Cheat Sheet

Data Management Planning Guide

WHY IT MATTERS

A clear and comprehensive data management plan ensures that research data are reproducible, accessible, and reusable, meeting funding agency requirements and adhering to scientific best practices. Implementing such a plan promotes transparency, facilitates effective collaboration, and supports long-term data stewardship

THINGS TO CONSIDER

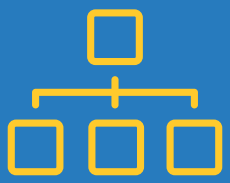
Data Planning

- What are the funding agency's expectations for data requirements, and what is included in the data management plan?
- What data will be included in the analysis?
- How will we document metadata, including dataset sources, permissions, and restrictions?
- What is the timeline for creating and publishing the data product?
- Who is responsible for maintaining metadata, updating data files, and coordinating data publication?

Cheat Sheet

Data Management Planning Guide

Data Organization & Sharing



- What storage platforms will we use (e.g., NCEAS Google Drive with daily backup, server, or other options)?
- Who will have access to different data types (e.g., raw vs. processed, sensitive vs. public)?
- How will we organize data into folder tiers (e.g., raw, clean, harmonized, analysis-ready)?
- What file naming conventions will we follow (e.g., Biomass_YYYYMMDD.csv)?
- What standard data formats will we use (e.g., CSV, GeoJSON, NetCDF)?
- How will we document column-level metadata (e.g., variables used for analysis, descriptions, units, missing data codes)?
- What platform will we use to discuss data issues and updates (e.g., GitHub issues)?

Data Publication

- How will we provide source data provenance (e.g., citing authors, year, dataset title, repository, and DOI in our data package)?
- Which trusted repository will we use for data publication (e.g., EDI, NCEI, Arctic Data Center)?
- What data license is appropriate (e.g., CC0, CC-BY)?
- How will we handle publication of sensitive data, if applicable?