

Appendix 2. Data Management Plan

NOAA's mission to understand environmental intelligence; to share knowledge and information; and to manage water ecosystems and resources depends on the ability to collect relevant information and on data management workflows that apply data stewardship techniques to data assets. Proper data stewardship techniques ensure that collected data are: 1) documented; 2) accessible by humans and machines; 3) and archived to ensure reusability and reproducibility. These stewardship principles parallel FAIR data principles: findable, accessible, interoperable, and reusable for future generations.

Proper management and stewardship of data crosses all disciplines and scientific research themes, with data interoperability a critical element of efficient science. The ability to access, analyze, and visualize increasing volumes of data, and to graphically compare disparate data sources and model predictions through on-demand cloud services, will be common in the next decade. But the ability to accomplish interoperability is predicated on proper data management techniques, data and metadata best practices, and adherence to community-supported data principles. To optimally leverage the societal and economic potential of rapidly increasing data volumes, user communities need to efficiently discover, access, and understand relevant data. Innovative data management techniques ensure extraction of maximum knowledge for each datum collected and, in the context of this proposal, power collaboration with NOAA to fulfil its mission.

2.1 Data Management Philosophy

CIROH is committed to a data management and sharing plan that addresses the intent of NOAA's 'Data and Publication Sharing Directive' and supports NOAA and DOC policies that require public accessibility of data collected under NOAA grants and cooperative agreements within two years of acquisition. Investigators will be encouraged to make data available as early as possible, and no later than two years after collection. Data will be provided free of charge or at no more than the cost of reproduction. Exceptions or extensions will be explicitly justified. Digital data will be made available in standard formats (such as NetCDF and CSV) that are machine readable, accessible through standard web services, and accompanied by machine and human readable metadata. Data will be archived, where possible, to NOAA's NCEI and allocated a Digital Object Identifier (DOI) to enable citation of the data when used by others. The proposed project will collect human dimension data, which will be afforded guarantees of privacy and confidentiality. These data consist of audio recordings of interviews, transcripts, tabulated workshop survey responses, and discussion notes. Results from human dimension studies will not contain personal identifiable information. All data collected during this research pertaining to individuals is confidential and restricted following the consent procedures and agreements set forth in the Consortium Member and Partner Institutional Review Board(s) (IRB) for the protection of human subjects. All PIs and their contributors working on any project with approved IRBs will strictly adhere to the requirements specified for data collection, data handling, and data disposition. Metadata will record the location, date, and other pertinent information. Tribe specific data generated by the project will belong to each tribe under separate data sharing agreements; such data will follow protocols of Free, Prior and Informed Consent. Use, sharing, and management of tribe-specific data are restricted to protocols to be set forth in legally binding agreements prior to starting research.

2.2 Data Sharing Policy

CIROH data collected during the Cooperative Award meet NOAA's definition of Environmental Data. While it is difficult to predict the evolution of data standards and formats, metadata and data produced through CIROH conform to existing community-accepted standards and conventions. When standards and conventions do not exist, data will be stored in the most appropriate accessible format and documented. Where applicable, software developed will be made available through open source repositories. This includes software developed for automated quality control using ML techniques, and visualization software to explore and analyze data. To support reproducibility in model development and modeling we will ensure that model code, model parameterization schemes, and initial conditions are archived and provided DOIs, where relevant. Data format standards will include NetCDF, ASCII CSV, and Open Geospatial Consortium (OGC) data formats such as Web Mapping

Service (WMS) and Web Coverage Service (WCS). Metadata conventions such as Climate and Forecast (CF), Darwin Core, Observations and Measurements (O&M) are all supported, as are metadata standards such as ISO 19115. As metadata conventions and standards evolve, CIROH will ensure that data conform to those standards and conventions. CIROH will strive to ensure that data collected conform to the FAIR data principles and continue to conform as those principles evolve.

Access to CIROH-created data will be provided through both human and machine accessible services. Access methods will include online data portals, interactive web pages, and other web tools. Machine-to machine data access will be available where appropriate and will be critical to support Machine Learning, cloud technologies and other advancements in computing infrastructure. We will support the Data Access Protocol (DAP) and new tools such as ERDDAP, Jupyter Notebooks, and collaborative communities such as Pangeo.

To support reproducibility of research, and reuse of data, CIROH will develop processes to ensure data are archived at a national data center (e.g. NCEI). Data and complete metadata will be archived and, where relevant, provided a DOI for later citation. Leveraging these principles CIROH will ensure that data from the new CI are discoverable through national data repositories and discovery tools (e.g. data.gov, NOAA OneStop).

Final pre-publication manuscripts of articles produced entirely or primarily with NOAA funding will be submitted to NOAA Institutional Repository after acceptance, and no later than upon publication. A catalog of peer-reviewed publications will be maintained and made publicly available on the consortium website.

2.3 Data Discovery, Access, and Dissemination

Specific research data, algorithms, and codes to be produced through CIROH tasks will be shared as follows:

1. Software: All software developed through this project will be made available on open-source online repositories. We expect CIROH will develop open-source code development (e.g., Github). A similar web-based repository will be established for CIROH to enable a community of co-developers across the Institute to contribute code.
2. Transcripts, videos, survey responses, observational notes, and photos: All data collected in social science research activities will be released pending participant consent and, where appropriate, de-identification procedures. Codes and codebooks will be publicly available.
3. Curricular materials for education and outreach: These consist of video lectures, slides, and other items and will be available upon request to the investigators.
4. Results will be disclosed to the scientific community through presentations in disciplinary scientific meetings, with long-term standard publication of scientific results in 1) peer-reviewed publications and 2) academic theses. Publication of data shall occur during, if appropriate, or at the end of individual tasks. If requested, data will be made available for sharing to qualified parties by the PIs as soon as reasonably possible, so long as such a request does not compromise intellectual property interests, interfere with publication, invade subject privacy, or betray confidentiality. Shared data will include notations needed to interpret the data, following accepted practices in the field. Following NOAA data policy guidelines, all data will be released within two years of collection.
5. Re-use, re-distribution, and derivative products: Publicly shared data will not contain any disclaimer for the use of the data in other publications or products. Curriculum materials will encourage re-use and the development of derivative products.