datasets once all metadata files, full data sets, and derived data products have been submitted and compiled into a unique dataset. Each dataset will be given a unique Digital Object Identifier (DOI) that will assist with attribution and discovery.

NSF Office of Polar Programs Guidelines

Will any of the data and/or related materials produced need provisions for appropriate protection of privacy, confidentiality, security, intellectual property, or other rights or requirements? If so describe them and detail any requested exceptions from the archiving requirements set for Arctic Sciences research.

Example Answer:

Survey data taken during interviews with the local residents are expected to need provisions for confidentiality due to ethical restrictions and the protection of indigenous knowledge. This sensitive data is governed by an Institutional Review Board policy. Additionally, this project deals with endangered species, so similarly sensitive data, particularly location data, will also be exempted from the archiving requirements set for Arctic Sciences research due to confidentiality and species protection.

Guidance:

- NSF Office of Polar Programs Guidelines
- Arctic Data Center Guidelines on who must submit

When is the approximate release date of the data products?

Note: Arctic Observing Network (AON) data must be deposited in a long-lived and publicly accessible archive within 6 months of collection, and Arctic Social Science Program (ASSP) research data must be deposited in a long-lived and publicly accessible archive within 5 years of the award date assuming no exceptions to the archiving requirements are requested.

Guidance:

NSF Office of Polar Programs Guidelines

Example Answer:

The data will be released within two years of data collection.

Policies for re-use and re-distribution

How do you anticipate the data for this project will be used? Consider the following:

1. Which bodies/groups are likely to be interested in the data?
2. What and who are the intended or foreseeable uses/users of the data?

Example Answer:

The salmon catch data are expected to be used by other researchers studying Arctic food web systems in addition to government agencies with regard to establishing catch limits in the area. Local fishermen and fish modellers may also make use of this data. The data will be added to a long-term data set to continue to observe changes in the region through time.

Will any permission restrictions need to be placed on the data? Consider the following:

1. Who will be allowed to use the data?
2. How will others be allowed to use the data?
3. Will others be allowed to disseminate the data.

Note: If you are planning on restricting access, use, or dissemination of the data, you must explain in this section how you will codify and communicate these restrictions.

Guidance:

Arctic Data Center Licensing and Data Distribution Policy: All data and metadata will be released under either the CC-0 Public Domain Dedication or the Creative Commons Attribution 4.0 International License (CC BY), with the potential exception of social science data that have certain sensitivities related to privacy or confidentiality. In cases where legal (e.g., contractual) or ethical (e.g., human subjects) restrictions to data sharing exist, requests to restrict data publication must be requested in advance and in writing and are subject to the approval of NSF, who will ensure compliance with all federal, university, and Institutional Review Board policies on the use of restricted data.

Example Answer:

All data will be accessible to the public and subject to usage and dissemination restrictions under the CC-0 Public Domain Dedication License.

Plans for archiving and preservation

What is the long-term strategy for maintaining, curating, and archiving the data?
Note: The Office of Polar Programs policy requires that metadata files, full data sets, and derived data products be deposited in a long-lived and publicly accessible archive.

Example Answer:

The data manager will follow the NSF Arctic Data Center guidelines to provide accurate and complete documentation for data preservation. The NSF Arctic Data Center will ensure that the data are curated in a relevant long-term archive and ensure data will be available after project funding has ended.

Guidance:

[Arctic Data Center Data Preservation Policy](): The Arctic Data Center ensures the long-term preservation of the data entrusted to the repository. The guiding principles for the

preservation plan follow:

1. Preserve the bits
2. Open science, open standards
3. Replicate data and metadata
4. Strong versioning
5. Frequent auditing
6. A wind down plan