## **Data Management & Sharing Plan Template** Pre-Award Phase

A Data Management and Sharing Plan (DMSP) is a formal document that outlines how research data will be handled, stored, shared, and preserved throughout the lifecycle of a project. The plan should demonstrate the researcher’s commitment to good data management practices (e.g., FAIR: Findable, Accessible, Interoperable, Reusable) and ensure that data are available for future research.

**Directions**: Please provide answers to the following questions about data your proposed work will collect, contribute, gather, analyze, and/or produce. Guidance and an example answer are provided to help shape your response. You may delete the guidance and example answer to replace it with your response. If selected and awarded for funding, ROSA and the project team will work together to complete a full, more detailed Data Management & Sharing Plan.

A glossary of terms can be found at the end of this document.

Project Information

Title of Proposal:

PI Name (Name, Title, Organization):

Assigned Data Manager (Name, Title, Organization):

**Expectations of the Data Manager**

If awarded, the designated Data Manager will perform the following duties:

* Complete a full Post-Award Data Management & Sharing Plan (DMSP) as provided by and in collaboration with ROSA within the first six months after project start date
* Update the DMSP at least once per year
* Communicate directly with ROSA about the data of the proposed work, data products produced, and data management throughout the project.

## Types of Data Produced & Contributed

1. Describe the types of data and derived data products to be collected and/or created during the project.

* What types of data, samples, collections, software, visualizations, apps, etc. will be **produced** during your project?
  + **Guidance:** A summary of all data and data products should be included here. This includes raw data, derived and processed data, maps and other visualizations, GIS layers, video files, photos, and sound recordings. This should be the most detailed and largest section of the plan.
  + **Example Answer:** This project will produce datasets, tissue samples, and an app. Datasets will include water temperature, water depth, water clarity/turbidity, dissolved oxygen, phytoplankton, zooplankton, fish morphology and weight. Tissue samples of fish will be collected and stored. An R Shiny app will be built to share data visualizations with a larger audience.
* What types of metadata will be collected during your project?
  + **Guidance:** List the types of metadata, such as sensor metadata, data file metadata, project metadata, tag metadata, etc.
  + **Example Answer:** We will create project-level metadata for FishFORWRD (e.g. objectives, lead entity, receptor, methodology, etc.), file-level metadata (e.g. units and definitions of column names for each column in the dataset, etc.), sensor metadata (e.g. info about sensor calibration and deployment/retrieval).
* What established protocols (e.g., sampling or collection) will be used to collect and analyze the data for each type of data and derived data product?
  + **Guidance:** List any established or developed protocols. Include references and protocols for quality assurance and control.
  + **Example Answer:** We will follow catch sampling protocols established for the VIMS NEAMAP trawl survey. Data will undergo quality control measures to verify accuracy and consistency. This includes data validation checks, peer review, calibration of sensors, and data verification. Data anomalies or errors will be flagged and corrected, and quality assurance logs will be maintained.
* Will any of the data require access restrictions to protect personal privacy, confidential or proprietary business interests, Indigenous data sovereignty, national security, intellectual property, protected species, or other rights or requirements.
  + **Guidance:** Describe any restrictions, embargoes, or other methodsthat will be applied to the data, their justification, and duration.
  + **Example Answer:** No, this project has no requirements to apply a specific restriction to any data. However, our institution does require that data be shared as openly as possible.

1. Describe the types of available data that will be **brought to** the project to perform the proposed work.

* What types of data, samples, collections, software, visualizations, apps, etc. will be brought and leveraged in the proposed project? How will they be accessed?
  + **Guidance:** A summary of all data and data products brought to the project should be included here. This includes raw data, derived and processed data, maps and other visualizations, GIS layers, video files, photos, sound recordings, etc.
  + **Example Answer:** This project will use previously collected bottom temperature in the \_\_\_\_\_\_\_\_\_\_\_\_ lease area held by and accessible from \_\_\_\_\_\_\_\_\_\_\_.
* Does the institution, repository, or organization where the data are stored have or require a specific data sharing agreement or license?
  + **Guidance**: Describe any licensing requirements from repositories or other organizations.
  + **Example Answer:** Yes, this project will use data from the \_\_\_\_\_\_\_\_\_\_\_\_ data repository, licensed under CC BY-NC-ND 4.0. The PI has protocols in place to maintain compliance with this license.

## Data Access, Sharing, and Reuse

1. Describe the access, sharing and reuse potential for data associated with the project.

* What types of data could be shared beyond the project and how will they be shared?
  + **Guidance:** Describe the types of data that will be shared.
  + **Example Answer:** All types of data collected by this project will be shared according to mechanisms agreed upon with ROSA in the Post-Award DMSP. All data brought to this project will be shared in accordance with the requirements of the data license described above.
* What other scientific question/problem could be addressed with your data?
  + **Guidance:** List anticipated cases for reuse of the project data.
  + **Example Answer:** These data could be reused for regional analyses in the future. Water quality parameter data might be useful for future studies looking at habitat for water column or benthic organisms. Fish detection data might be useful in a larger compilation of biogeography or species presence in the study area.

## Plans for Stewardship & Preservation

1. Describe the plans for data stewardship and preservation.

* What is the long-term strategy for maintaining, curating, and archiving the data?
  + **Guidance:** Describe the plans for stewarding the data for the next several decades.
  + **Example Answer:** Initially, data will be stored locally on PI’s computer hard drives, with an additional copy on hard drives on the main campus. Data will be published in repositories that have strong sustainability plans and will preserve data for future use.
* Which scientific data from the project will be preserved? Provide the rationale for this decision.
  + **Guidance:** List the types of data that will be preserved.
  + **Example Answer:** All scientific data collected by this project will be preserved. The rationale is that all of the data could be useful for other studies and scientific research in the future, and therefore we will make them available through publication in repositories or according to mechanisms agreed upon with ROSA in the Post-Award DMSP.

## Glossary

#### **Data License** - A legal framework that specifies how data can be accessed, used, and reused.

#### **Data Management** - The tasks involved with documenting, maintaining, updating, and publishing data.

#### **Data Products** - Models, maps, predictions, analysis results, visualizations and other syntheses that are produced from raw and/or derived data.

#### **Data Repository -** A persistent, findable, searchable entity that provides infrastructure for long-term storage and access to data. It should provide for data publication by data holders, as well as access for using/reusing data. Datasets in a repository are described with metadata that provides essential information about the data and enables efficient search and reuse. It should also employ unique identifiers, have redundancy, and have clear guidelines for what types of data it accepts. May or may not be federated with other repositories. See the Encyclopedia of Big Data.

#### **Data Sharing Agreement** - A legal agreement between two or more parties that defines the terms and conditions for sharing non-public data.

#### **Data Stewardship** - The actions involved in supporting, promoting, and guiding the use and reuse of data throughout the data lifecycle.

#### **Derived Data** - Existing data that have been processed, combined, assembled for the purposes of analysis, visualization, or publication. Derived data often refers to data resulting from the compilation or assembly of raw data into analysis-ready data to answer a particular question. Sometimes also referred to as processed data, though that often has more limited meaning. Examples include text or data mining, and compiled datasets.

#### **Metadata** - Data about the data. The information that describes the data, both at the file-level (column headings, etc.) and the project level (data creator, geographic location, year, etc.), and enables someone to know enough about the data to be able to reuse it. *Federal Government Definition:* Metadata is structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage an information resource (NISO 2004, ISBN: 1-880124-62-9).

#### **Raw Data** - Data from a sensor, observation, or experiment that has not been cleaned or processed. This data may be uncalibrated, unpacked and compressed, and not QAQC’d.