NOAA Data Management Planning

Procedural Directive

Version 1.0

VERSION NOTE: This is version 1.0 of this NOAA procedural directive. Before you proceed with implementation of this directive, we recommend that you check to be sure this is the most recent version available. You can check to see what the current version is, download any updates and access additional implementation resources at the following permanent URL:

https://www.nosc.noaa.gov/EDMC/PD.DMP.php

NOAA Environmental Data Management Committee

October 2011

VERSION HISTORY

Version	Implemented By	Revision Date	Approved By	Approval Date	Reason
1.0	James Sargent	Initial Draft	EDMC	2011-10-24	Approved by EDMC (was draft version 0.7.2)

Table of Contents

1. Purpose

NOAA Administrative Order (NAO) 212-15, Management of Environment Data and Information, states that environmental data is to be managed based upon a lifecycle that includes developing and following a data management plan. The purpose of this procedural directive is to expand upon this requirement for Data Management Plans (DMPs), direct managers of all data production projects and systems to write data management plans, and provide detailed guidance on who should write them and how to write them. It contains a template with questions to be considered and addressed in planning all data production projects.

The goal of the Data Management plan is to ensure that data are properly collected, documented, made accessible, and preserved for future use in a NOAA Data Center or other long-term archive facility.

2. Directive

- Managers of all environmental data production projects, including observing systems, projects, programs, surveys, collections, interviews, investigations, studies, and research that produce environmental data, shall write data management plans before starting to acquire data. The plans are to be followed and maintained throughout the full lifecycle of the data. The scale and complexity of the plan will vary with the scale, complexity, and significance of the data being produced and will address the questions included in the template (Appendix A) at an appropriate level of detail.
- Data Management Plans shall be stored in the Data Management Plan Repository established by NOAA's Technology Planning and Integration for Observations (TPIO) program.
- Key elements of the plan are to be summarized on one slide for all investments presented to the NOAA Information Technology Review Board (NITRB), as described in NOAA IT Standard (NISN) 3.007, Appendix H.

3. Background

According to NOAA's *2010 Next Generation Strategic Plan*, our mission is "to understand and predict changes in climate, weather, oceans, and coasts and to share that knowledge and information with others, ..." Environmental information is therefore NOAA's primary product, and it is our responsibility to produce that information, manage it as a national asset, and share it with the American people and the world.

This Procedural Directive has been developed by the Environmental Data Management Committee (EDMC) in collaboration with leaders of NOAA's observing systems, data managers, and other key stakeholders. It is the EDMC's intention to provide clear guidance on what needs to be done, how to do it, and how to get help. The authority of the EDMC to issue directives such as this one derives from the NOAA Chief Information Officer Council (CIOC) and the NOAA Observing Systems Council (NOSC). The directive will be revised after one year based on experience and feedback gathered during implementation.

4. Overview of Data Management Plans

4.1. What is a Data Management Plan?

In the context of this directive, a data management plan is a document that specifies key attributes of the lifecycle of environmental data. This information may be provided in the form of answers to the questions in the NOAA Data Management Plan Template (Appendix A), or in an equivalent document that answers these questions while satisfying other project requirements. Environmental data is defined by NAO 212-15 as recorded and derived observations and measurements of the physical, chemical, biological, geological, and geophysical properties and conditions of the oceans, atmosphere, space environment, sun, and solid earth, as well as correlative data, such as socio-economic data, related documentation, and metadata.

Data management plans should be flexible and may be hierarchical. "Flexible" means they can be modified relatively easily in response to changes in technology, policies, and business processes. "Hierarchical" means they inherit content from higher-order plans or formally replace inherited content with alternative guidance when appropriate. Generic higher order plans provide guidance, context and common requirements for plans that are developed within their specific spheres of influence. For instance, if NOAA chooses to develop a NOAA-wide DMP, all other plans would incorporate the guidance and standards articulated in that plan. Similarly, NOAA Line Offices and Staff Offices may choose to develop plans that provide templates for all data acquisitions in that LO. Program offices may do the same to ensure consistency and interoperability among their data production projects.

Plans must not be ignored after development. Rather, they must be followed and maintained throughout the lifecycle of the project.

4.2. When Should a Plan Be Developed?

The Data Management Plan should be initiated in the overall project-planning phase and updated as appropriate throughout the data management lifecycle. Plans should be developed for all data production projects. The level of detail of these plans may vary depending on the scale of the project. Comprehensive data management plans, which should address all the questions in the template, are needed for:

- New investments that involve significant data collections, such as a new observing systems and research projects.
- Legacy data collections that have not completed the full lifecycle.
    - Major IT Investments tracked by the OCIO that produce data.

In the planning phase of new data production projects, at least a subset of the questions in the template must be addressed to ensure that adequate resources are allocated to support the full lifecycle of the data from planning and acquisition to preservation. Data collection/acquisition should not be started unless the full life cycle has been planned. It may not be practical to answer some questions until later in the project. Unanswered questions should be noted as either TBD (to be determined later in project) or NA (not applicable to this project).

4.3. Where Will Plans be Stored?

Copies of all NOAA Data Management Plans will be stored in the NOAA DMP Repository. The repository will provide secure, managed storage for these critical agency documents. The plans therein will serve as a resource for those who are developing new plans.

The Repository will be on the Global Earth Observation Integrated Data Environment (GEO-IDE) Wiki at https://geo-ide.noaa.gov/wiki/index.php?title=Data_Management_Plans. This space is managed by NOAA's Technology Planning and Integration for Observations (TPIO) program. TPIO also supports the EDMC and the NOAA Data Management Architect.

5. Review and Approval of the Plan

The reviewers and potential audience of a particular Data Management Plan depend on the phase, scope, and nature of the project and the data involved. Below are some examples of reviewers and audiences for different types of data production projects:

- New observing systems – Data management plans for new and or enhanced observing systems may need to be reviewed or approved by the NOAA Observing Systems Council and the Chief Information Officers Council.
- New or enhanced IT investments that produce data – As part of NOAA IT investment capital planning and investment control policies, all major IT investments (above $10M) are reviewed by the IT Review Board. The content of the review package includes a summary of the data management plan.[1] Investments below $10M require Line Office or Staff Office CIO approval.
- Project-Specific Data Management Plans – High level plans developed by NOAA Line Offices and Staff Offices as templates for project-specific plans may be reviewed by the EDMC for feedback at the request of the Offices.
- Line Office/Staff Office Data Management Plans – Review and approval of DMPs for specific projects will be reviewed and approved by supervisors of the project managers.

6. Writing the Plan – The Steps

1. **Write the Plan.** Use the NOAA Data Management Plan Template as a context and resource. Responding to questions may involve collaborating with your supervisors, data managers, Line Office or Staff Office CIO, IT support, and Data Centers, among others.
2. **Vet the Plan.** Circulate for review and comment by the appropriate NOAA staff, management, Boards and Councils.
3. **Store the Plan** into the Data Management Plan Repository
4. **Implement the Plan.** Identify an individual responsible for ensuring the plan is followed. Align the steps in the plan with the overall project schedule.

[1] NOAA IT Review Guidance (NOAA IT Standard Number 3.007), 2011.

5. **Maintain the Plan.** As situations, technology, budget, or staff change, keep the Plan up-to-date throughout the life cycle of your project while the data are collected, documented, managed, stored, distributed, and preserved. Update the copy stored in the DMP Repository.

7. Sources, Dependencies and Interfaces

Data management planning should be done in the context of other NOAA policies, procedures, and best practices. Below are some particular examples:

- NOAA Data and Information procedural directives. Data management plans must conform to all procedural directives that implement NAO 212-15.
    - Procedure for Scientific Records Appraisal and Archive Approval Procedural Directive. This "what to archive" document defines the procedure to identify, appraise, and decide what scientific records are preserved in a NOAA archive.
    - Data Documentation Procedural Directive [link to be added when available] – currently under development by the EDMC
    - NOAA Data Sharing Policy for Grants and Cooperative Agreements [link to be added when available] currently under development by the EDMC and NOAA management
    - Future Procedural Directives
- Paperwork Reduction Act (PRA) Requests – All PRA requests involve some level of discussion on how the information collected will be managed. DMPs should reflect this discussion if applicable.
- NOAA's IT Security Plans – All NOAA IT systems have approved Security Plans that are required by the Office of Management and Budget (OMB) Circular A-130, *Management of Federal Information Resources* and Public Law 107-347, the Federal Information Security Management Act (FISMA). These plans contain a wealth of information about the IT infrastructures in which NOAA data reside and provide information applicable to confidentiality, integrity, and availability requirements. Confidentiality, integrity, and availability requirements need to be specified and given special consideration in DMPs. Contact your LO IT Security Officer for more information.

8. Where Can I get Help?

There are a number of sources for support in developing data management plans including:

The Data Management Planning Wiki [link to be added when available] – The wiki will be developed over time by discovering and documenting what works and doesn't work in writing plans. The wiki will contain Frequently Asked Questions (FAQs) to assist LOs in future data management planning.

The NOAA Data Management Plan Repository – As NOAA data management plans mature the repository will be a source of examples of proven plans, contact points, and other information for developing new plans.

The Data Management Integration Team (DMIT) - The DMIT is a NOAA-wide technical group established by the EDMC. DMIT can either answer specific questions or provide a forum for discussion, testing and prototyping. To be added to the DMIT mailing list, contact noaa.tpio.dma@noaa.gov.

The NOAA Data Management Architect – The Data Management Architect in the NOAA Technology, Planning and Integration for Observations (TPIO) program can provide detailed information, advice, and guidance on developing data management plans. Contact the DMA office at noaa.tpio.dma@noaa.gov.

EDMC Line Office/Staff Office Representatives – Your representative can provide specific guidance on implementing this directive in your LO.

9. Implementation of this Procedural Directive

Implementation of this directive will be the responsibility of data production project managers and their supervisors and management - overseen and coordinated by the NOAA Data Management Architect and the EDMC.

Implementation will occur in the following phases:

Phase	Description	Outputs	Outcomes
Phase 1 - Establish and socialize the directive	Vet and refine the guidance Establish and develop SOPS for the DMP Repository and the Wiki. Establish measurements of success for DMP Plans	DMP Repository established and standard operating plans (Sops) in place DMP Wiki initiated with metrics and SOPs in place NOAA-wide DMP in Repository (if appropriate)	Data Management Plan Position description and initial plans in place NOAA's Data Management process socialized and initiated
Phase 2. Collect or develop plans for major observing systems and new initiatives	Find existing DM Plans across NOAA and store copy in Repository. Identify observing systems lacking plans. Select new initiatives for this phase, and develop the corresponding DM plans. Establish process for reviewing DM Plans as described in Section 5.	DMPs for major Observing Systems LO-wide DMPs in Repository (if appropriate)	Plans for major observing systems and new initiatives in place

Phase 3. Fully mature NOAA DMP process	Develop DMPs for all data production projects Performance plan elements in place for all staff involved in environmental data management	SOPs for an ongoing sustainable NOAA DMP process Data Management Performance plan elements	DMPs for all NOAA data production projects' Data in Repository Fully sustainable data management planning process

10. Roles and Responsibilities

Following are roles and responsibilities for key stakeholders in NOAA's Data Management planning process:

EDMC – Developing and refining this directive.

EDMC Line Office/Staff Office Representatives – Determining how this directive will be implemented in their LO and whether or not the LO will develop a generic DMP to serve as the basis for other plans within the LO. LOs are encouraged to factor out common considerations into a parent plan, with additions or exceptions enumerated in subsidiary plans (see discussion of hierarchical plans in Section 4).

Data Management Architect – Supporting the NOAA-wide implementation of this directive. The NOAA Data Management Architect works closely with the NOAA Enterprise Architect.

Data Production Project Managers– Developing, maintaining, and following data management plans in accordance with this directive.

Supervisors and Management of Data Production Project Managers – Reviewing and approving data management plans for data production projects under their purview. Also, determining whether to develop program-wide data management plans to serve as a basis for project-level plans within the program.

Appendix A. Data Management Plan Template

This Template is provided as an example to assist DM Plan developers. It may be downloaded from the DMP Repository. The Repository also includes other plans and templates as examples.

Note: Some of this information may be used to populate fields in a formal metadata record about the data collection. These are indicated for illustrative purposes only with [→ field name].

1. General Description of Data to be Managed

1.1. Name of the Dataset or data collection project [→ gmd:title].

1.2. Keywords that could be used to characterize the data, and vocabulary from which those keywords were obtained (e.g., GCMD, CF Conventions, etc.) [→ gmd:MD_Keywords]

1.3. Summary description of the data to be generated [→ gmd:abstract].

1.4. Anticipated temporal coverage of the data [→ gmd:EX_TemporalExtent].

1.5. Anticipated geographic coverage of the data [→ gmd:EX_Extent]

1.6. What data types will you be creating or capturing? (e.g., digital numeric data, photographs, video, acoustic records, database tables, spreadsheets, paper records, physical samples, etc.)

1.7. How will you capture or create the data? (e.g., satellite, airplane, unmanned aerial system, radar, weather station, moored buoy, research vessel, autonomous underwater vehicle, animal tagging, manual surveys, enforcement activities, etc.)

1.8. Where will this plan be stored electronically besides in the NOAA DMP Repository?

1.9. What volume of data is anticipated to be collected in the Project Time Frame?

1.10. Will the data contain Personally Identifiable Information or any information whose distribution may be restricted by law or national security?

2. **Points of Contact** (Give name, title, location, e-mail address, phone number and mailing address, as appropriate.) [→ gmd:CI_ResponsibleParty]

2.1. Who can, or could, represent this data collection project on NOAA's Data Management Integration Team (DMIT)? Current members of DMIT are listed at https://geo-ide.noaa.gov/wiki/index.php?title=DMIT_Membership.

2.2. Who is the overall point of contact for the data collection?

2.3. Who is responsible for verifying the quality of the data?

2.4. Who is responsible for answering questions about the data collection?

2.5. Who is responsible for data documentation and metadata activities?

2.6. Who is responsible for the data storage and data disaster recovery activities?

2.7. Who is responsible for ensuring adherence to this data management plan, including ensuring that appropriate resources are available to implement the data management plan?

3. Data Stewardship

3.1. What quality control procedures will be employed?

3.2. What is the overall lifecycle of the data from collection or acquisition to making it available to customer?

4. Data Documentation

4.1. Which metadata repository will be used to document this data collection?

4.2. In addition to discovery-level metadata, what additional metadata or other documentation is necessary to fully describe the data and ensure its long-term usefulness? How will that metadata be collected and updated?" Is there a requirement to document this data collection in other metadata repositories?

4.3. What standards will be used to represent data and metadata elements in this data collection. *Note:* The EDMC Data Documentation Procedural Directive calls for the use of ISO 19115 and related standards for data documentation.

5. Data Sharing

5.1. Will the data be made available to the public? If so, what is the expected date of first availability? Is this a one-time data collection, or an ongoing series of measurements? Will there be a Principal Investigator hold or other delay between data collection and publication, and if so for how long? [Note: the NOAA Data Sharing Policy for Grants and Cooperative Agreements, which is currently under development, provides useful guidance for sharing data in a timely manner.]

5.2. If the data are not to be made available to the public, explain why and under what authority distribution may be restricted. NOAA policy states that Environmental data will be visible, accessible and independently understandable to users, except where limited by law, regulation, policy (such as those applicable to personally identifiable information or protected critical infrastructure information or proprietary trade information) or by security requirements.[2]

5.3. Will users be subject to any access conditions or restrictions, such as submission of non-disclosure statements, special authorization, or acceptance of a licensing agreement?

5.4. What data access protocols will be used to enable data sharing? The use of open-standard, interoperable, non-proprietary web services is recommended (for example, OPeNDAP, or Open Geospatial Consortium (OGC) web services).

5.5. In what catalogs will these services or data be made registered to enable discovery by users and other Catalogs?

---

[2] NOAA Administrative Order 212-15, "Management of Environmental Data and Information" (2010), https://www.nosc.noaa.gov/EDMC/internal/nao_212-15.php.

6. Initial Data Storage and Protection

6.1. Where and how will the data be stored initially (i.e., prior to being sent to a long-term archive facility)?

6.2. How will the data be protected from accidental or malicious modification or deletion? Discuss data back-up, disaster recovery/contingency planning, and off-site storage relevant to the data collection.

6.3. If there will be limitations to data access, how will these data be protected from unauthorized access? How will access permissions be managed? What process is to be followed in the event of unauthorized access?

7. Long-Term Archiving and Preservation

Note: NOAA's "What-to-archive" policy[3] describes how to identify, appraise and decide what scientific records are to be preserved in a NOAA archive.

7.1. In what NOAA Data Center (NODC, NCDC, NGDC) will the data be archived and preserved? Have you begun discussions with that Data Center regarding your intended submission?

7.2. If you have not identified a NOAA Data Center, what is your long-term strategy for maintaining, curating, and archiving the data?

7.3. How will the costs of long-term data archiving be provided and maintained?

7.4. What transformations or procedures will be necessary to prepare data for preservation or sharing? (e.g., quality control, format conversion, anonymization of personally-identifiable information, etc.). What related information will be submitted to the archive to enable future use and understanding of the data [e.g., metadata, references, reports, research papers, algorithms, audio or video codecs, special character sets or fonts, etc.]) .

Identify the Record Schedule applicable to these data and provide the retention time for these data.

[3] NOAA Procedure for Scientific Records Appraisal and Archive Approval: Guide for Data Managers (2008), https://www.nosc.noaa.gov/EDMC/internal/NOAA_Procedure_document_final_12-16-1.pdf.

Appendix B. Glossary of Terms

Note: Some terms and phrases, such as *significant* as used in "significant data collections" and *comprehensive* as used in "comprehensive data management plans," are intentionally left open to interpretation. This is to allow project leaders and managers to appropriately scale plans for the data being produced. *Significant* is intended to take on the meaning as defined at http://en.wiktionary.org, "the extent to which something matters, important". Only project managers and their supervisors can determine the importance of a data set and asses the detail, extent, and comprehensiveness of the plan that is required.

Data production projects: Observing systems, projects, programs, surveys, collections, interviews, investigations, studies and research that produce environmental data.

Data production project managers: Principal investigators, project leaders, program managers, and observation system managers who are primarily responsible for data production projects.

Data management lifecycle (data lifecycle), includes:

- Determining what environmental data are required to be preserved for the long term and how preservation will be accomplished
- Developing and maintaining metadata throughout the environmental data lifecycle that comply with standards
- Obtaining user requirements and feedback
- Developing and following data management plans that are coordinated with the appropriate NOAA archive for all observing and data management systems
- Conducting scientific data stewardship to address data content, access, and user understanding
- Providing for delivery to the archive and secure storage
- Providing for data access and dissemination
- Enabling integration and/or interoperability with other information and products

[From NAO 212-15]

Appendix C. Frequently Asked Questions (FAQs)

This appendix is to address questions folks may have during the review of the procedural directive.

Q. Does completion of the "full lifecycle" mean that data collection has ended?

A. NO - lifecycle continues with long-term preservation and final disposition after data collection has ended.

Q. If a data collection is expected to continue for several more years. Does the directive indicate we need to write a plan for this mature collection and archive process?

A. Yes

Q. Why is directive not more directive? It doesn't explicitly state what I must do to comply.

A. The directive states that if you are a manager of a project or system that produces data you must write and follow a data management plan and that the plan must be stored in the data management plan repository. It leaves open to interpretation the scale of the plan, which should reflect the data production project managers and supervisors in chain of command's best judgment.

Q. Appendix A can be treated minimally or comprehensively. Why is this?

A. Appendix A provides questions to be considered in all data management plans. The directive does not state the all questions must be answered - only that they should be considered.

Q. Can LO's or systems develop their own DMP template?

A. Yes. The questions in the directive's template are to provide a list of items that should be addressed.

Q. Why are there loopholes in who develops a comprehensive plan?

A. In the directive, terms such as *comprehensive*, *significant* etc., are purposely left open to interpretation so that data production project managers and their chain of command may scale the plan to their particular needs.