NOAA Data Documentation

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.DD.php

NOAA Environmental Data Management Committee

October 2011

VERSION HISTORY

Version	Implemented By	Revision Date	Approved By	Approval Date	Reason
1.0	Ted Habermann Lewis McCulloch	Initial Draft	EDMC	10-28-2011	Approved by EDMC

Table of Contents

Purpose

NOAA Administrative Order (NAO) 212-15, Management of Environmental Data and Information, as revised in November 2010, describes the NOAA data life cycle and requires that: "Environmental data will be visible, accessible and independently understandable to users…" It also lists "Developing and maintaining metadata throughout the environmental data life cycle that comply with standards" as the second element of this life cycle. This Procedural Directive provides background information and outlines responsibilities for documenting NOAA's environmental data and information using International Standards.

Scope

This Procedural Directive applies to metadata and documentation for all existing and new NOAA environmental data, information and services[1] and to the personnel and organizations that collect and manage them, unless exempted by statutory or regulatory authority.

Specifically:

- All NOAA data collections, and products derived from these data shall be documented.
- Services that provide NOAA data and products shall be documented.
- Data collections funded by NOAA, and products derived from these collections that are funded by NOAA shall be documented.
- Data collections currently in progress and products derived from these data shall be documented.
- All active and planned data collection programs shall be documented.

This Procedural Directive considers metadata, other documentation and links between them (see Appendix A). All three will likely be needed to address all of NOAA's documentation needs.

Standards

This Procedural Directive establishes a metadata content standard (International Standards Organization [ISO] 19115 Parts 1 and 2) and a recommended representation standard (ISO 19139) for documenting NOAA's environmental data and information.

Roles and Responsibilities

Cross-NOAA Responsibilities

- *Encourage* and *support* participation in the ISO and Open Geospatial Consortium (OGC) standards development and evolution processes

[1] Underlined words and terms are defined in Appendix E.

- *Develop* and *implement* common metadata management tools including mechanisms for evaluating the completeness and quality of data documentation
    - Utilize rubrics to establish the baseline and monitor progress.
    - Engage users in providing feedback on data documentation efforts and opportunities.

- *Promote* and *highlight* good examples of documentation and the individuals involved in their creation.

- *Support training* specifically targeted at improving NOAA's data documentation.

- *Initiate teams* to work on "special documentation problems" that cross Line and Staff Offices (See Appendix D for suggested topics).

See Appendix B for details on Cross-NOAA Responsibilities.

Environmental Data Management Committee (EDMC) Responsibilities

- Review this Procedural Directive twice a year to evaluate effectiveness and monitor progress.

- Work with the CIO Council and the NOAA Observing Systems Council to implement and monitor progress on the Cross-NOAA responsibilities listed above.

- *Encourage* and *support* partnerships with external organizations in the process of migration of metadata from FGDC to ISO Standards.

Line and Staff Office and Program Responsibilities

Plans and resources required for implementation of the improvements envisioned in this Directive will vary greatly with the diversity of existing situations and needs. The real work required for improving documentation of NOAA data, products, and services will be carried out in the Line and Staff Offices and Programs and they are responsible for planning and resourcing those efforts following these steps.

Step 1: Identify documentation expertise

- Establish Data Stewardship Teams to facilitate documentation creation and improvement for appropriate organizational units or around programmatic needs.

- Data Stewardship Teams should include the following expertise/skills:
    - Data Collectors/Providers
    - Data Users
    - Data Stewards
    - Standards Experts

Step 2: Assess the current state of documentation

- Identify existing sources of documentation (Data Collectors/Providers)

- Classify existing documentation into following categories (Data Stewards and Standards Experts):
    - Metadata for Discovery
    - Metadata for Use
    - Metadata and Documentation for Understanding
    - Documentation of Collections
    - Documentation of Datasets
    - Documentation of Services

- Identify high-priority targets for improvement (All members of Data Stewardship Team)

- Highlight best practices, successful teams and individuals (Line and Staff Office Management)

Step 3: Create and Improve Metadata (Data Collectors/Providers, Data Stewards and Standards Experts)

- Translate/transform existing metadata into the recommended representation (ISO 19139)

- Create metadata for undocumented data and information

- Use spiral approach for improving metadata

Step 4: Publish Metadata (Data Stewards)

- Publish metadata record in new or existing Web-Accessible Folders or using a standard catalog service. This will make it possible to connect metadata records to various discovery portals using standard services

Step 5: Preserve Documentation (Data Stewards)

- Work with NOAA Archives to ensure that documentation and metadata will be preserved for the long-term

See Appendix C for details on Line and Staff Office and Program Responsibilities.

Measuring Progress

Effectiveness of this Procedural Directive will be measured by the following:

- An increase in the amount of NOAA environmental data and information that is well documented and discoverable via national and international discovery portals.

- Improvement in the quality of Line and Staff Office and Program data documentation processes.

Appendix A. Metadata Background

Data collected and produced by NOAA scientists and managers form the basis for characterizing and understanding important aspects of the global environment. These irreproducible observations form the foundation for future generations to understand the current state of this environment. NOAA's core data collections, and the results or products derived from them, need to be credible and authoritative now and in the future. High quality documentation must accompany these data and analyses and be readily accessible and understandable, so the data will be trusted and easily integrated into the international data fabric. If detailed documentation that meets well-defined standards is not available, the data will not be accepted or used.

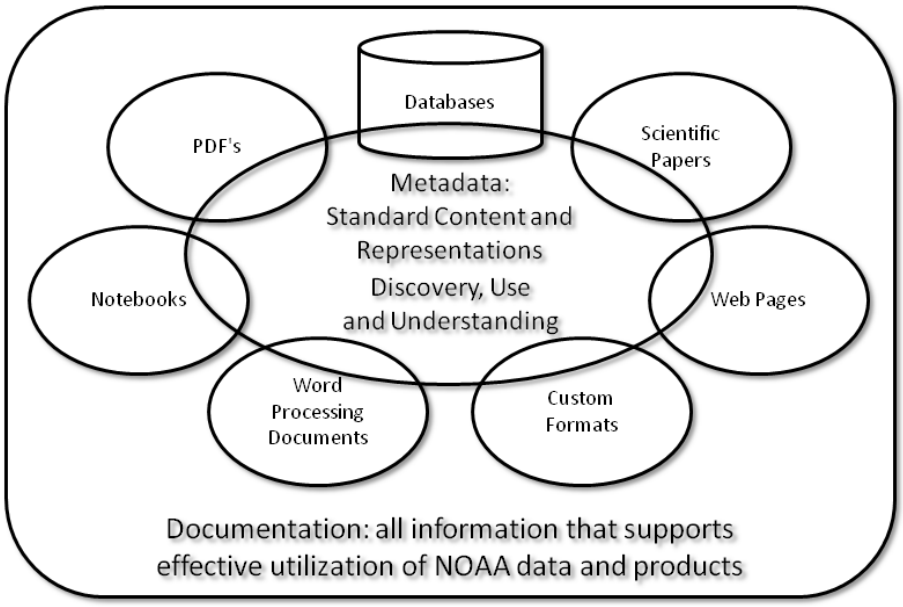
During the last several years, a series of international (ISO) metadata standards have emerged to replace those that were developed by the U.S. Federal Geographic Data Commission (FGDC) and included in National Spatial Data Infrastructures around the world. These new standards form the foundation for current and future documentation efforts. The adoption of these standards in the United States will involve a significant transition in the way the U.S. environmental community documents data and in the ways humans and applications use metadata. The impacts will extend significantly beyond the data discovery role that has motivated metadata developments in the United States over the last several decades to include detailed descriptions of lineage (provenance), processing and data quality. The focus will be on ensuring that observations are independently understandable by many diverse users.

This transition creates exciting opportunities and challenges for all NOAA Programs that collect, document, analyze and preserve environmental observations. This document outlines a collaborative effort to build capabilities and expertise across NOAA and help the entire organization effectively address this transition. It provides background to support shared understanding for documentation and metadata discussions (Appendix A) and outlines expectations and processes for creating documentation that ensure the future value of NOAA data collections and analytical products.

Documentation vs. Metadata

Many NOAA datasets and products are documented using approaches and tools developed by data collectors to support their analysis and understanding. This documentation exists in notebooks, scientific papers, web pages, user guides, word processing documents, spreadsheets, data dictionaries, PDF's, databases, custom binary and ASCII formats, and almost any other conceivable form, each with associated storage and preservation strategies. This custom, often unstructured, approach may work well for independent investigators or in the confines of a particular laboratory or community, but it makes it difficult for users outside of these small groups to discover, use, and understand the data without consulting with its creators.

Metadata, in contrast to documentation, helps address discovery, use, and understanding by providing well-defined content in structured representations. This makes it possible for users to access and quickly understand many aspects of datasets that they have not collected. It also makes it possible to integrate metadata into discovery and analysis tools, and to provide consistent references from the metadata to external documentation.



Metadata standards provide standard element names and associated structures that can describe a wide variety of digital resources. The definitions and domain values are intended to be sufficiently generic to satisfy the metadata needs of various disciplines. These standards also include references to external documentation and well-defined mechanisms for adding structured information to address specific community needs.

This Procedural Directive considers all three of these components: structured metadata, references to external documentation, and structured extensions to the metadata. All three will likely be needed to address all of NOAA's documentation needs.

Metadata Standards/Dialects

The purpose of metadata is to ensure that users can discover, use, and understand data and information in the present and the future. Achieving this goal across a diverse community of data producers and users is difficult and data comparisons are practically impossible if documentation for each dataset is written and organized in different ways. Many communities address this problem by adopting and adapting standards and developing conventions that enable transparent access to comprehensible, structured information (metadata). Two types of standards are important. Content standards describe what elements and structures users can expect to find in metadata and the meaning of those elements. Representation standards control how that content is arranged and formatted, so they can be read and understood by users and machines. This Directive describes a specific metadata content standard and a general representation approach for NOAA documentation.

A variety of detailed content standards exist for environmental metadata. The most comprehensive and broadly applicable are the ISO Standard for Metadata for Geographic Information (19115)

and related standards (see Appendix F). These standards are being adopted throughout the global environmental community and were officially endorsed as US Standards by the Federal Geographic Data Committee (FGDC) during September, 2010. The adoption of the ISO Standards by the U.S. Federal Government, and by many national and international NOAA partners, coupled with their well-defined governance and breadth, make them the clear choice as the core standards for current and future NOAA metadata efforts.

Extensible Markup Language (XML) has become the universal format for organizing and representing metadata content. NOAA metadata must be available in well-formed XML documents that are valid with respect to a published and openly available XML schema in order to be integrated into the international data arena. The ISO 19139 standard provides an open and available XML representation for the content included in ISO 19115 and other related content standards. It is the preferred XML representation for NOAA metadata. If a different schema is used for some metadata, an XSL style sheet must be provided that translates between that schema and 19139. If elements exist in NOAA metadata that are outside of the ISO standards, they must be described using the standard mechanism for extending the ISO Standards.

Metadata and Documentation Types

This directive applies to documentation for all NOAA observations and products, regardless of the purpose of the documentation or the granularity of the data. This section describes documentation that serves a variety of purposes and exists at many granularities. The classifications described here are very general and the boundaries between them are very fuzzy. They should be viewed as illustrative examples rather than hard and fast boundaries.

Metadata for Discovery

Discovery metadata allows users to search and find NOAA data holdings using text, keyword, temporal, and spatial queries, and to locate a contact person for the data they discover. These metadata address the following questions:

- Does a dataset on a specific topic exist ('what')?
- For a specific place ('where')?
- For a specific date or period ('when')?
- Where can I obtain the data and whom can I ask about them ('who')?
- Why were the data collected ('why')?

Popular dialects traditionally used for this type of metadata include: FGDC Content Standard for Digital Geospatial Data (CSDGM), NASA Directory Interchange Format (DIF), and Unidata NetCDF Attribute Conventions for Data Discovery. All of the discovery elements have straightforward mappings to the ISO Standards. Metadata in these dialects are shared with and supported by major discovery portals (e.g., Geospatial One-Stop, data.gov, Global Earth Observing System of Systems (GEOSS), Global Change Master Directory (GCMD), etc.).

Metadata for Use

Use metadata are those that allow tools and services to identify appropriate data and perform fundamental display, analysis, and comparison operations. These metadata might address the following questions (as well as others):

- Are these data a recognized type?
- What format are the data in? How is metadata stored in that format?
- What axes/units should be used to create a standard display for this data type?
- What parameters are included in the dataset and what are their units and ranges?
- What projection, map scale, exchange format, compression type, and data formats are available?

Popular dialects for this type of metadata include CF-Conventions for netCDF, OPeNDAP Data Descriptions in XML (ddx), netCDF Markup Language (NcML), and Climate Science Modeling Language (CSML). Most use metadata is used in native formats, like netCDF, and much of the use content can be translated to ISO if necessary for sharing.

Metadata and Documentation for Understanding

Discovery and use can generally be facilitated with complete well-structured metadata. Understanding, on the other hand, can require diverse information that can be difficult to fit into a generic structure. It is essential that derived and analytical products, tools and services be documented with enough metadata and documentation to support independent understanding by anticipated and unanticipated users. The goal of metadata for understanding is to provide a structured summary of this diverse information, where possible, with references to more complete documentation where available or necessary. Together this enables users to analyze and compare NOAA data in web-based and desktop tools. Metadata and documentation for understanding:

- provides information about the quality and accuracy of resources, the sources used to create integrated datasets and models, instrumentation and collection methodologies, processing history, and archival procedures required for users to understand the data and trust the decisions made using them,
- supplies adequate descriptions of the data parameters allowing potential users to assess their suitability for other purposes,
- allows users to determine whether they want to access the raw or derived data asset, and to understand the potential limitations of data usage,
- supports the preparation, publication, and assessment of scientific reports,
- ensures that data shared inside and outside NOAA are readily and independently understandable by identifying and providing access to related documentation.

Documentation of Series (Collections)

Many NOAA data sets are divided into multiple files for collection, management and, or distribution convenience. These files could contain all of the observations for a single satellite orbit or all of the observations collected during a single deployment of an in-situ ocean sensor. These files often share contact and distribution information, collection and processing history, and many other metadata elements. It is also common for documentation, e.g. processing descriptions and scientific papers, to refer to collections of files rather than single files.

These shared elements have traditionally been referred to as collection, directory, or dataset metadata. In ISO 19115 and 19115-2 they are referred to as series metadata. A single series metadata record can describe a dataset that includes many thousands of files. They play an important role in the data discovery process and are generally shared across discovery portals (GeoData.gov, data.gov, GCMD …). It makes sense to manage these shared elements in a way that recognizes their quasi-static and re-useable nature.

Documentation of Datasets (Granules)

While series metadata can include many important elements, the files or granules that make up the series also have important individual characteristics that need to be described in metadata or documentation for specific files. This can include spatial and temporal extents, specific quality assessments, specific processing versions, and other elements. This metadata has traditionally been termed granule metadata. In the ISO standard it is referred to as dataset metadata. It is generally served in specialized systems that support detailed queries for granules with specific characteristics. The CLASS discovery system is an example of a NOAA system that provides collection and granule searches.

Documentation of Data Services

The last several years have seen a significant increase in the utilization of services that provide data in response to standard requests. Many of these services are Open Geospatial Consortium web services (e.g. Web Map Service, Web Coverage Service, Web Feature Service, Sensor Observation Service, …) or OPeNDAP data services. The ISO 19119 Standard extends ISO 19115 to include metadata for these services.

Current Documentation States and Workflows

As discussed above, current documentation for NOAA data and products exists in many analog and digital forms and locations. This documentation can be difficult to access and understand, particularly for people who did not create it. All the same, documentation in this state needs to be identified and organized. The heterogeneity of this documentation will make this process difficult and time consuming. The needs are particularly urgent for documentation created by scientists who are close to retirement. As they leave, they take with them information that could be critical to understanding the data they collected while at NOAA. Losing this information significantly decreases the future value of their work.

It is important to note that there is significant overlap between the content of much of this documentation and the standard content models. Line and Staff Offices and Programs can maximize the benefit of this unstructured documentation for NOAA data and products by 1) identifying standard content throughout NOAA and making that content available through translation and/or reformatting using international standards and 2) identifying non-standard content and referencing it, when possible, from standard metadata.

There are large collections of NOAA metadata that conform to the FGDC Content Standard for Digital Geographic Metadata (CSDGM) with and without remote sensing or biological extensions. These metadata need to be translated to ISO and expanded to include information in ISO that is not in CSDGM. The digital and well-structured nature of these metadata make this process significantly

more straightforward than dealing with the unstructured documentation described above. There are many other organizations with considerable collections of FGDC metadata. NOAA should partner with them in the migration and improvement process.

Finally, NOAA observing systems create many real-time data streams that have special documentation and metadata needs. These can include on-going summarization, quality assessments and real-time user notification. It may be possible to create this real-time metadata in standard formats or it may be necessary to reference services that can provide it.

Appendix B. Cross-NOAA Responsibility Details

It will be critical to provide over-arching support applied consistently across the organization and to nurture and support the documentation and metadata creation processes as they spread across the organization.

Supporting Standards Development and Evolution

The ISO Standards that NOAA is implementing evolve and improve over time under the guidance of the International Standards Organization. These improvements are driven by new requirements that emerge from the global environmental data community that is implementing the standards. A similar process takes place in the Open Geospatial Consortium (OGC), an important ISO partner (NOAA is a principal member of the OGC). Ensuring that requirements of NOAA and our partners are considered in these standards development and evolution processes is critical to effective utilization of these standards in our community.

Positive returns from investments in geospatial standards have been clearly demonstrated. These standards increase the value and utility of our data significantly and leverage internal development resources across the international community. In order to reap these benefits, NOAA must encourage and support participation in ISO and OGC by experts in all aspects of data and product development and documentation throughout the organization.

As ISO Standards are adopted across NOAA, areas that need adaptation and extension will almost certainly be identified. NOAA must support processes that facilitate sharing extension needs and solutions across the Line and Staff Offices and Programs. Similar needs should be identified and addressed with similar solutions in order to encourage interoperability.

NOAA-Wide Tool Implementations

Achieving the goals of this directive across NOAA will certainly involve the adoption, adaption and development of metadata translation, creation, and management tools. Many of the requirements addressed by these tools will be shared across multiple Line and Staff Offices and Programs. The team described below (see Metadata Creation and Management Tools section in Appendix D) will be responsible for identifying tools that serve those needs. In some cases it will be most cost-effective to address those needs at the NOAA-wide level. Examples might include enterprise licensing for widely used COTS tools, shared discovery portals, and support for web-accessible folders and metadata quality and improvement tools.

Evaluating Documentation

NOAA will develop and implement mechanisms for evaluating the completeness and quality of existing documentation and the tools for applying those mechanisms consistently across the organization. The spiral development approach outlined above is amenable to evaluations using rubrics that clearly describe goals and measure progress towards those goals. A rubric that can quantitatively measure documentation completeness across NOAA will be developed based on the spiral development approach outlined above. Information about the rubric and additional metrics will be available in the GEO-IDE Wiki.

The Open Archival Information System (OAIS) Reference Model describes a high bar for monitoring the quality of metadata and documentation. Representatives of the communities that use the data are identified as designated community representatives and they are asked to evaluate whether the metadata and documentation provided with the data are sufficient to make the data independently understandable to their communities. All NOAA Data Centers have adopted the OAIS Reference Model as a guide to their operations. This approach should be extended to all NOAA groups that distribute data and they should be required to include identification of designated community representatives, regular metadata and documentation evaluations, and mechanisms for feedback from their user communities.

Finally, the ISO Standards NOAA is adopting include a mechanism for recording limitations of data, products, metadata and documentation identified by users. This mechanism should be implemented across all groups in NOAA that provide data and products to users. This will extend the designated community representative concept across a broad sample of users and user communities that may identify problems missed by the initial designated community representatives.

Identifying, Sharing and Training Good Examples, Experiences, and Practices

One of the primary goals of evaluating published documentation throughout NOAA is the identification of good examples at every step of the improvement process. These examples, and the practices used to create them, need to be broadly shared in order to foster improved documentation and, equally important, processes used to create it. Further, good examples need to be identified in the NOAA and Department of Commerce awards process and the award winners need opportunities to share their experiences to increase capabilities throughout the organization and our national and international partners.

In addition, NOAA should support a variety of training efforts including shared web resources (the GEO-IDE Wiki), on-line training (Commerce Learning Center and others), and face-to-face workshops designed to foster regional partnerships and communities of practice across NOAA. This effort should include specifics of the standards being used to document NOAA data and observations as well as the underlying technologies and tools that support managing and sharing that information (e.g. XML).

Appendix C. Line and Staff Office and Program Responsibility Details

Successful migration of NOAA documentation towards high-quality metadata that are compliant with international standards will require significant efforts across all NOAA Line and Staff Offices and Programs. The goal of the Line and Staff Office and Program efforts will be the creation of high-quality, web-accessible documentation that includes ISO metadata for all data and products. That metadata will be made available through appropriate discovery portals and will include references to other documentation whenever possible. Achieving this goal will require a significant emphasis on sharing data and information with broadening user communities. Individuals involved in creating and improving documentation should work with their Line and Staff Offices and/or Programs to identify specific steps towards the goals outlined in this document.

Planning and Resourcing

Plans and resources required for implementation of the improvements envisioned in this Directive will vary greatly with the diversity of existing situations and needs. The real work required for improving documentation of NOAA data, products, and services will be carried out in the Line and staff Offices and Programs and they are responsible for planning and resourcing those efforts. The common goals and standards will give rise to many opportunities for collaborating and sharing resources across NOAA. The communication mechanisms described below as Cross-NOAA responsibilities should support that collaboration and sharing.

Step 1. Data Stewardship Teams

Creating and maintaining high quality metadata for NOAA data and products is best done by ongoing collaboration within teams that include active representation of several groups:

Data Collectors/Providers are the group responsible for collecting and processing observations. They design instruments, observing systems and processing systems and operate all three. Some of these people are scientists working on research projects on the cutting edge of environmental science. Some are parts of operational teams trying to keep legacy equipment and systems running just a little bit longer! They are the foundation of the documentation process, understanding the details that affect the observations and creating the original source materials (either physical or mental) for the documentation system.

Data Stewards are the people that take long-term responsibility for sharing observations with users and for ensuring that the users can understand the data they receive. In many cases, they represent the Data Collectors/Providers to the users. They continue in this role after the Data Collector/Provider has moved on to other problems and, sometimes, to other careers or retirement. Preserving data and understanding is a difficult process of communicating with the future. Data Stewards orchestrate that communication.

Data Users can be defined in countless ways, but all share a need to understand data sufficiently well to apply the data with confidence to their own problems. Users are where the documentation rubber meets the road, and user effectiveness in finding, using, and understanding data will be the bottom-line success metric for this entire endeavor. Called the Designated Community in the Open

Archival Information System Reference Model (OAIS-RM), users bring real-world experience in how data actually are employed, and they can offer deep insight into documentation requirements.

None of the above groups have the time (or, typically, the interest) to learn the arcane details of international standards for sharing or documenting data. *Standards Experts* are the people who are familiar with the details of the standards and, more importantly, the practices being used throughout the global environmental community to apply those standards in real-world situations. Many times these people are involved in the important work of bringing NOAA requirements into the standards development process and sharing feedback about how other groups around the world are addressing similar requirements. They understand the content standards and the details of standard representations. They are also familiar with community documentation practices being used by NOAA and NOAA partners and the connections between those practices and the standards.

Increasingly all of these groups understand that high-quality documentation in standard forms increase the value of data and form a foundation for the trust that fosters use of data by people who did not collect it. Even understanding the need for documentation and standards, most NOAA Offices do not have the resources to address the need alone. NOAA Line and Staff Offices and Programs are responsible for identifying members of these groups at appropriate levels within their ranks and forming them into effective data stewardship teams that will work together ,and with similar groups in other parts of NOAA, to improve documentation of datasets they are familiar with. The Line and Staff Offices and Programs are also responsible for connecting with or developing Standards Experts that can help Stewardship Teams meet their goals.

Step 2. Documentation Assessment and Gap Analysis

The first step in the documentation improvement process for each Line and Staff Office and Program is to assess the current state of documentation for their data and products relative to the goals of this directive. Existing documentation resources should be identified and classified using the types outlined above (discovery, use, understanding, series, granule, …) and the format (custom or standard, analog or digital). The analysis should also include estimates of the number of datasets and products that currently do not have documentation.

The resource identification and classification phase of this task is primarily the responsibility of the Data Collectors/Providers, although others may be involved for historic datasets. Once existing documentation is collected and accessible, the Data Stewards and Standards Experts come into the picture to characterize the existing documentation and the requirements that it serves. They bring experience with standard approaches used to address general documentation requirements and apply that experience to the existing documentation.

When possible, users should be included in the assessment process in order to provide input on high priority targets for focused metadata improvement efforts. They can also identify important datasets that are missing critical documentation..

Outstanding documentation and metadata examples will play a critical role throughout this entire process. The assessment and gap analysis phase is the first opportunity for Line and Staff Offices to identify good examples to share with others involved in this process throughout NOAA. The

analysis should include identification of well documented datasets and the people responsible for creating them. These are the leaders that have already developed practices that support the goals of this directive. They have a critical role to play in helping others to move forward. At the same time, Line and Staff Offices and Programs need to identify critical gaps in staff knowledge and skills that need to be addressed through training and, more importantly, through sharing of the knowledge, skills and practices that created the examples identified as being at the top of the gaps.

In many cases this assessment will result in the identification of existing collections of metadata that comply with the CSDGM core and extensions. The initial focus for this large collection of metadata will be on translation to ISO without compromising content. See the Dialect Translation/Presentation section below for specifics.

Step 3. Creating and Improving Metadata

The Assessment phase will increase awareness of existing documentation and the requirements that it serves (discovery, use, or understanding) as well as identifying undocumented data and products. Metadata can be created from the existing documentation by identifying standard content and transforming it into a standard structure. The details of this transformation will depend on the specifics of each situation. If the existing documentation is in a database, an output report that produces structured metadata may be the most reasonable approach. If the existing documentation is in XML, a stylesheet that transforms it into a standard structure may be necessary. If the current documentation is in a custom format and there are tools to read it, those tools may be extended to output a standard metadata format.

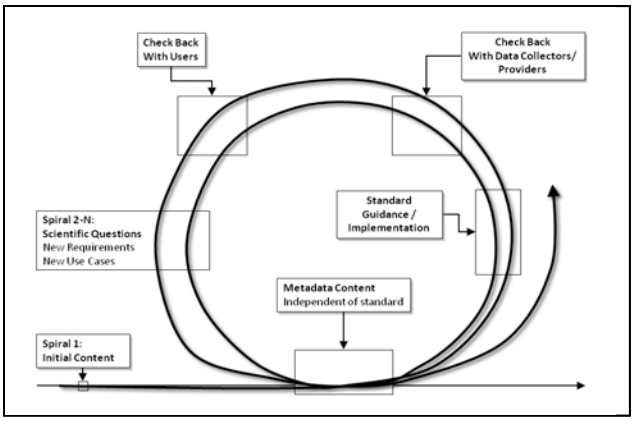
Metadata will need to be created for the undocumented products. Again, the details of this process will vary with the specifics of the situation. Approaches similar to those outlined above and lessons learned during their application can play a significant role in simplifying the metadata creation process.

The metadata creation process will require Data Stewards and Standards Experts working together to identify standard content in the documentation and to map that content into the appropriate places in the standards. As this stage progresses, check-backs with the Data Collectors/Providers should occur in order to make sure that the translation is correct and as efficient as possible.

The focus for the initial metadata creation efforts should be on discovery metadata that can be shared with national and international portals to help new users connect with NOAA data and products that they have not previously been aware of. This should be accomplished without forgetting the greater goal of robust metadata for use and understanding.

Adding to the breadth of this foundation built above will be an on-going process as new data and products are developed or undocumented datasets uncovered. There will also be many opportunities for building on this foundation with use and understanding metadata as well as unstructured documentation that is referenced.

The software engineering community has recently been very successful at envisioning and implementing the software development process as a series of spirals, each of which addresses a small



set of user requirements. Each spiral involves several phases: requirements collection and prioritization, implementation, testing, and, most importantly, on-going interaction with users. Each spiral builds on previous work and requirements are addressed through a series of on-going iterations, each of which results in a more capable system. This Figure shows how this spiral concept can be applied to the metadata improvement process.

Like a multi-spiral software development process, the creation of high-quality, complete documentation is an on-going interaction among several groups. It is an end-to-end process that encompasses the complete data life cycle. Cross-NOAA responsibilities outlined below include shaping and supporting this spiral development effort and sharing success stories and examples. Those responsibilities also include identifying special documentation problems and identifying and supporting teams to develop cross-NOAA solutions to those problems.

Step 4. Publishing Metadata

The metadata creation phase will result in a significant increase in the number of NOAA data and products that are documented for discovery through national and international portals (e.g. Data.gov, GEOSS Portal, GCMD, …). In order to complete the loop, the metadata must be published to those portals using various protocols and services that may vary from portal to portal. Typically, this process can be facilitated by storing the metadata records in directories that are accessible through the Web (Web-Accessible Folders). This also facilitates translation of the metadata into dialects that are appropriate for each portal in addition to systematic evaluation and quality-control of the metadata, which are described below as Cross-NOAA responsibilities.

Step 5. Preserving Documentation

The Assessment phase will also increase awareness of existing documentation in many forms and formats that falls outside of the metadata content standards. This documentation forms a critical part of the information required for long-term preservation of NOAA data and products so steps must be taken to ensure that it is preserved along with the data and products. This documentation will also be referenced from the metadata, so mechanisms for persistent identifiers and URL's will need to be developed and deployed.

Appendix D. Special Problems

This directive identifies the suite of ISO 191* Standards as the adoption target for NOAA documentation. Effective utilization of a family of standards as broad as the ISO 191* family involves adoption, adaption and development. There are a number of important documentation and metadata challenges across NOAA that will require research, experimentation, and guidance for adapting these standards to specific needs. Additionally, tools may need to be adopted and/or developed to support consistent implementation of this guidance. NOAA needs to initiate teams tasked with addressing these challenges by identifying and/or creating good examples. These examples can include content and/or tools.

Dialect Translation/Presentation

NOAA currently uses a variety of standard and custom documentation dialects. Translating between these dialects is critical to extending the utility and re-usability of documentation across multiple communities. In the XML arena, these translations are typically done using Extensible Stylesheet Language Transformations (XSLT). Many XSLs (the instructions for an XSLT) have already been developed both within NOAA and by NOAA partners and are being used operationally. These translations need to be shared, tested with various inputs and evolved as new needs emerge.

The translation between FGDC CSDGM and ISO is of special interest because of the large collection of FGDC-compliant metadata that exists in NOAA and around the world. NOAA metadata experts at the National Coastal Data Development Center (NCDDC) have taken the lead in developing and implementing this translation. The NCDDC tools work very well, but there are many details that inevitably emerge and complicate the process in specific cases. This translation will ultimately be used throughout NOAA and the global environmental data community. It needs to be tested and improved as new special cases are identified. A reverse translation (ISO to FGDC) that focuses on metadata for discovery has been developed at the National Geophysical Data Center (NGDC) and will also continue to be important in supporting legacy tools and portals.

Transformations between netCDF or OPeNDAP documentation dialects and ISO are also important for many NOAA data providers. Initial versions of these translations have developed at NGDC and will soon be incorporated in the Thematic Real-Time Environmental Distributed Data Services (THREDDS) Data Server. These translations will also need testing and evolution as new requirements emerge.

Metadata Creation and Management Tools

Metadata creation has historically involved manual editing of records by Data Collectors / Providers or Stewards using tools developed specifically for this task. There are a number of open-source and commercial desktop and web-based tools available for creating and editing ISO metadata (e.g. GeoNetwork, CatMDEdit, GeoPortal Toolkit). These need to be tested and adopted for use within particular communities in NOAA. Communication needs to be facilitated within groups that are using particular tools and adoptions that support best practices need to be shared. Groups that are developing those adoptions need to be aware of common requirements that can be addressed as code is developed.

Utilization of XML brings many standard tools into the metadata creation and management arena. These are sophisticated integrated tools that have significant capabilities (e.g. Altova XMLSpy, Oxygen XML Editor). These tools are already being used in many NOAA Offices. Supporting these tools with enterprise licenses and facilitating sharing of expertise is critical to cost-effective utilization across NOAA.

Catalogs and Portals that support discovery are another important part of the metadata landscape. NOAA is currently taking advantage of several external catalogs and portals (Geospatial One-Stop, Data.gov, GEOSS, GCMD, USGS National Biological Information Infrastructure (NBII) Clearinghouse…) and is developing and maintaining the Climate Portal at the National Climatic Data Center (NCDC). NOAA needs to support the open source communities and companies that are developing these tools with input on existing and anticipated requirements.

Finally, NOAA should develop tools that support the development and use of the web accessible folders recommended for publishing Line and Staff Office metadata. These tools could include automated version control, checking for broken or unresponsive links in the metadata, translation of the metadata into multiple views and dialects, and automated validation and evaluation of the completeness of the metadata.

Reusable Documentation Components

Many metadata elements are repeated in related records from particular collections or Data Centers. The ISO Standards allow these reusable components to be managed in ways that centralize and simplify maintenance and reuse. The components can be connected to many records using the standard xlink protocol which allows links like those used in Web pages to be added to XML documents. Tools that support this approach must be developed, tested and deployed across NOAA.

Hierarchical Documentation and Metadata

Many NOAA datasets are naturally organized into hierarchical structures that are often reflected in directories and sub-directories that contain related files. Similar hierarchical structures are available for organizing documentation and metadata that are related to different aspects of a collection at several levels. Using these hierarchies supports intuitive metadata organizations that parallel the data organization, but it introduces complexity into many management and display tools. Nevertheless, these structures are commonplace, so users can take advantage of them intuitively. NOAA needs to develop guidance and examples of using these metadata structures effectively and tools that connect metadata to appropriate levels of data collections.

Granules and Collections

Many NOAA datasets are made up of collections of granules (files) that share many documentation elements and have others that are specific to each granule (see discussion above). These granules are written in a variety of formats, many of which support the inclusion of some documentation and metadata. That documentation can facilitate interoperability if it is written using well-known shared conventions like the Climate-Forecast and Data Discovery Conventions for netCDF. Preliminary mappings of those conventions to the ISO Standards exist and are being integrated into granule access tools. Methods for connecting the granules to more detailed collection metadata need to be developed, tested, and deployed in groups across NOAA that are using these formats.

Similar approaches need to be explored for other data formats that are used across multiple groups in NOAA.

Datasets and Services

Many NOAA Data Centers use various kinds of services to provide data and information to users. The ISO 19119 Standard provides a mechanism for describing such services. These consistent descriptions can facilitate discovery, use, and eventual chaining of these services. Connecting the metadata for a service to the metadata for the datasets it serves is straightforward for simple services that include a small number of datasets. Describing the connections in cases where a service includes many datasets is not as straightforward.

Resource Lineage and Data Quality

NOAA serves many datasets, products, and model results that are created by integrating multiple observations made using a variety of instruments and processing systems. Keeping track of the input data, the instrumentation, and the processing systems is critical for understanding changes in these products that have occurred or might occur in the future. Consistent quantitative determinations of data quality are also important.

The ISO Standards include the capability to record instrumentation, data quality, and processing histories (lineage). Connecting this capability to existing processing and quality control systems will provide a powerful foundation for understanding future variations in NOAA products and demonstrating that those variations reflect real changes in the environment rather than changes in the systems used to collect and process data.

Information Quality as defined by Office of Management and Budget (OMB) consists of utility, integrity and objectivity of the information and data made publicly available. The goal of the U.S. Information Quality Act (IQA) is to ensure and maximize the quality, objectivity, utility, and integrity of information disseminated to the public. NOAA has developed guidelines to meet the IQA. These goals and guidelines must be considered as we develop tools and guidance for documenting our data.

Appendix E. Definition of Terms

Data Collectors/Providers: The group responsible for collecting and processing observations. They design instruments, observing systems and processing systems and operate all three.

Data Stewards: The people that take long-term responsibility for sharing observations with users and for ensuring that the users can understand the data they receive

Data Users: The people that use data and share a need to understand the data sufficiently well to apply them with confidence to their own problems

Discovery Portals: Online mechanisms (e.g. Data.gov, GCMD, GEOSS Portal) to allow users to search and discover environmental data and information resources.

Documentation: All structured and unstructured information that can be used to discover, use, and understand a datasets, products, and services. Includes metadata, scientific papers (published and unpublished), reports, web pages.

Documentation of Collections: Information that describes a complete collection or series of data and that is shared by all items in that series.

Documentation of Datasets: Information that describes a single item or granule in a collection or series of data and that is specific to that item.

Documentation of Services: Information that describes a service, many times a web-service, that provides data. In many case these are standard services defined by the Open Geospatial Consortium or other standards body.

Environmental data and information: 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. Media, including voice recordings and photographs, may be included (NAO 212-15).

Metadata: The structured and standard subset of documentation that conforms to content and representation standards.

Metadata Content Standard: A Standard that defines elements and structures users can expect to find in metadata and the names and meaning of those elements.

Metadata for Discovery: Information that allows users to search and find data holdings using text, keyword, temporal, and spatial queries, and to locate a contact person for the data they discover.

Metadata and Documentation for Understanding: Information that allows independent understanding of data by anticipated and unanticipated users.

Metadata for Use: Information that allows tools and services to identify appropriate data and perform fundamental display, analysis, and comparison operations.

NOAA Archives: The three national Data Centers, NCDC, NGDC, and NODC.

Representation Standard: A Standard or Specification that defines the format of elements and structures used to represent elements from a content standard.

Rubric: A rubric is an explicit set of criteria used for assessing a particular type of work or performance. A rubric usually also includes levels of potential achievement for each criterion, and sometimes also includes work or performance samples that typify each of those levels. Levels of achievement are often given numerical scores. A summary score for the work being assessed may be produced by adding the scores for each criterion. The rubric may also include space to describe the reasons for each judgment or to make suggestions for the author.

Service: An online process that enables applications to submit standard requests and responds with metadata and data in standard formats.

Special Documentation Problems: Important documentation and metadata challenges facing NOAA that will require research, experimentation, and guidance for adapting the recommended standards to specific needs.

Standards Experts: The people who are familiar with the details of the standards and, more importantly, the practices being used throughout the global environmental community to apply those standards in real-world situations.

Spiral Approach: An approach to accomplishing a task by breaking it into multiple iterations or spirals, each of which addresses a small set of critical requirements, and including a feedback cycle with users and other stakeholders prior to the next iteration.

Web-Accessible Folders: Directories on Web Servers that contain files that are accessible through the Web. In metadata applications these generally contain related metadata records in xml files.

Appendix F. ISO TC211 Standards

ISO Technical Committee 211 (TC211) aims to establish a structured set of standards for information concerning objects or phenomena that are directly or indirectly associated with a location relative to the Earth. Understanding what these standards cover and how they fit together can be difficult. The list below was extracted from the TC211 guide to the Technical Specifications and Standards that includes short descriptions of each standard they are responsible for (http://www.isotc211.org/Outreach/Standards_Guide.htm).

STANDARDS THAT SPECIFY THE INFRASTRUCTURE FOR GEOSPATIAL STANDARDIZATION

ISO 19101 Geographic information — Reference model

ISO/TS 19103 Geographic information — Conceptual schema language

ISO/TS 19104 Geographic information — Terminology

ISO 19105 Geographic information — Conformance and testing

ISO 19106 Geographic information — Profiles

STANDARDS THAT DESCRIBE DATA MODELS FOR GEOGRAPHIC INFORMATION

ISO 19109 Geographic information — Rules for application schema

ISO 19107 Geographic information — Spatial schema

ISO 19137 Geographic information — Core profile of the spatial schema

ISO 19123 Geographic information — Schema for coverage geometry and functions

ISO 19108 Geographic information — Temporal schema

ISO 19141 Geographic information — Schema for moving features

ISO 19111 Geographic information — Spatial referencing by coordinates

ISO 19112 Geographic information — Spatial referencing by geographic identifiers

ISO 19156 Geographic information – Observations and measurements (Currently a Draft International Standard)

STANDARDS FOR GEOGRAPHIC INFORMATION MANAGEMENT

ISO 19110 Geographic information — Methodology for feature cataloguing

ISO 19115 Geographic information — Metadata (Revised International Standard expected 2013-05)

ISO 19115-2 Geographic information — Metadata — Part 2: Extensions for imagery and gridded data

ISO 19113 Geographic information — Quality principles (superceded by ISO 19157)

ISO 19114 Geographic information — Quality evaluation procedures (superceded by ISO 19157)

ISO 19131 Geographic information — Data product specifications

ISO 19135 Geographic information — Procedures for item registration

ISO/TS 19127 Geographic information — Geodetic codes and parameters

ISO/TS 19138 Geographic information — Data quality measures (superceded by ISO 19157)

ISO 19157 Geographic information — Data quality (International Standard expected 2013-01)

STANDARDS FOR GEOGRAPHIC INFORMATION SERVICES

ISO 19119 Geographic information — Services

ISO 19116 Geographic information — Positioning services

ISO 19117 Geographic information — Portrayal

ISO 19125-1 Geographic information — Simple feature access — Part 1: Common architecture

ISO 19125-2 Geographic information — Simple feature access — Part 2: SQL option

ISO 19128 Geographic information — Web map server interface

ISO 19132 Geographic information — Location based services — Reference model

ISO 19133 Geographic information — Location based services — Tracking and navigation

ISO 19134 Geographic information — Location base services — Multimodal routing and navigation

STANDARDS FOR ENCODING OF GEOGRAPHIC INFORMATION

ISO 19118 Geographic information — Encoding

ISO 6709 Standard representation of geographic point location by coordinates

ISO 19136 Geographic information — Geography Markup Language (GML)

ISO/TS 19139 Geographic information — Metadata — XML schema implementation

STANDARDS FOR SPECIFIC THEMATIC AREAS

ISO/TS 19101-2 Geographic information — Reference model — Part 2: Imagery

ISO/TS 19130 Geographic information — Imagery sensor models for geopositioning

Appendix G. Resources

A variety of information is available for creating and managing Metadata. They are developed by and aimed at a wide variety of audiences. There are web-based and desktop tools and some that work in multiple environments. This is a partial list of resources that are available:

NOAA's Data Management Integration Team (DMIT) - https://geo-ide.noaa.gov/wiki/index.php?title=DMIT_Membership

NOAA's GEO-IDE wiki - https://geo-ide.noaa.gov/wiki/index.php?title=Main_Page

OAIS RM - http://en.wikipedia.org/wiki/Open_Archival_Information_System

CF Conventions - http://cf-pcmdi.llnl.gov/

ISO TC211 Standards Guide: http://www.isotc211.org/Outreach/ISO_TC_211_Standards_Guide.pdf

CatMDEdit: http://catmdedit.sourceforge.net/

GeoNetwork: http://geonetwork-opensource.org/

Oxygen XML Editor: http://www.oxygenxml.com/

XMLSpy: http://www.altova.com/xmlspy.html

ESRI GeoPortal Server: http://www.esri.com/software/arcgis/geoportal/index.html