≋USGS

NATIONAL CLIMATE CHANGE & WILDLIFE SCIENCE CENTER & CLIMATE SCIENCE CENTERS DATA MANAGEMENT MANUAL

Prepared by:

NCCWSC/CSC Data Management Working Group
US Geological Survey

February 23, 2015

Version 3

More Information: NCCWSC Data Policies

Point of Contact: Emily Fort, Data and IT Coordinator for the NCCWSC and the CSCs, efort@usgs.gov

ACKNOWLEDGEMENTS

Many, many thanks to the following individuals who contributed a great deal to the development of this document and the other NCCWSC/CSC Data Management document.

Name	Organization Representing
Emily Fort	NCCWSC, Data & IT Coordinator
Dave Blodgett	USGS, Center for Integrated Data Analytics
Sean Finn	Great Northern LCC Science Coordinator
Greg Gollberg	Northwest CSC, University of Idaho
Tom Heinrichs	Alaska CSC, University of Alaska – Fairbanks
Viv Hutchison	USGS, Core Science Analytics and Synthesis
Tim Kern	USGS, Fort Collins Science Center and North Central CSC
Natalie Latysh	USGS, Core Science Analytics and Synthesis
Scott Lewein	USGS, Center for Integrated Data Analytics
Stan Smith	USGS, Alaska Science Center and Alaska CSC
Lisa Zolly	USGS, Core Science Analytics and Synthesis

TABLE OF CONTENTS

On September 14, 2009, Secretary Salazar signed a Secretarial Order (No. 3289) entitled, "Addressing the Impacts of Climate Change on America's Water, Land, and Other Natural and Cultural Resources." The Order established the Department's first-ever coordinated strategy to address current and future impacts of climate change on our land, water, wildlife, cultural heritage, and tribal resources. The Secretarial Order directed that eight regional Climate Science Centers (CSCs) and 21 Landscape Conservation Cooperatives (LCCs) be established. (The Caribbean LCC was added at a later point in time bringing the total to 22 LCCs.) The CSCs are aimed at addressing the regional climate science and information needs of a full range of natural and cultural resource managers. The LCCs are management-science partnerships that inform integrated resource management actions addressing climate change and other stressors within and across landscapes.

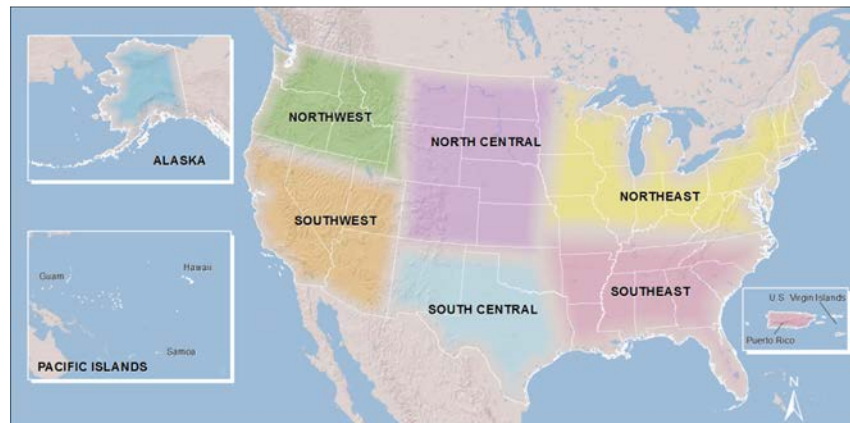Figure 1. Map of Climate Science Center Regions

Congress established the National Climate Change and Wildlife Science Center (NCCWSC) at the U.S. Geological Survey (USGS) in 2008. The NCCWSC serves as DOI's agent for the establishment, management, and coordination of the network of CSCs; delivering scientific and technical information to help natural resource managers cope with a changing climate.

Collectively, the NCCWSC, CSCs, and LCCs will generate, integrate, and disseminate data and derived data products that will help resource managers develop adaptation strategies in response to a changing climate. These managers need data management policy and guidance to ensure that appropriate standards, consistent guidelines, and common strategies are used, providing linkages to and consistency with other similar systems, and fostering a true national network.

NCCWSC/CSC data management policies and guidance will evolve as NCCWSC, CSC and LCC implementation and operations progress. The NCCWSC/CSC Data Management working group was formed to address data management across the NCCWSC/CSC enterprise. Membership of the group includes representatives of each CSC, LCC representatives, and USGS staff with expertise in data management. The NCCWSC/CSC Data Management working group will continue to identify needs, set priorities, and make adjustments to the data management strategy as the CSCs and LCCs progress. As changes occur, the group will update policies, guidance, and this Manual as needed.

IMPORTANCE OF DATA MANAGEMENT

NCCWSC and the CSCs invest in data management and associated tools and resources for many reasons. In order for resource managers to make decisions based on good science, that information needs to be available and accessible. Accessibility helps move the state of the science forward because colleagues can access data and leverage it as part of their work. Data sharing helps researchers get credit for their work and raise its visibility. Data management helps the scientist and research teams. When data are well-managed, they are easier to find, understand, and access. Data loss is also prevented with additional security and back-up procedures. Many funding organizations and groups require data management plans as part of their proposals, and the NCCWSC and the CSCs are consistent with those requirements and the series of laws and policies related to data management. More information on the laws and policies relevant to data management can be found in Attachment A.

LINKS TO OTHER ORGANIZATIONS AND INITIATIVES

The NCCWSC/CSC data management policies and guidance focus primarily on supporting the scientific work of the NCCWSC and CSCs. Data management actively works to create appropriate linkages, minimize duplication, and maximize standards-based interoperability with partner organizations. Policies, guidance, and technology investments will leverage the capabilities and work done by other USGS and partner organizations as much as possible. Collaboration with partners and related activities will be an ongoing, high priority activity.

ORGANIZATION OF AND AUDIENCE FOR THIS MANUAL

A priority for the NCCWSC/CSC enterprise is to make the results of funded scientific research, analyses, and modeling available to the scientific and resource management community as well as the public. Sharing of data and derived data products is vital to the study of climate change and its cascading impacts on resource management.

In order to provide guidance and support for data sharing, the NCCWSC/CSC Data Management Working Group has developed three documents. The first document, a NCCWSC/CSC Data Sharing Policy, requires data and associated products be publically shared and also requires the development of a data management plan. The second document, NCCWSC/CSC Data Management Plan Guidance, provides details to researchers on what should be included in the data management plan that is required as part of the proposal and for funded projects. Finally, the third document (this document), the NCCWSC/CSC Data Management Manual is intended to provide CSC Data Stewards and researchers funded by the NCCWSC or a CSC information on a variety of topics including roles and responsibilities, metadata, file formats, and data standards. Each CSC has a Data Steward to work with funded PIs on the development of the Data Management Plan, answer any questions related to data management, and assist with the transition of data and information products into the NCCWSC/CSC repository.

NCCWSC/CSC ENTERPRISE AND REPOSITORY

The NCCWSC/CSC has developed an enterprise capability to support data sharing and interdisciplinary collaboration. Each CSC and the NCCWSC will function individually and as part of the larger network to ensure that resources, tools, standards, and data are shared across the enterprise. The NCCWSC is committed to defining and building a flexible framework to help researchers focus on science. During the lifespan of a project, the focus will be on supporting the data management of the research and associated products. After the completion of a project, the data and associated products will be transferred to the NCCWSC project and data repository which will permanently store and serve that information. A Principal Investigator (PI) may choose to make their data available through other repositories; however, they are still obligated to provide the NCCWSC/CSC with a copy of datasets and derived data products. In some cases, the NCCWSC may allow data and data products to reside in established third-party repositories, but that exception must be approved by the NCCWSC Data and Information Coordinator.

An agreed upon and mandated data integration framework composed of data, metadata, and web service standards will link the CSCs and the NCCWSCs data and derived data products. Elements of the NCCWSC/CSC data integration framework are defined later in this document.

LAWS AND POLICIES

All activities conducted by the federal government and its employees or contractors are subject to all applicable federal laws and policies. All activities conducted by non-federal government entities (e.g., CSC universities and researchers) are subject to federal law and the terms of the agreement with the federal government. See Attachment A for a listing of relevant laws and policies.

PROJECT TRACKING

Central to the NCCWSC/CSC enterprise is information on funded projects including descriptions of the work, project deliverables, and information on the geographic area of the research (e.g., a shapefile), keywords, schedule, and budget. This information allows tracking and aggregation of project information for both internal and external audiences.

The USGS ScienceBase application will be the primary repository for NCCWSC and CSC funded projects. ScienceBase provides:

- Tools for science and management teams to find, organize, manage and track project activities and deliverables;
- Data uploading and documentation;
- Selective, secure data sharing;
- Integration with other USGS systems such as the Publications Warehouse and BASIS+; and
- Serving data and metadata through standards-compliant Web services.

NCCWSC and CSC funded researchers should review their project information, available on the NCCWSC website. Any corrections or additional information (pictures, publications, data, websites, etc.) that can be shared should be provided to Emily Fort (efort@usgs.gov) or Holly Padgett (hpadgett@usgs.gov). As projects develop reports, publications, or websites, the project information should be updated to reflect those accomplishments and information.

NCCWSC and CSC project and data information will be available via partner sites that are using a standards-based, integrated approach for sharing project metadata. These sites include the Northwest Knowledge Network (NKN) and the Geo Data Portal (GDP). If other programs or sites would like to partner with the NCCWSC and the CSCs to share metadata, please contact Emily Fort (efort@usgs.gov).

ROLES AND RESPONSIBILITIES

The NCCWSC is centrally funding data stewards that can assist NCCWSC and CSC funded researchers with their data management plans and implementation questions. The Data Stewards contact all funded PIs at the beginning of their projects to facilitate the completion of a data management plan and answer any questions. As projects near completion, these stewards will assist the research teams in the transition of data and derived data products into the NCCWSC/CSC repository. Contact information for the data stewards can be found in Attachment C, and on the NCCWSC website.

The tables below identify various activities and the person responsible for those activities. There will likely be exceptions to these generalizations. Those exceptions should surface during the data management planning and can be discussed and clarified at that time.

Table 1. Roles and Responsibilities Between Researchers and Data Stewards

Activity	Role	
	Researcher	CSC Data Steward
Collect data or reuse existing datasets	Responsible	Supportive
Perform data quality review	Responsible	Supportive
Describe scientific workflow/process	Responsible	Supportive
Provide standards-compliant metadata	Responsible	Supportive
Submit data and data products	Responsible	Supportive
Preserve data and data products	Consulted, Informed	Responsible
Provide formats (e.g., web services, NetCDF, etc.) for data discovery and integration	Consulted, Informed	Responsible

An additional responsibility of researchers is citing other's data and data products as appropriate when preparing documentation, reports, or references of their work. To assist researchers, this manual provides information on standards developed in the earth science community that can be used when citing data and datasets in Attachment B.

COLLECT DATA OR REUSE EXISTING DATASETS

Researchers funded by the NCCWSC or the CSCs will collect new data in the field and/or reuse existing datasets. The NCCWSC/CSC Data Management Guidance has been organized to collect information on each dataset – including both new and existing. During the proposal process, a subset of the DMP is required for the Data Management Plan (DMP). If the proposal is selected for funding, the Data Management Plan (DMP) must be completed in full by the research team within one month of receiving funding and submitted to the Data Steward for that CSC. A list of CSC Data Stewards is in Attachment C and on the NCCWSC website.

As outlined in the NCCWSC/CSC Data Management Guidance, PIs are required during the data input phase to carefully track and document the following:

- Existing Dataset(s) or Data Source(s): Source for data, any restrictions on its reuse, and processing or workflow steps that transform the existing data into a new dataset
- New Dataset(s): Details on how data were collected, prepared, and reviewed

Information from the DMP is an important first step in collecting project and data-level documentation. Careful tracking of these components will facilitate more thorough documentation of processes and products in the required project and dataset-level metadata records (see below for more information).

PERFORM DATA QUALITY REVIEW

The usefulness of a particular dataset can only be determined based on documentation of quality and other dataset characteristics. Quality assurance and checks may come in several forms, including software tests, validation tests, tests for data anomalies or outliers, and output reproducibility. As part of the DMP, researchers are asked to describe their steps to ensure their data and associated data products are of high quality. The CSC data stewards can help answer any questions, share examples, and assist with documentation. See Attachment C for CSC data steward contact information.

DESCRIBE SCIENCE WORKFLOW/PROCESS

Analytical workflow documentation is critical to understanding data outputs and products and to allow reproduction of the transformations applied to the data inputs – existing or new – into outputs. This documentation can be done using visualization and/or software tools such as UML diagrams, VisTrails, Kepler, Visio, etc. or in other formats. Visualizations need to be supported by written descriptions such that they can be easily interpreted. Researchers are responsible for describing the process but the CSC Data Steward will assist with questions. Workflow documentation will be incorporated into the project or data-level metadata record.

PROVIDE STANDARDS COMPLIANT METADATA

INTRODUCTION TO METADATA

Metadata is a detailed description of a dataset that includes information about the content, quality, structure, accessibility and other characteristics of the data. Detailed and robust metadata must document the project, datasets, and services in such a way that data can be transmitted, interpreted, reused, and understood. Use of metadata standards is critical for data discovery, integration, and sharing.

Standards for metadata vary in scope and purpose, including metadata written by people to be read and understood by others while other metadata is machine-generated and describes web services to enable automation, data integration, and discovery. There are also numerous standards that focus on particular scientific domains' unique data description needs. Though many metadata standards share elements, one might be the more appropriate choice for a specific dataset, depending on the data being described.

The Federal Government is required to follow the Federal Geographic Data Committee (FGDC) metadata guidance to document spatial datasets developed with federal funding. FGDC is in the process of transitioning from the current standard, FGDC Content Standard for Digital Geospatial Metadata (CSDGM) to the International Organization for Standardization (ISO) 19115 and associated standards. The NCCWSC/CSC preference is for ISO metadata. However, during the transition, we will continue to accept the current FGDC CSDGM standard. With the use of ISO standards and service oriented data access, metadata for machine interpretation is becoming increasingly important. Recognizing this, the NCCWSC/CSC program requires that data published using web services be implemented following applicable attribute conventions such as those described by Network Common Data Format – Climate and Forecast (NetCDF-CF) convention or Open Geospatial Consortium (OGC) Web Service standards. For more information on the various metadata standards, see Attachment D.

PROJECT METADATA

Project-level metadata describes such attributes as the actors, scope, purpose, methods, timeline, and geographic footprint of the project. NCCWSC/CSC requires a compliant FGDC Content Standard for Geospatial Metadata record to describe project-level metadata. Much of the information required for a project level metadata record can be created using information submitted during the proposal submission process. Additional information will be gathered from the research team by the CSC Data Steward. This information will be entered into ScienceBase, the NCCWSC/CSC project tracking tool and data repository. Using ScienceBase's web services, project information, including related products and data, will be displayed on the NCCWSC and relevant CSC websites. There are several good examples for creating Project Metadata such as the Alaska Data Integration Working Group's Briefing and Findings Paper on Project Metadata.

METADATA FOR UNDERSTANDING

Metadata content standards (e.g., ISO 19115, FGDC CSDGM) used for documentation are primarily intended for human understanding of the dataset or project. Metadata for understanding ensures that another researcher can interpret datasets properly and thus data can be reused or understood in the future. Each NCCWSC or CSC funded project must provide dataset metadata that complies with the federal standards. The USGS Data Management website lists several tools that can assist with the development of metadata. The CSC Data Steward will assist the research team with any questions.

DATASET METADATA

Metadata must be available to describe each dataset generated by a project. For the purposes of cataloging, metadata should summarize datasets at what is commonly referred to as the "collection" level. This is the highest level of granularity (most coarse) that can be used to document a dataset uniquely. This approach is critical to the ability of data consumers to search for and identify datasets where the important, unique characteristics of a dataset are clear and distinguishable from many similar datasets. Metadata creation and cataloging tools are available as well as the CSC Data Stewards to assist in creating this metadata content. NCCWSC will accept dataset metadata that follows any broadly accepted interoperable dataset metadata standard that is applicable to the data in question.

SERVICE METADATA

Data should be published using appropriate web service standards whenever possible. The metadata to describe these services can be manually or automatically added to a dataset metadata record as a separate, standalone but related component. In addition to dataset metadata, a service-level metadata record should include contact information (contact name, email, physical address and position), a title, keywords, an abstract, and any access constraints for the data service itself. This allows not only machine-to-machine interaction with the data service, but also human understanding of the data service.

INTEGRATED CATALOG AND DISCOVERY

ScienceBase and the Geo Data Portal, dataset metadata catalogs capable of handling web services and the data they hold, are available from NCCWSC. These catalogs rely on ISO content and encoding standards with support for numerous scientific domain and data-type specific meta-attributes (from several metadata standards). These catalogs and associated data discovery tools will be designed in cooperation with third party data archives. While this system of metadata catalogs and data repositories forms develop and mature, guidance and requirements regarding data publication and documentation will evolve.

SUBMIT DATA AND DATA PRODUCTS

At the conclusion of CSC and NCCWSC funded research, the PI is responsible for submitting all new data collections, derived data products, and links to publications that were developed as part of the funded research to the NCCWSC and/or the sponsoring CSC. Before information is submitted, it must be peer reviewed. Proposals with USGS researchers must follow the Fundamental Science Practices for both publications and data. This includes using the appropriate USGS tools and resources such as the Information Product Data System (IPDS) and the Science Publishing Network (SPN). Proposals with university researchers must also follow appropriate scientific peer review principles.

After the research proposal has been awarded, the CSC Data Steward can assist the PI in identifying the specific data collections and derived data products that will be required for submission at the conclusion of the project, including the use of recommended data and metadata standards. As the project nears completion the PI should once again consult with the CSC Data Steward to review and assist in formatting of data and preparation of metadata for transmission to the NCCWSC and sponsoring CSC.

Attachment E of this document provides guidance to researchers about supported data and metadata standards by data type. The PI may format his/her data for transmission using any of the supported formats for a data type. All exceptions to formats listed in Attachment E must be pre-approved by the NCCWSC and sponsoring CSC Data Stewards.

PRESERVE DATA AND DATA PRODUCTS

The NCCWSC/CSC network assumes full responsibility for the long-term care and preservation of all data collections and derived data products that result from its funding. Once the PI has transmitted the project data to the NCCWSC and/or sponsoring CSC the PI is cleared of further responsibility for the data's preservation. The NCCWSC/CSC network may make portions of the data available to the public through portals or other distribution points of its choosing according to procedures outlined in section "Provide Mechanisms for Data Discovery and Integration" of this document. The PI does retain the right to maintain his/her copies of the research data for future use including distribution at their discretion.

PROVIDE MECHANISMS FOR DATA DISCOVERY AND INTEGRATION

In addition to preserving data and data products, the NCCWSC/CSC network is developing mechanisms to enable data discovery and integration. To encourage data discovery, the NCCWSC/CSC network has a master project catalog and data repository displayed on the NCCWSC website as well as the CSC websites.

To allow for data integration, the NCCWSC/CSC enterprise is adopting common, broadly accepted standards and ensuring access to these data sources. The NCCWSC strongly encourages the use of open data encoding standards but acknowledges that some scientific communities commonly work with proprietary formats. In these situations, community norms should dictate allowed formats. If a project works with a proprietary format (e.g., MS Excel), the project needs to consider alternative file formats for product presentation. At the outside,

the data should be submitted in a non-proprietary format as well as the proprietary format (e.g., production of a comma delimited text file in addition to representation in a MS Excel spreadsheet). In all cases, any non-standard or proprietary file format produced must be identified and justified in the DMP. Details on the common standards and data encoding services can be found in Attachment F.

The Laws and Policies relevant to management of data, standards, and computer applications in the guardianship of or contracted by a federal agency are shown in the tables below. These abbreviated descriptions are summarized from the full text of the U.S. Geological Survey Data Management Plan Framework – Layer II – Part A, Laws and Policies. Please reference that document for additional information. This list is not intended as an exclusive list. Rather, it highlights particular laws and policies that are of interest.

1.1.   FEDERAL

LAWS

Laws are an encapsulated set of rules or acts passed by a legislative body to govern the affairs of a 'community' which are enforced by a political authority and legal system.

Identifier	Title	Brief Description	Data Management Relevance
FOIA	Freedom of Information Act	The FOIA grants anyone the right to access any DOI records unless DOI reasonably foresees that the release of the information would harm a protected interest or falls under one of the nine exemptions.	The public and other researchers have the right to all research data upon request.
	Link: http://www.doi.gov/foia		
CCA	Clinger-Cohen Act of 1966 a.k.a. "Information Technology Reform Act"	Rules for information technology acquisition and management (primarily for hardware and software)	Agencies must track their information technology spending and manage its assets.
	Link: https://www.fismacenter.com/Clinger%20Cohen.pdf		

Identifier	Title	Brief Description	Data Management Relevance
E-Gov Act	E-Government Act of 2002 a.k.a. "Title II - Federal Management and Promotion of Electronic Government Services § 207 – Accessibility, usability, and preservation of government information"	The purpose of the E-Gov Act is to improve the methods by which Government information, including information on the Internet, is organized, preserved, and made accessible to the public.	Provides guidance and structure for sharing government information on the Internet.
	Link: http://www.archives.gov/about/laws/egov-act-section-207.html		
CIPSEA	Confidential Information Protection and Statistical Efficiency Act of 2002	CIPSEA provides strong confidentiality protections for statistical information collections and statistical activities such as data analysis and sample design that are sponsored or conducted by Federal agencies.	Data that contains PII or other sensitive information must be summarized to abstract the sensitive information.
	Link: http://www.whitehouse.gov/omb/memoranda_m03-22/		
Section 508	Section 508 of the Rehabilitation Act of 2000	Provide open access to information products, with provisions for serving people with disabilities	Project data and communication products of significance to the general public must be made 508-compliant. 2011 revisions are available in draft.
	Link: http://www.section508.gov/index.cfm?fuseAction=1998Amend		

Identifier	Title	Brief Description	Data Management Relevance
PRA	Paperwork Reduction Act of 1995	This Act is a subpart of the Coordination of Federal Information Policy, and is intended to streamline and standardize government IT planning, purchasing, operations, and activities.	Requires data created or managed by the federal government is to effectively and efficiently be made accessible and usable to the public.
	Link: http://www.gpo.gov/fdsys/pkg/PLAW-104publ13/html/PLAW-104publ13.htm		
GPEA	Government Paperwork Elimination Act of 1998	GPEA requires federal agencies to allow individuals or entities that deal with the agencies the option to submit information or transact with the agency electronically. The Act specifically states that electronic records and their related electronic signatures are not to be denied legal effect.	Data Management plans must provide for external researchers and labs to submit findings electronically. Standards can still be applied to requirements.
	Link: http://www.whitehouse.gov/omb/fedreg_gpea2/		

Identifier	Title	Brief Description	Data Management Relevance
Bayh Dole Act	Patent and Trademark Law Amendments Act (the Bayh Dole Act) of 1980	In 1980, the Bayh-Dole Act (PL 96-517, Patent and Trademark Act Amendments of 1980) created a uniform patent policy among the many federal agencies funding research. As a result of this law, universities retain ownership to inventions made under federally funded research. In return, universities are expected to file for patent protection and to ensure commercialization upon licensing.	Research conducted by universities is subject to the Bayh Dole Act requirements.
	Link: http://grants.nih.gov/grants/bayh-dole.htm		

EXECUTIVE ORDERS

An Executive Order is a directive issued by the executive head of a government office, generally the President of the United States. Executive orders are considered to carry the full weight of the law since they are usually issued in conjunction with implementing certain acts of Congress. Executive Orders may also be issued by Governors or Mayors.

Identifier	Title	Brief Description	Data Management Relevance
CSDGM	Content Standard for Digital Geospatial Metadata – Executive Order 12906	CSDGM is the federal standard for the documentation of geospatial data. The standard was defined by the Federal Geospatial Data Committee.	Projects must adhere to metadata requirements and geographic data management guidelines. In 2012 the 1993 (revised 1998) metadata standard is to be replaced by ISO 19115-2.
	Link: http://www.archives.gov/federal-register/executive-orders/pdf/12906.pdf		

OMB CIRCULARS

A government circular is a written statement of government policy. It provides information, guidance, rules, and often background information on the reasoning behind a policy. OMB Circulars are instructions or information issued by the Office of Management and Budget to Federal Agencies. These are expected to have continuing effect for two years or more.

Identifier	Title	Brief Description	Data Management Relevance
A-16	Coordination of Geographic Information and Related Spatial Data Activities - OMB Circular A-16 (1990, 2003, 2010)	This Circular provides direction for federal agencies that produce, maintain or use spatial data either directly or indirectly in the fulfillment of their mission.	This Circular established a coordinated approach to electronically developing the National Spatial Data Infrastructure and establishes the Federal Geographic Data Committee (FGDC).
	Link: http://www.whitehouse.gov/omb/circulars_a016_rev/		

Identifier	Title	Brief Description	Data Management Relevance
A-110	Requirements for Grants and Agreements - OMB Circular A-110 (1999, 2003) a.k.a. "Uniform Administrative Requirements for Grants and Agreements with Institutions of Higher Education, Hospitals, and Other Non-Profit Organizations"	This Circular sets forth standards for obtaining consistency and uniformity among Federal agencies in the administration of grants to and agreements with institutions of higher education, hospitals, and other non-profit organizations.	
	Link: http://www.whitehouse.gov/omb/circulars_a110/		
A-119	Federal Participation in the Development and Use of Voluntary Consensus Standards and in Conformity Assessment Activities – OMB Circular A-119 (1998)	The revised Circular establishes policies on Federal use and development of voluntary consensus standards and on conformity assessment activities; and authorized the National Institute of Standards and Technology to coordinate conformity assessment activities of the agencies.	Agencies must use common standards when creating and managing data resources, including use of subject area domains and common data elements. New standards and methods should be well-documented.
	Link: http://www.whitehouse.gov/omb/circulars_a119/		
A-130	Management of Federal Information Resources – OMB Circular A-130	General policies that apply to the information activities of all agencies of the Executive Branch of the Federal government.	This is a set of mandates for effective stewardship of data and information. Establishes requirement for Data Management.
	Link: http://www.whitehouse.gov/omb/circulars_a130_a130trans4		

The table below contains policy and guidance issued by parts of the Executive Office of the President, which includes OMB's E-Government Office and the Office of Science and Technology Policy.

Identifier	Title	Brief Description	Data Management Relevance
DRM	Data Reference Model, Version 2.0	The DRM is a framework whose primary purpose is to enable information sharing and reuse across the federal government via the standard description and discovery of common data and the promotion of uniform data management practices.	The DRM provides a standard means by which data may be described, categorized, and shared. These are reflected within each of the DRM's three standardization areas: data description, data context, and data sharing.
	Link: http://www.whitehouse.gov/sites/default/files/omb/assets/egov_docs/DRM_2_0_Final.pdf		
National Strategy for Information Sharing and Safeguarding	National Strategy for Information Sharing and Safeguarding	Provides guidance for effective development, integration, and implementation of policies, processes, standards, and technologies to promote secure and responsible information sharing	Provides principles and goals to guide information sharing between federal organizations and partners
	Link: http://www.whitehouse.gov/sites/default/files/docs/2012sharingstrategy_1.pdf		
Digital Government Strategy	Digital Government: Building a 21st Century Platform to Better Serve the American People	Provides strategies and guidance for making government information more accessible to its citizens	Encourages the use of standards, new technologies, interoperability, and openness

Identifier	Title	Brief Description	Data Management Relevance
	Link: http://www.whitehouse.gov/sites/default/files/omb/egov/digital-government/digital-government-strategy.pdf		
Access to the Results of Federally Funded Scientific Research	Memorandum for the Heads of Executive Departments and Agencies, Increasing Access to the Results of Federally Funded Scientific Research	Provides guidance to agencies to release the results of scientific research funded by the federal government publically	Requires public access to information
	Link: http://www.whitehouse.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf		

1.2. DOI POLICIES

Policies are a course of action or policy intended to influence the actions or decisions made in the course of doing business.

Identifier	Title	Brief Description	Data Management Relevance
375 DM 12	Department of the Interior Departmental Manual Information Resources Management, Part 375 IRM Program Management, Chapter 12: Information Resources Standards Program	The Information Resources Standards Program coordinates the development, adoption, implementation, and review of information management, automated data processing, and telecommunications standards.	Standards play an important role in ensuring interoperability and compatibility between systems.
	http://elips.doi.gov/ELIPS/DocView.aspx?id=1208		

Identifier	Title	Brief Description	Data Management Relevance
378 DM 1	Department of the Interior Departmental Manual Information Resources Management, Part 378 Data Resource Management, Chapter 1: Program Description and Objectives	The Data Resource Management Program encompasses the process of planning managing, controlling and protecting DOI data assets while supporting DOI business functions and goals.	Promotes the adoption of standards and practices that encourage the sharing and exchange of information to further enhance mission and business performance for DOI.
	http://elips.doi.gov/ELIPS/DocView.aspx?id=1237		
Internet Accessibility Policy	Agency-wide Internet Accessibility Policy	The CIO web page does not restate Section 508, but does provide guidelines for 'Best Practices', responsibilities, and tools to assist in complying with section 508 requirements.	
	Link: http://www.section508.gov/agency-wide-it-accessibility-policy		
IP6	Transition to IPv6	This CIO Memorandum states the Department's commitment to the operational deployment and use of Internet Protocol version 6 (IPv6).	
	Link: https://cio.gov/innovate/transition-to-ipv6/		

Identifier	Title	Brief Description	Data Management Relevance
NIST 800-53	National Institute of Standards and Technology – (NIST) Special Publication 800-53	All agencies are required to manage the security of servers, networks (LAN, WAN), and computers must conform to security guidelines, and administrative rights to servers and computers should be limited to those staff that are required for operational use. All users with administrative rights must have an approved request, with justification, on file with the agency's Office of Chief Information Office.	Security must be managed effectively at all levels of development, implementation, and deployment. Approvals for persons with administrative rights for computers within the agency are documented and filed with the CIO.
	Link: http://csrc.nist.gov/publications/PubsSPs.html#800-53		

1.3. USGS POLICIES

Identifier	Title	Brief Description	Data Management Relevance
Survey Manual Section 502.1	U.S. Geological Survey Manual Section 502.1 – Fundamental Science Practices.	USGS Fundamental Science Practices govern how scientific investigations, research, and activities are planned and conducted and how information products are reviewed and approved for release and dissemination.	There are minimum data and information management requirements to which all USGS organizational units must adhere.
	Link: http://www.usgs.gov/usgs-manual/500/502-1.html		

1.4. NCCWSC

NCCWSC's data management policies and guidance can be found here: https://nccwsc.usgs.gov/content/data-policies-and-guidance .
This website includes the Data Sharing Policy, the Data Management Plan Requirement and Guidance, and the Data Management
Manual.

ATTACHMENT B: DATA CITATION

Data citation is an emerging and critical practice in science publishing. As data is shared with more frequency, data citation provides numerous advantages including reproducibility through direct reference to the data used in a research study; providing credit to data creators and authors, research provenance, and ability for researchers to track the use of their datasets in other studies.

Data must be cited and referenced in publications and other media using a bibliographic reference similar to that used for journal articles. Individual researchers or data set owners may define data citation requirements for their data and research; however, the expectation of NCCWSC/CSC is that certain specific elements are included. These elements are:

- Author/Principal Investigator/Data Creator
- Release Date/Year of Publication – year of release, for a completed dataset
- Title of Data Source – formal title of the dataset
- Version/Edition Number – the version of the dataset used in the study
- Format of the Data – physical format of the data
- 3rd Party Data Producer – refers to data accessed from a 3rd party repository
- Archive and/or Distributor – the location that holds the dataset
- Locator or Identifier – includes Digital Object Identifiers (DOI), Handles, Archival Resource Key (ARK), etc.
- Access Date and Time – when data is accessed online
- Subset of Data Used – description based on organization of the larger dataset
- Editor or Contributor – reference to a person who compiled data, or performed value-added functions
- Publication Place – city, state, and country of the distributor of the data
- Data within a Larger Work – refers to the use of data in a compilation or a data supplement (such as published in a peer-reviewed paper)

These guidelines are based on the work of Earth Science Information Partners (ESIP). More information, examples, and guidance can be found at the Earth Science Information Partners (ESIP) website: http://wiki.esipfed.org/index.php/Interagency_Data_Stewardship/Citations/provider_guidelines.

ATTACHMENT C: CONTACT INFORMATION FOR THE CSC DATA STEWARDS

Click here for the CSC Data Steward contact information.

More information on various metadata standards can be found below.

ISO 19115/19139:
The 19115 standard, developed by the International Organization for Standardization (ISO) is an internationally accepted standard for documenting scientific datasets. ISO 19115 defines the schema required for describing geographic information and services. It provides information about the extent, the quality, the spatial and temporal schema, spatial reference, and distribution of digital geographic data.

ISO 19115 adds functionality not found in the FGDC standard, for example, in areas such as multi-lingual data sharing, topic categories for high-level metadata classification, unique identifiers for metadata records, roles and responsibilities for a dataset, descriptions of geospatial service metadata.

 ISO 19139 defines the XML schema implementation derived from 19115.

More information:
http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=26020).

ISO 19115-2
This ISO standard includes extensions for imagery and gridded data elements.

More information: http://www.fgdc.gov/metadata/geospatial-metadata-standards or www.iso.org and from NOAA: https://geo-ide.noaa.gov/wiki/index.php?title=Main_Page

FGDC Content Standard for Digital Geospatial Metadata (with accompanying profiles and/or extensions to include the Biological Data Profile, Shoreline Data Profile, and the Remote Sensing Extension.)

The CSDGM was developed by the FGDC in 1994 and revised in 1999. The standard has a primary focus on geospatial data, but can also be used for tabular data. The CSDGM can be used with accompanying profiles and/or extensions. A profile/extension contains additional elements that are incorporated into the core metadata standard to fill the need of a particular community. The Biological Data Profile, Shoreline Data Profile, and the Remote Sensing Extension are three examples of extended elements developed to suit a particular scientific domain.

More Information: To download guidance about the standard, profiles and extensions, visit www.fgdc.gov/metadata.

Ecological Metadata Language (EML):
This community-driven standard focuses primarily on documenting projects in ecology. The standard was built using modules to create an extensible functionality, such that users of the standard can determine which modules are most pertinent to describing their data, literature, and software

resources. These modules can then be linked through a crosswalk to other established standards. Ecological Metadata Language (EML) was developed by a collaboration of individuals at National Center for Ecological Analysis and Synthesis (NCEAS), the Long Term Ecological Research Program, and the Joseph W. Jones Ecological Research Center.

More information: http://knb.ecoinformatics.org/software/eml/

The NCCWSC has approved the following formats for transmitting research data to the NCCWSC and sponsoring CSC at project close. Search the left most column of the table below to locate the data type for the collection you need to transmit and read the row to find supported data and metadata formats.

The NCCWSC and CSC enterprise is in early stages of development. If a format is missing that a researcher needs, please contact a CSC Data Steward to suggest additions.

Type of data	Acceptable formats for transmission to NCCWSC/CSC	Acceptable metadata formats
Quantitative tabular data with extensive metadata *a dataset with attribute labels, code labels, defined missing values, and attribute definitions in addition to the data matrix*	Structured data from a database such as MySQL, PostGres, and MS Access [.mdb	.accdb]	ISO 19115-2:2009
Quantitative tabular data with minimal metadata *a dataset with or without attribute labels but no other metadata in addition to the data matrix*	ASCII comma separated values (CSV) [.csv] ASCII tab delimited file [.tab]	ISO 19115-2:2009
Geospatial data *vector and raster data*	.shp, .tif/geotiff, Keyhole Mark-up Language (KML) [.kml]	ISO 19115-2:2009
Qualitative data *textual*	ASCII Rich Text Format [.rtf] ASCII plain text data [.txt]	
Gridded data	Network Common Data Form (NetCDF) [.nc	.cdf]	ISO 19115-2:2009

Type of data	Acceptable formats for transmission to NCCWSC/CSC	Acceptable metadata formats
Digital image data	Joint Photographic Experts Group (JPEG) [.jpg	.jpeg] Adobe Portable Document Format (PDF) [.pdf] TIFF version 6 uncompressed [.tif]	ISO 19115-2:2009
Digital audio data	MPEG-1 Audio Layer 3 [.mp3] Waveform Audio Format (WAV) [.wav]	
Digital video data	Moving Picture Experts Group Standard definition: MPEG-4 [.mp4	AVC/H.264] High Definition: [AVCHD/H.264]	
Documentation, publications, and scripts	ASCII Rich Text Format [.rtf] Adobe Portable Document Format (PDF) [.pdf]	

Note: Exceptions to this list must be approved by the NCCWSC and the responsible CSC Data Steward.

DATA ENCODING STANDARDS AND FORMATS

NCCWSC/CSC will promote data sharing and integration through the use of data standards such as NetCDF-CF, OGC Standards, and commonly used formats, like shapefile and geotiff. Unifying software packages that support numerous file formats matching a common data model such as GDAL, for geospatial gridded data; OGR, for geospatial vector data; and NetCDF-Java, for file-encoded time series data will be relied on to make connections to the numerous file formats used in scientific practice.

Network Common Data Format with Climate and Forecasting (NetCDF-CF) metadata conventions: NetCDF is "a set of interfaces for array-oriented data access and a freely-distributed collection of data access libraries for C, Fortran, C++, Java, and other languages. The netCDF libraries support a machine-independent format for representing scientific data. Together, the interfaces, libraries, and format support the creation, access, and sharing of scientific data." (Unidata http://www.unidata.ucar.edu/ Accessed June 13, 2010) NetCDF functionality has been developed primarily for the atmospheric science community. The Climate and Forecasting (CF) metadata conventions have been developed by the atmospheric science community to describe NetCDF compatible datasets. Combined, the NetCDF and CF data standards provide a fully described common data model accessible by programmatic tools that support data discovery, access, and integration.

Open Geospatial Consortium (OGC) Standards: OGC standards are internationally developed interchange methods, information models, and xml application schemas. Primarily, OGC has focused on data that could be conveyed on a map or analyzed in a GIS setting. Increasingly, OGC domain working groups are defining information models and xml application schemas for domain specific observational data. These domain specific information models will offer internationally recognized standardization for information such as hydrologic model output or groundwater monitoring network observations. Using a combination of OGC standards, a wide range of NCCWSC related data formats and use cases can be accessed and satisfied, respectively.

De-facto Community Standards: Depending on the community developing the dataset, products may fall outside the previous mentioned data encoding standards and file formats. The NCCWSC strongly encourages the use of open data encoding standards but acknowledges that some scientific communities commonly work with proprietary formats. In these situations, community norms should dictate allowed formats. If a project works with a proprietary format (e.g., MS Excel), the project needs to consider alternative file formats for product presentation. At the outside, the data should be submitted in a non-proprietary format as well as the proprietary format (e.g., production of a comma delimited text file in addition to representation in a MS Excel spreadsheet). In all cases, any non-standard or proprietary file format produced must be identified and justified in the DMP.

DATA EXCHANGE STANDARDS

Web Services provide a standard way for software applications to interoperate using the Hypertext Transfer Protocol (HTTP). Building on HTTP with infrastructure, architecture, and core technologies under the auspices of the World Wide Web Consortium (W3C), web service standards have been defined that enable systems to be created by chaining networked services together. Since the network may be the entire Internet, web service technology can expedite the creation of entirely new applications by assembling discrete, possibly geographically distributed services.

The Catalog Service for the Web (CSW) standard is an OGC metadata and other catalog resource web service protocol specification that allows catalogs of metadata to be searchable and accessed by any client software (e.g., ScienceBase, Geo Data Portal) that has implemented the standard. NCCWSC/CSC data stewards and partners are actively involved in efforts to standardize the usage of metadata and catalog service for the web implementations across the geographic and atmospheric science communities.

The Web Map Service (WMS) standard is a widely implemented OGC specification for exchange of map images rendered to a requested size and resolution. It provides a mechanism to request previews of data resources that can be rendered and delivered for immediate display. The WMS standard is especially useful for data discovery and evaluation as it provides efficient visualization of potentially massive data sources. Data servers implemented by the geospatial and atmospheric data communities support the protocol.

The Web Feature Service (WFS) standard provides functionality to query and retrieve geospatial features and associated attributes. Geospatial servers implement the WFS standard as a mechanism to make shapefiles and simple geodatabase content available. Use of WFS for visualization is somewhat limited in that service-consuming software can become overwhelmed by the volume of data delivered in response to a naïve request. For this reason, WFS are commonly coupled with WMS to provide rich functionality and rapid display.

The Web Coverage Service (WCS) standard allows requests for gridded data with server-based handling of geospatial coordinate systems. A data consuming application can request data for an area of interest using common latitude longitude geospatial coordinates. Depending on geospatial server capabilities, the WCS specification provides methods to request data in one of any available geospatial coordinate systems, data formats, and resolutions. As the web coverage service standard and software implementing it evolve, handling for time series and specialized server-performed transformations will improve (or decrease handling time).

The Open Source Project for a Network Data Access Protocol (OPeNDAP) standard provides extremely generalized access to structured data. Coupled with appropriate metadata standards, virtually any type of data can be represented. The highly generalized nature of the standard puts much of the data

integration burden on data consuming software; however, the standard is very useful as data can be self-describing and is easily accessed by project-specific user-written software.