NOAA NCCOS Federally-Funded Project General Data Management Guidance

This document details general data management guidance, requirements, and checklists for ensuring that datasets and publications produced by projects awarded under a NOAA National Centers for Coastal Ocean Science (NCCOS) Federal Funding Opportunity (FFO) are publicly available past the life of the funded project.

Data Management Planning

All proposals are required to include a data management plan (DMP) that includes the following information (refer to the FFO for details):

- clear intent to share data in a timely manner
- data deliverables:
    - expected types of data, formats, and total volume
    - metadata standards, format, and content
    - data access approach:
        - submission to NOAA NCEI for long-term preservation (archiving) **OR**
        - submission to another appropriate data facility
- data milestones:
    - data collection dates
    - derived data product completion dates
    - QA/QC completion dates
    - data delivery dates
- prior experience in data management

Proposal writers may request a consultation with the NCCOS Scientific Data Coordinator (nccos.data@noaa.gov) during the proposal development phase in order to better understand the data management requirements and improve their DMP.

Data Management Reporting

Semi-annual progress reports and final reports should include details about data deliverables and milestones:

- Status (not started, data collection/production in progress, undergoing QA/QC, in preparation to be submitted to a data facility, submitted, publicly available, or limited release)
- Access location of all publicly available datasets and data services (data facility, internet address, accession number, and/or DOI)

Data Delivery

Final delivery of data to NOAA (or other appropriate data facility) should be completed according to the agreed-upon DMP data submission date(s), which may be up to two years after the original end date of the award. Some projects may also implement additional data sharing approaches, such as web services, or additional data delivery to partner or regional data facilities - all such approaches should be completed during the life of the project.

Project leads and data managers are encouraged to request a consultation with the NCCOS Scientific Data Coordinator (nccos.data@noaa.gov) during the project implementation phase for specific guidance on data formatting, documentation, organization, and delivery.

Data Format

Datasets should be delivered in a non-proprietary format:

- Tabular data (spreadsheets):
    - CSV – comma-separated value (preferred)
    - TSV – tab-separated value
    - ASC, TXT – ASCII
- Geospatial data (ArcGIS):
    - Vector data: shapefiles (SHP) and associated ancillary files:
        - SHX – shape index (mandatory)
        - DBF – attribute format (mandatory)
        - PRJ – projection (recommended)
        - SBN, SBX - spatial index
        - CPG – character set
        - XML – metadata taken from the file
    - Raster data: GeoTIFF (TIF) files and associated ancillary files:
        - TFW – GeoTIFF world file (mandatory)

        - OVR – GeoTIFF overview (mandatory)
        - AUX – auxiliary information
        - XML – metadata taken from the file
- Acceptable complex data file formats: BUFR, FITS, GRIB 1, GRIB 2, HDF4, HDF5, HDF-EOS, IMMA, McIDAS Area, McIDAS Grid, netCDF-3, netCDF-4, PNG, SVG
- Acceptable compression algorithms: bzip2, compress, gzip, netCDF-4, HDF5

Data Documentation

All data deliverables must include data documentation (*i.e.*, metadata) that ensures potential future users can discover, use, and understand datasets. Data documentation may be delivered as plain text, ISO XML, or other appropriate format (please note that FGDC XML is acceptable but deprecated and superseded by ISO XML format). In general, data documentation should contain the following information:

- Project Information
    - Project Description: Project overview, purpose, and partnerships
    - Grant Funding Statement: For example, "Data in this accession was funded by the NOAA RESTORE Science Program under award XXXXXXXXXXXXXX to the University of Alaska."
    - Principal Investigator(s): Name, email, affiliation
    - Funding: List all funding entities that supported this dataset
    - Related Projects and Project Webpages (if any)
- Dataset Information
    - Dataset Title
    - Dataset Author(s) and Primary Point of Contact: Name, email, affiliation
    - Collaborators: Name and affiliation of any individuals not otherwise listed who should be recognized for their contribution to this dataset
    - Description: Brief description of the dataset, including spatial and temporal extent (if applicable), parameters, and format
    - Purpose: Brief explanation of why this dataset was collected or derived
    - Methods: Brief description of observation sampling methods and/or model approach, including reference to any publications
    - Time Period: Start date, end date (YYYY-MM-DD)
    - Location: North, south, west, east boundary latitude/longitude (dd.ddddd)
    - Cited Publications, External Data Sources, and Associated Web Services (if any)
    - Keywords:
        - Scientific keywords
        - Water Bodies, U.S. States and Territories, Marine Protected Areas

        - Vessels, Platforms, and Instruments
- File Information
    - Data Files: Provide the file name, size, format, compression, resolution, GIS projection (if applicable)
    - Data Dictionaries: Each column of a spreadsheet or layer of a GIS dataset should be described, including column/layer name, label/code, definition, units, range; similar files may have a single data dictionary
    - Preview Graphics: An image file (JPG or PDF) of a representative graph (*e.g.*, time series) may be included to help users preview tabular data, and/or an image showing a representative geospatial layer with symbology may be included to help users recreate GIS layers as intended
    - Documentation: Provide the file name(s) for all documentation files, including plain text documentation (PDF), reports (PDF), machine-readable metadata (XML), data dictionaries (PDF, CSV), and preview graphics (JPG, PDF)
- Parameter Information: Each major observed or derived parameter should be described
    - Parameter Description: Name, units, sampling instrument (if applicable)
    - Sampling Method: Measurement, collection, and/or sampling methodology
    - Data Quality Method: Data processing and analysis methodology

Data Organization

Once properly formatted and documented, data packages of data files and documentation should be assembled for final delivery. Use any folder structure and file name convention, but do not use any spaces in your folder or file names (use dashes or underscores instead). Data packages must also include a browse graphic that exemplifies the data (JPG format, less than 500KB and 1000x1000 pixels).

Example complete data package:

- PILastName_DataPackage.ZIP (zipped file):
    - File Folder 1: Dataset1_Observed-parameters
        - Data1_Parameter1_CruiseLeg1.CSV
        - Data2_Parameter1_CruiseLeg2.CSV
        - Data3_Parameter2_CruiseLeg1.CSV
        - Data4_Parameter2_CruiseLeg2.CSV
        - DataDictionary1_Parameter1.CSV
        - DataDictionary2_Parameter2.CSV
        - Documentation1_CruiseReport.PDF
        - Documentation2_InstrumentMetadata.XML

        - PreviewGraphic1_TimeSeries.JPG
        - PreviewGraphic2_ShipTrack.JPG
    - File Folder 2: Dataset2_Model-output
        - Data5_ModelScenario1.GRIB
        - Data6_ModelScenario2.GRIB
        - DataDictionary3_ModelOutputHeader.TXT
        - Documentation3_ModelDocumentation.PDF
        - PreviewGraphic3_Scenarios.PDF
    - File Folder 3: Dataset3_Derived-data-product
        - Data7_GISProduct.ZIP (zipped file)
            - GISProduct.TIF
            - GISProduct.TFW
            - GISProduct.OVR
        - Documentation4_GISProductDocumentation.PDF
        - PreviewGraphic4_Layer1.PDF
        - PreviewGraphic5_Layer2.PDF
        - PreviewGraphic6_Layer3.PDF
    - BrowseGraphic.JPG

Data Access

Pre-release data may be shared with collaborators, partners, and stakeholders at the discretion of the PI via any local data access methods, *e.g.*, FTP, web service, data portal, *etc*. In rare cases, data or derived products may be limited for public release by law, regulation, policy (such as those applicable to personally identifiable information or protected critical infrastructure information or proprietary trade information), security requirements, commercial or international agreements, or valid technical considerations. PIs must request permission from the Science Program to not to make any awarded project data publicly accessible; Data Access Waivers must be approved by NOAA (request via email to nccos.data@noaa.gov), updated on an annual basis, and revised or superseded as needed.

Finalized data that has undergone quality control should be formatted, documented, and organized into data archive packages as described above. Packages may be submitted to NOAA (via email to nccos.data@noaa.gov) for archiving at NCEI; upon acceptance by NCEI, archived data packages will be made available to the public with an Accession Number and digital object identifier (DOI). Data packages may submitted to another appropriate data facility (*e.g.*, NIH GenBank for genomics data) that makes the data publicly available; these should be

documented with appropriate metadata, and a copy of the documentation should be delivered to NOAA (as an attachment to a semi-annual or final report).

Additional data sharing approaches, such as web services (see below) or web applications, should be documented with service-level metadata and a copy of the documentation should be delivered to NOAA (as an attachment to a semi-annual or final report). Non-archivable digital data such as video, still photographs, *etc*., should be documented and delivered to NOAA (via email to nccos.data@noaa.gov), which may make those data available in the future as archiving capabilities expand. Non-digital data such as biological specimens, preserved samples, paper or analog records, *etc*., should be documented (as appropriate to the type of data) and held by the PI's institution or other appropriate facility, and a copy of the documentation should be delivered to NOAA (as an attachment to a semi-annual or final report).

Web Mapping Service Access

Geospatial data archived at NCEI will be publicly available for bulk file download. While this is adequate in terms of meeting minimum NOAA requirements for data management, adding Web Mapping Service (WMS) access makes geospatial map image data more widely available through online interfaces that allow users to visualize the data in map form, along with other map layers. The two main steps to the process are (1) create the map document .mxd file; and (2) create a map-service metadata record .xml. WMS packages (.mxd file and .xml file) may be submitted to NOAA (via email to nccos.data@noaa.gov) for hosting on the NCCOS MapServer maintained by NCEI-Stennis.

The required file format for saving maps is the Esri map document file, or ***.mxd**. The .mxd file contains information needed to render your map, including layers, symbology, projection information, etc. Note that an .mxd, and the web map service created from it, can contain a single map layer or multiple map layers. For example, it may be useful to bundle several maps into a project-level map service. From a performance standpoint, fewer layers per service is desirable, so there may need to be an evaluation of the tradeoffs. Below are some **best practices** for preparing .mxd files within the ESRI ArcGIS environment that lead to better results:

- Fewer layers in one .mxd project file are better to optimize retrieving and drawing of initial map service view. If possible, group map services by data type especially if there is a layer that is used across several different map services. If that layer is in one service instead of multiple map services, it will be easier to maintain.
- Make sure your .mxd file has been saved out using the WGS 1984 Web Mercator (auxiliary sphere) coordinate system (102100 (3857)). For reasons of efficiency when the

map is rendered, it is advisable (but not mandatory) that the data layers within the .mxd are in this projection as well. The projection uses should be noted in the service metadata template as well.

- The name of an .mxd file can not contain any blanks, dashes, or other special characters. The exception to this rule is the use of underbars (_).
- Data layer names within the .mxd can contain spaces, underbars, and dashes. Parentheses are forbidden. This is more of an OGC restriction (WMS) than an ArcServer restriction.
- The field should contain user-friendly field aliases (i.e. for "T_COMPLEX", use "Topographic Complexity"). Also, include units of measurement ("Area (m²)" instead of "Area"), especially if it is different than the units of the data's projection.
- Turn off all unnecessary fields. Sometimes, extraneous field names accumulate as data are exported to a shapefile or geodatabase and back, or are the product of geoprocessing tools. There might be a field "Shape_Length" and "Shape_Leng". Choose which field to display by opening the Layer Properties dialog, selecting the Fields tab and unchecking the fields you don't want to show. The OBJECTID and SHAPE fields must be turned on.
- Make sure the OBJECTID field is called "OBJECTID" (and not by using a field alias)
  Do not incorporate symbol levels when defining the layer properties. That is an advanced feature in ArcMap, that results in a dramatic slowdown in rendering the map. A better option, if symbol levels are required, is to include multiple instances of the same layer in the .mxd file with different definition queries applied. This approach has less impact on drawing speed.
- Make sure to fill out as much detail as you can within each of the following sections of the .mxd. These will be used to create the service-level metadata:
    - ap Document Properties: make sure to fill out title, description, author, credits, and tags. These are used within the REST service pages and are great references to the map service itself.
    - ata Frame Properties: fill in description and credits. These also can provide information useful in the REST service pages. The description could be the same as the map documents properties or it can be different.
    - Layer Properties: make sure to fill in the description and credits. Again useful to the REST service pages and can be unique to the layers themselves.
- Remove basemap and reference layers. These are not able to be published to a map service.
- Check the source of all layers to ensure they are pointing to a single network location. Do not point to any data on a local drive. This location will be copied to a separate

network drive accessible by NCEI, so remove any extraneous data not associated with the mxd from that location.

A service-level metadata record has general citation, contact, and identification information, with an added section describing service properties as identified from the .mxd file. Use the service template (example) to create your web service metadata.

Publications

Journal articles should include a funding acknowledgement including the relevant Grant Program and Award Number, for example: "This article is a result of research funded by the NOAA RESTORE Act Science Program under Award XXXXXXXXXXXXXX to the University of Alaska and Award XXXXXXXXXXXXXX to the University of Hawaii." All publications, including peer-reviewed journal articles, book chapters, NOAA Technical Memoranda, conference proceedings, *etc*., resulting from the awarded project should be delivered to NOAA (as an attachment to a semi-annual or final report) with the following:

- Full citation, including DOI if applicable
- PDF:
    - Open access publications: final published "post-print" PDF
    - Copyrighted publications: both the final published "post-print" PDF **AND** the draft "preprint" PDF; the draft version of the manuscript after it has been peer-reviewed and revised by the author, but before the publishers formatting has been added
- Note additional information:
    - whether the publication is open access or copyrighted
    - whether the publication was peer-reviewed
    - any supplementary data published with the publication

Data Delivery Checklist

Status	Item	Date Completed
	1. Consultation with the NCCOS Scientific Data Coordinator (nccos.data@noaa.gov)	
	2. Data collection and/or derived data product development completed, and:	
	2A. Pre-release data shared with collaborators, partners, and/or stakeholders (if applicable)	
	2B. Limited-release data approved via Data Access Waiver (if applicable)	
	3. QA/QC completed	
	4. Complete data package assembled, including:	
	4A. Final QA/QC'd data files created in non-proprietary format	
	4B. Data documentation created	
	4C. Data dictionaries created (if applicable)	
	4D. Preview graphics created (if applicable)	
	4E. Browse graphic created	
	5. Data package delivered to nccos.data@noaa.gov (for submission to NOAA NCEI) or other appropriate data facility, and:	
	5A. Data package accepted by data facility	
	5B. Accession Number and/or DOI created by data facility	
	5C. Data available from data facility	
	6. Full data citation (with internet location, accession number, and/or DOI) provided in semi-annual or final report, and:	
	6A. Attached metadata submitted to a non-NOAA data facility (if any)	
	6B. Attached web service metadata (if any)	
	6C. Attached non-digital data documentation (if any)	
	7. Non-archivable data documented and delivered to nccos.data@noaa.gov (if any)	

Publication Delivery Checklist

Status	Item	Date Completed
	1. Funding acknowledgement included in manuscript	
	2. Final revised manuscript accepted for publication	
	3. Full publication citation (with DOI) provided in semi-annual or final report, and:	
	3A. Indicated whether peer-reviewed or not	
	3B. Indicated whether supplementary data was included	
	3C. Indicated whether open access or copyrighted	
	3D. Attached final published “post-print” PDF	
	3E. Attached final revised manuscript "preprint" PDF (copyrighted publications only)	