# 2010 SURVEY OF EPA AND OTHER FEDERAL AGENCY SCIENTIFIC DATA MANAGEMENT POLICIES AND GUIDANCE

**U.S. Environmental Protection Agency**
**Office of Research and Development (ORD)**
**Office of Science Information Management (OSIM)**

Contract No: GS-10F-0299K
EPA Order No: EP06H000698

**April 30, 2010**

# Survey of EPA and Other Federal Agency Scientific Data Management Policies and Guidance 2010

U.S. Environmental Protection Agency
Office of Research and Development (ORD)
Office of Science Information Management (OSIM)

April 30, 2010

# Contents

**List of Tables** ..... v

**Section 1 Introduction** ..... 1

1.1 Report Purpose and Approach ..... 2
1.2 Report Contents and Organization ..... 5

**Section 2 Manage Scientific Data as Enterprise Assets or Liabilities (Policy Area #1)** ..... 6

2.1 EPA SDM Policy Information (Policy Area #1) ..... 7
2.2 Other Federal Agency SDM Policy Information (Policy Area #1) ..... 7

**Section 3 Develop a Scientific Data Management Plan that Covers the Full Data Life Cycle (Policy Area #2)** ..... 10

3.1 EPA Policy Information ..... 10
3.2 Other Federal Agency Policy Information ..... 13

**Section 4 Identify Scientific Data with Metadata to Enable Needed Business Operations (Policy Area #3)** ..... 13

4.1 EPA Policy Information ..... 13
4.2 Other Federal Agency Policy Information ..... 19

**Section 5 Manage Scientific Data for Appropriate Control (Policy Area #4)** ..... 19

5.1 EPA Policy Information ..... 23
5.2 Other Federal Agency Policy Information ..... 23

**Section 6 Maintain Version and Change Control on Data Sets (Policy Area #5)** ..... 23

6.1 EPA Policy Information ..... 28
6.2 Other Federal Agency Policy Information ..... 28

**Section 7 Retain Data Commensurate with Its Value (Policy Area #6)** ..... 32

7.1 EPA Policy Information ..... 32
7.2 Other Federal Agency Policy Information ..... 35

**Section 8 Ensure that Scientific Data Management Processes Are Integrated with Knowledge Management Initiative (Policy Area #7)** ..... 35

8.1 EPA Policy Information ..... 35
8.2 Other Federal Agency Policy Information ..... 39

**Section 9 Conclusions** ..... 39

9.1 Resources by Policy Area and Agency ..... 39
9.2 Key Resources ..... 41
9.3 Information Gaps ..... 41
9.4 Next Steps ..... 44

**Appendices**

A Summary of Findings by Office and Policy Area – EPA
B Summary of Findings by Office and Policy Area – Other Federal Agencies
C References
D Additional Resources

# Tables


1 Comparison of SDM principles and policy areas ..... 4
2 Manage scientific data as an enterprise asset or liability: EPA documents and resources (Policy Area #1) ..... 8
3 Manage scientific data as an enterprise asset or liability: Other federal agency documents and resources (Policy Area #1) ..... 9
4 Develop a scientific data management plan that covers the full data life cycle: EPA documents and resources (Policy Area #2) ..... 11
5 Develop a scientific data management plan that covers the full data life cycle: Other federal agency documents and resources (Policy Area #2) ..... 14
6 Identify scientific data with metadata to enable needed business operations: EPA documents and resources (Policy Area #3) ..... 16
7 Identify scientific data with metadata to enable needed business operations: Other federal agency documents and resources (Policy Area #3) ..... 20
8 Manage scientific data for appropriate control: EPA documents and resources (Policy Area #4) ..... 24
9 Manage scientific data for appropriate control: Other federal agency documents and resources (Policy Area #4) ..... 26
10 Maintain version and change control on data sets: EPA documents and resources (Policy Area #5) ..... 29
11 Maintain version and change control on data sets: Other federal agency documents and resources (Policy Area #5) ..... 30
12 Retain data commensurate with its value: EPA documents and resources (Policy Area #6) ..... 33
13 Retain data commensurate with its value: Other federal agency documents and resources (Policy Area #6) ..... 36
14 Number of references to SDM documents and resources by policy area, applicability rating, agency type, and level ..... 40
15 SDM documents and resources with three-star ratings by agency and policy area ..... 42

# 1. Introduction

The U.S. Environmental Protection Agency's (EPA's) Office of Science and Information Management (OSIM) in the Office of Research and Development (ORD) is developing a policy framework and ultimately, a set of policies and related guidance for managing the scientific data created and used by ORD across the entire life cycle of the data (e.g., from initial planning stages, to data gathering, organization, and analysis, to data publishing, to data archiving, potential re-use, and destruction). Developing this comprehensive scientific data management (SDM) policy will involve a long-term (e.g., two years or more) effort requiring the assessment and resolution of many complex information management, information technology, and other issues, and resulting in a series of individual SDM policy statements and related guidance on how to implement these policies. This process will entail the following general approach:

1. *Develop a SDM policy framework.* This framework incorporates seven general policy areas to be covered by the ORD SDM policy, and specific policy and guidance topics within each policy area. This framework will also identify gaps and conflicts that need to be addressed further, and define and prioritize next steps. Information for the SDM policy framework will be based on:

    a. A review of existing EPA and other federal agency SDM policies and guidance in order to collect sample policy approaches, categorize these documents by policy area and other characteristics, and identify information gaps.

    b. A series of workshops that bring EPA, other federal agency officials, and SDM experts together to discuss requirements and best practices for developing SDM policy and guidance.

2. *Develop ORD SDM policies, guidance, and tools*. For each policy area, as identified and prioritized in the preceding task:

    a. Identify what ORD policy, guidance, and tools must be developed for each policy area.

    b. Review and validate the information collected for each policy area, and determine which materials are relevant to ORD's SDM development approach.

    c. Determine additional sources of information, as appropriate (e.g., through additional literature and Internet reviews, workshops, and conversations with EPA and non-EPA agency staff who have developed similar policies).

d. Convene working groups, composed of appropriate EPA headquarters and regional offices and staff [e.g., Office of Environmental Information (OEI), General Counsel, Human Resources], for each policy area. OSIM will work with these groups to assess options within each policy area and develop SDM policies, including defining roles and responsibilities and developing related guidance and training materials.

e. Develop a policy implementation plan that considers issues such as changes in work culture, impacts on automated systems, development of supporting tools, timing of implementation, impacts on large programs, approaches for raising awareness about the need for the SDM plan and for publicizing and promoting the plan, and ensuring that the SDM policies, guidance, and tools are easy to understand and use.

## 1.1 Report Purpose and Approach

This report presents the results of the first step in the process of developing a SDM policy – identifying and summarizing SDM policies and/or guidance documents developed by EPA and other federal agencies. The main purpose is to determine the usefulness of these resources to ORD's goal of developing its own SDM policies, and identify gaps where more research is needed. The approach for developing this report involved the following activities:

- *Conducted an initial Internet and EPA Intranet review of SDM policies.* The first step involved conducting a general Internet and EPA Intranet review to identify literature on data management in general and to begin to identify federal agency policies and guidance that focus on SDM.

- *Defined seven SDM policy areas.* The text box entitled *Recommended ORD SDM policy areas* summarizes seven policy areas listed by general order of the phases of the scientific data life cycle. These policy areas were identified during initial review of SDM documents, based on two sources that focus extensively on defining key

**Recommended ORD SDM policy areas**

1. Manage scientific data as an enterprise asset or liability.
2. Develop a SDM plan that covers the full data life cycle.
3. Identify scientific data with metadata to enable needed business operations (e.g., access control, discovery, linking to products).
4. Manage scientific data for appropriate control [e.g., intellectual property (IP), data rights, proprietary data].
5. Maintain version and change control on data sets.
6. Retain data commensurate with its value.
7. Ensure that SDM processes integrate with KM initiatives.

principles that underlie an effective SDM program. The American National Standards Institute's (ANSI's) *ANSI/GEIA 859 Data Management* defines nine principles of a high-quality data management program. The second source of SDM principles was compiled by the National Research Council (NRC), which impaneled a committee to provide advice on how to archive and provide access to environmental data collected by the National Oceanic and Atmospheric Administration (NOAA) and its partners. The committee published a report, *Environmental Data Management at NOAA*, which identifies nine general principles for effective environmental data management. Based on a review of both sets of principles, a set of seven "policy areas" were developed that address OSIM's main SDM concerns, and that were used to categorize the SDM documents. Table 1 presents a summary of the ANSI and NRC principles, grouped according to similar SDM issues. The seven policy areas defined for this study are shown in the third column of the table, matched up with the relevant ANSI and NRC principles.

- *Identified federal agencies for targeted investigation and conducted research.* Based on the initial Internet and EPA Intranet reviews, research was focused in two areas: (1) existing EPA policies and guidance related to SDM, and (2) SDM practices by other federal agencies that are similar to EPA in terms of size (e.g., small- to moderate-sized agencies) and/or mission (e.g., protecting human health and the environment, developing scientific data). EPA program and office Internet and Intranet sites were searched to identify existing SDM documentation. The goal was to collect examples of current EPA best practices and to ensure that OSIM does not duplicate or reinvent existing policies. Two EPA offices that have developed a variety of SDM policy materials are OEI and the Office of Solid Waste and Emergency Response (OSWER).

  Potential non-EPA federal agencies were identified based on an Internet search, discussions with EPA staff, and a review of the CENDI [formerly called the Commerce, Energy, National Aeronautics and Space Administration (NASA), Defense Information Managers Group] web site. CENDI is an interagency group composed of the scientific and technical information managers from 11 federal agencies engaged in scientific and technical research and development. Based on these sources, several agencies were identified that met the criteria described above and each agency's web site was searched to identify policy and guidance documents related to SDM. Five agencies offered relevant SDM policy and guidance information: the U.S. Department of Energy (DOE), NASA, National Institutes of Health (NIH), NOAA, and National Science Foundation (NSF). More limited information was also captured from other federal agencies, including the U.S. Geological Survey (USGS) and the U.S. Department of Agriculture (USDA).

**Table 1. Comparison of SDM principles and policy areas**

| ***ANSI-GEIA 859 Data Management* principles** | **NOAA principles** | **Related OSIM SDM policy areas** |
|---|---|---|
| Define the enterprise-relevant scope of data management. | | Manage scientific data as an enterprise asset or liability (Policy Area #1). |
| Plan for, acquire, and provide data responsive to customer requirements. | ▸ Data-generating activities should include adequate resources to support end-to-end data management.<br>▸ Environmental data management activities should recognize user needs.<br>▸ Effective interagency and international partnerships are essential (e.g., sharing data). | Develop a SDM plan that covers the full data life cycle (Policy Area #2). |
| Develop data management processes to fit the context and business environment in which they will be performed. | ▸ Effective data management requires a formal, ongoing planning process. | Develop a SDM plan that covers the full data life cycle (Policy Area #2). |
| Identify data products and views so their requirements and attributes can be controlled. | ▸ Metadata are essential for data management.<br>▸ Data and metadata require expert stewardship.<br>▸ An effective data archive should provide for discovery, access, and integration. | Identify scientific data with metadata to enable needed business operations (e.g., access control, discovery, linking to products) (Policy Area #3). |
| Control data, data products, data views, and metadata using approved change control processes. | | Maintain version and change control on data sets (Policy Area #5). |
| Establish and maintain a management process for IP, proprietary information, and competition-sensitive data. | | Manage scientific data for appropriate control (e.g., IP, data rights, proprietary data) (Policy Area #4). |
| Retain data commensurate with value. | ▸ Environmental data should be archived and accessible.<br>▸ A formal ongoing process with broad community input is needed to decide what data to archive and what not to archive. | Develop a SDM plan that covers the full data life cycle (Policy Area #2).<br>Retain data commensurate with its value (Policy Area #6). |
| Continuously improve data management. | | Develop a SDM plan that covers the full data life cycle (Policy Area #2). |
| Effectively integrate data management and knowledge management (KM). | | Ensure that SDM processes integrate with KM initiatives (Policy Area #7). |

- *Identified SDM policy levels*. The targeted research identified several "levels" of SDM policy documents and resources, ranging from broad goals, vision statements, and principles, to policy statements, to general and specific guidance for carrying out the policies. These different policy levels are shown in the text box, at right, *SDM policy levels*.

- *Summarized findings*. A series of tables was developed to summarize the research findings and to categorize these findings by agency, policy area, and policy level. This report presents these tables, along with a description of the SDM policy and guidance documents identified, the relative usefulness of these documents to ORD's goal of developing its own SDM policies, and information gaps where more research is needed.

**SDM policy levels**

- *Goals and vision statements* describe the broad objectives that an agency wants to achieve through its SDM.
- *Principles* are high-level descriptions and statements regarding the development and maintenance of high-quality SDM.
- *Recommendations for policies*, often resulting from an agency committee investigation of SDM issues, present general descriptions of SDM policies that the agency needs to develop.
- *Policies* are relatively brief documents that define a rule for enhancing SDM, and often include related information such as definitions, roles and responsibilities, and references.
- *General guidance* includes general instruction on how a policy should be implemented.
- *Specific guidance* includes more specific, step-by-step instructions on how to implement a SDM policy.
- *Other* includes types of SDM documents and resources that do not fall into any of the preceding categories [e.g., a research analysis, internal audit report, or Government Accountability Office (GAO) report].

## 1.2 Report Contents and Organization

The remainder of this report presents the study findings and conclusions. As shown in the text box, *Report organization*, Sections 2 though 8 summarize study findings by SDM policy area. Each section provides (1) a definition of the policy area and (2) examples of the types of policy and guidance that could be developed within this policy area, followed by brief descriptions of the EPA and other federal documents found by policy level and associated summary tables. Each summary table organizes the SDM policy and guidance documents for the relevant policy area by policy level. For each table, information is provided on the agency and/or office that developed the policy/guidance, the document title and date, a brief description of the document, an "applicability rating," and a link to the source of information. The applicability rating is shown as one, two, or three stars. A one-star rating (∗) means that the information is related to ORD's SDM policy framework, but is expected to be of limited value in developing its policies and guidance. A two-star rating (∗∗) means that the

information is somewhat relevant to ORD's policy framework and/or is presented in limited detail. A three-star rating (∗∗∗) indicates that the information is directly applicable to ORD's SDM policy/guidance goals and could serve as a model for ORD.

Section 9 presents study conclusions, including an overview of the key types of SDM policy information that is available from EPA and other federal agencies, information gaps, and suggested next steps.

Appendices A and B provide tabular summaries of findings organized by office and policy area for EPA and other federal agencies, respectively. Appendix C provides references reviewed for this report. Appendix D provides potentially relevant resources that have been recently identified and have not been incorporated in the findings.

**Report organization**


| | |
|---|---|
| Section 1: | Introduction |
| Section 2: | Manage Scientific Data as Enterprise Assets or Liabilities (Policy Area #1) |
| Section 3: | Develop a Scientific Data Management Plan that Covers the Full Data Life Cycle (Policy Area #2) |
| Section 4: | Identify Scientific Data with Metadata to Enable Needed Business Operations (Policy Area #3) |
| Section 5: | Manage Scientific Data for Appropriate Control (Policy Area #4) |
| Section 6: | Maintain Version and Change Control on Data Sets (Policy Area #5) |
| Section 7: | Retain Data Commensurate with Its Value (Policy Area #6) |
| Section 8: | Ensure that Scientific Data Management Processes Are Integrated with Knowledge Management Initiative (Policy Area #7) |
| Section 9: | Conclusions |
| Appendix A: | Summary of Findings by Office and Policy Area – EPA |
| Appendix B: | Summary of Findings by Office and Policy Area – Other Federal Agencies |
| Appendix C: | References |
| Appendix D: | Additional Resources |


# 2. Manage Scientific Data as Enterprise Assets or Liabilities (Policy Area #1)

A key principle underlying effective SDM policy, and the first policy area described in this report, is the concept that scientific data are EPA assets and liabilities, and should be managed as such. Scientific data developed with ORD resources (such as funding, staff, computers and other equipment) belong to the taxpayer and are governed by EPA. These data have value, which may be positive (i.e., assets) or negative (i.e., liabilities). An example of an asset is a data set that will be reused for trend analysis; an example of a liability is a data set that will never be reused but ORD continues to incur costs for its maintenance.

The development of this policy area will involve, for example, establishing policies or rules for how others can use the data and who owns the data during its initial development. There can also be polices on how to treat data appropriately by defining when it is an asset and when it is a liability. It is important to note that this policy area often overlaps with other areas, such as developing SDM plans, managing data for appropriate control, and data retention and valuation.

## 2.1 EPA SDM Policy Information (Policy Area #1)

Table 2 presents EPA documents related to Policy Area #1, manage scientific data as an enterprise asset or liability. As shown in the table, limited EPA documentation was found for this policy area. One OSWER document addresses this issue in terms of establishing principles and general guidance for treating data as an asset or liability, and one OEI reference also provides general guidance. Both documents are directly applicable to EPA/ORD (i.e., they all have a three-star rating). Three documents discuss specific SDM policies, or provide recommendations for policies. Two of these documents (ORD's *Scientific Data Management Strategy* and OEI's *National Geospatial Data Policy*) contain directly applicable information and one document provides information that would be considered potentially relevant to ORD (i.e., two stars). No supporting documentation was found for "Goals, Vision Statements" or "Specific Guidance."

## 2.2 Other Federal Agency SDM Policy Information (Policy Area #1)

Table 3 presents non-EPA federal documents and resources that relate to Policy Area #1, manage scientific data as an enterprise asset or liability. As shown in the table, eight documents, developed by five federal agencies (DOE, NASA, NIH, NOAA, and NSF), were identified. These agencies recognize that scientific data are an asset that must be managed and made freely available to the public. DOE and the NSF have developed principles regarding the need to make data available. NASA, NIH, and NOAA have policies on retaining valuable data into the future, and two NASA offices have developed specific guidance on how to make data available to the public. Two documents contain information that is directly appropriate for EPA/ORD (i.e., three-star ratings), while the remaining information is considered somewhat relevant for the office's needs (i.e., two-star ratings).

| Table 2. Manage scientific data as an enterprise asset or liability: EPA documents and resources (Policy Area #1) | | | | | |
|---|---|---|---|---|---|
| **Level** | **Agency name/office** | **Document title/date** | **Description/pertinent aspects** | **Applicability[a]** | **Link (or reference)** |
| **Goals, vision statements** | | | | | |
| **Principles** | Office of Solid Waste and Emergency Response (OSWER) | System Life Cycle Management Guidance Part 3 Practice Paper: Data Management During the Life Cycle. January 1989. | This document states: "Data is a valuable resource. Data is collected, stored, and used to support critical OSWER program activities and decisions, making accurate and timely data an important OSWER resource." OSWER data "is used to make decisions affecting public health and safety, environmental quality, and the use of public funds. Without this information OSWER could not perform its mission. The data collected, stored, processed and disseminated by OSWER systems are used to create the information OSWER needs to operate."<br><br>This document also provides details for project managers concerning their responsibilities for data management under OSWER System Life Cycle Management Guidance. The practice paper describes "data management during the system life cycle, and provides guidance concerning major topics that should be addressed by project teams."<br><br>Specifically, page 4 lists several benefits for increasing the focus on data management. | *** | http://www.epa.gov/oswer/docs/oswerlcm/00000021.pdf |
| **Recommendations for policies** | Office of Research and Development (ORD) | Scientific Data Management Strategy. 2007. | An objective of the strategy is to identify and prioritize SDM projects by determining where there are “hidden” data management projects, some of which add significant value to the agency. It identifies others as "pet projects"' that add no value. | *** | EPA, 2007. |
| **Policies** | Office of Environmental Information (OEI) | National Geospatial Data Policy. CIO Policy Transmittal 05-002. 8/24/2008. | The policy states that all EPA investment in geospatial data should be leveraged for enterprise use and managed through enterprise architecture guidance. | *** | Email communication with Lynne Petterson, 6/10/09. |
| | National Health and Environmental Effects Research Laboratory (NHEERL) | NHEERL Data Management Policy and Practices: Genomics and Related High Throughput Data. | The policy states that data collected from human subjects presents a challenge in that sharing of data can only occur if the confidentiality of the subjects has been assured. Assurance must be obtained from the NHEERL Human Subjects Research Official prior to the entry of such data into a centralized data base. | ** | EPA, Undated. |
| **General guidance** | Office of Environmental Information (OEI) | EPA Enterprise Architecture Target Data Architecture. 6/23/2009. | The successful management of information and data as an enterprise asset is of critical importance. To achieve the vision of maximizing the value of enterprise data assets, EPA will establish an Enterprise Data Architecture (EDA) Program to create a proactive, enterprise service organization focusing specifically on critical data management issues and challenges faced by EPA programs and their partners. | *** | Email communication with Kevin Kirby. 7/14/09. |
| | Office of Solid Waste and Emergency Response (OSWER) | System Life Cycle Management Guidance Part 3 Practice Paper: Data Management During the Life Cycle. January 1989. | This document states: "If you choose an approach that doesn't address data dictionary issues as part of a large, high impact project, you will increase the risk of time and cost overruns for your project." | *** | http://www.epa.gov/oswer/docs/oswerlcm/00000021.pdf |
| **Specific guidance (e.g., how to interpret and use policies)** | | | | | |
| **Other (specify)** | | | | | |

a. The applicability rating is shown as one, two, or three stars. A one-star rating (*) means that the information is related to ORD’s SDM policy framework, but is expected to be of limited value in developing its policies and guidance. A two-star (**) rating means that the information is somewhat relevant to ORD’s policy framework and/or is presented in limited detail. A three-star rating (***) means that the information provided is directly applicable to ORD’s SDM policy/guidance goals and could serve as a model for ORD.

| Table 3. Manage scientific data as an enterprise asset or liability: Other federal agency documents and resources (Policy Area #1) | | | | | |
|---|---|---|---|---|---|
| **Level** | **Agency name/office** | **Document title/date** | **Description/pertinent aspects** | **Applicability[a]** | **Link (or reference)** |
| **Goals, vision statements** | | | | | |
| **Principles** | Department of Energy (DOE) – Oak Ridge National Laboratory (ORNL) | Developing Data Management Policy and Guidance Documents for your NARSTO Program or Project. 2005. | The document states that at some point there is a legal obligation for data collected with government funds to be freely available. | *** | http://cdiac.ornl.gov/programs/NARSTO/DM_develop_guide.pdf |
| | National Science Foundation (NSF) | Implementation of the NSF Data Sharing Policy. April 2002. | The document states that it is the responsibility of organizations to make results, data, derived data products, and collections available to researchers in a timely manner and at a reasonable cost. | ** | http://www.nsf.gov/geo/ear/EAR_data_policy_204.pdf |
| **Recommendations for policies** | | | | | |
| **Policies** | National Aeronautics and Space Administration (NASA) – Heliophysics Great Observatory | NASA Heliophysics Science Data Management Policy. 2007. | The paper recognizes that NASA observational data represent an asset that must be retained in a usable state into the indefinite future. | ** | http://hpde.gsfc.nasa.gov/Heliophysics_Data_Policy_2007June25.pdf |
| | National Institutes of Health (NIH) - National Heart, Lung, and Blood Institute (NHLBI) | Policy for Distribution of Data. Undated. | Data collected by NHLBI constitute an important scientific resource. Its full value can only be realized if it is made available (with the informed consent of individual participants) to the largest possible number of qualified investigators. The policy covers the responsibilities of investigators seeking access to data, the responsibilities of investigators in preparing data sets in response to requests, and procedures for protecting privacy for data sets. | ** | http://www.nhlbi.nih.gov/resources/deca/policy_new.htm |
| | National Oceanic and Atmospheric Administration (NOAA) | NOAA Administrative Order: 216-101. Ocean Data Acquisitions. 7/9/1990. | The order states that retrospective access to data is required by the research community through designated national data management centers. | ** | http://www.corporateservices.noaa.gov/~ames/NAOs/Chap_216/naos_216_101.html |
| **General guidance** | | | | | |
| **Specific guidance (e.g., how to interpret and use policies)** | NASA - Office of Space Science and Applications | Guidelines for Development of a Project Data Management Plan (PDMP). 1993. | The guidelines state that any agreements regarding exclusive rights to data should be stated, along with summary timelines for when the data will be released to the public. | *** | http://nssdc.gsfc.nasa.gov/nssdc/pdmp_guidelines_march93.rtf |
| | NASA - Jet Propulsion Laboratory | Cassini/Huygens Program Archive Plan for Science Data. 2004. | The document states that archives must be accessible to the public on-line. In addition, the office is responsible for filling large delivery orders to the science community, and making data available to foreign investigators, educators, and the general public. | ** | http://trs-new.jpl.nasa.gov/dspace/bitstream/2014/14261/1/00-0674.pdf |
| **Other (specify)** | DOE - Office of Scientific and Technical Information (OSTI) – Lawrence Livermore National Laboratory (LLNL) | The State of Data Management in the DOE Research and Development Complex. 7/14-15/2004. | The report briefly discusses issues such as data ownership and DOE rights of re-use that compound the problem of how to manage data. | ** | http://www.osti.gov/publications/2007/datameetingreport.pdf |

a. The applicability rating is shown as one, two, or three stars. A one-star rating (*) means that the information is related to ORD's SDM policy framework, but is expected to be of limited value in developing its policies and guidance. A two-star (**) rating means that the information is somewhat relevant to ORD's policy framework and/or is presented in limited detail. A three-star rating (***) means that the information provided is directly applicable to ORD's SDM policy/guidance goals and could serve as a model for ORD.

# 3. Develop a Scientific Data Management Plan that Covers the Full Data Life Cycle (Policy Area #2)

Developing a SDM plan provides the opportunity to focus on scientific data, including how the data will be gathered, processed, and analyzed. This plan can involve all stakeholders (e.g., users and potential users of the data) who can assess the value of the data for both current projects and potential future uses and reuses, even beyond the life of the projects. For ORD projects, a data management plan is sometimes developed as part of the quality assurance (QA) plan.

Development of this policy area will include a wide range of issues, including reviewing project strategy and planning to determine the general needs of the data throughout its life cycle, developing guidance on SDM planning for all projects, and tailoring planning for specific projects (which involves determining data needs and identifying users of the data). The contents of a data management plan often include the types of data to be authored, formatting standards, archiving and preservation provisions, metadata, and plans for transitioning or terminating the data. The end result may be a consolidated list of data products and tools that are needed to support the entire life cycle of the project. As with Policy Area #1, developing a SDM plan overlaps with several other policy areas.

## 3.1 EPA Policy Information

Table 4 describes 15 individual EPA documents and resources related to developing a SDM plan.[1] About one-half of these documents provide general guidance on issues such as SDM plan contents and successful data management strategies. One document provides a recommendation for policies, six documents are policies or specific guidance, and one document is classified as an "other" reference, which is a document containing links to several supporting one-star documents (i.e., of limited value to ORD). No supporting documentation was found for "Goals, Vision Statements" or "Principles." The research identified five examples of three-star documents that are directly relevant to EPA/ORD and 11 documents with two-star (i.e., somewhat relevant) and one-star (i.e., of limited value) ratings.

1. Note that some documents provide information at more than one policy level (e.g., a single document might provide policy recommendations and general guidance). Sections 4–8 report the number of individual documents (not the double-counted references) that provide information on each policy area.

| Table 4. Develop a scientific data management plan that covers the full data life cycle: EPA documents and resources (Policy Area #2) | | | | | |
|---|---|---|---|---|---|
| **Level** | **Agency name/office** | **Document title/date** | **Description/pertinent aspects** | **Applicability[a]** | **Link (or reference)** |
| **Goals, vision statements** | | | | | |
| **Principles** | | | | | |
| **Recommendations for policies** | ORD | Scientific Data Management Strategy. 2007. | The strategy states an objective to define an SDM organizational structure. The structure needs to be "tuned" to the specific needs of each L/C/O. | *** | EPA, 2007. |
| **Policies** | OEI | Information Resources Management (IRM) Policy, Chapter 19 Information and Data Management. 2001. | Section 5, Policies, of this document lists the U.S. Environmental Protection Agency (EPA) policies on information and data management. Note that this document has expired and has not yet been updated. | ** | Email communication with Lynne Petterson, 6/10/09. |
| | OEI | Records Management. 12/11/2009. | This policy states: "The Records Management Policy establishes principles, responsibilities, and requirements for managing EPA's records to ensure that the agency is in compliance with federal laws and regulations, EPA policies, and best practices for managing records. This Agency-wide policy provides the framework for specific guidance and detailed operating procedures governing records management organization and implementation." | * | http://www.epa.gov/records/policy/index.htm. |
| | OEI | National Geospatial Data Policy. CIO Policy Transmittal 05-002. 3/24/2008. | The policy establishes specific requirements for all EPA program offices and labs regarding the planning, collecting, acquiring, processing, documenting, storing, accessing, maintaining, and retiring of geospatial data. | *** | http://www.epa.gov/esd/ggc/pdf/epa_natl_geo_data_policy.pdf |
| **General guidance** | OSWER<br><br>Brownfields and Land Revitalization Technology Support Center | Management and Interpretation of Data Under a Triad Approach – Technology Bulletin. May 2007. | The triad approach produces flexible, but rigorous project plans; data management is key to rapid collection and analysis of data gathered. "A successful data management strategy depends on input not only from data management specialists but also from those who will be generating and using the data, including vendors, geoscientists, chemists, and other technical specialists. The data management plan must address how data from different sources will be integrated to support decisions." | * | http://www.brownfieldstsc.org/pdfs/Management_and_Interpretation_of_Data.pdf |
| | Western Regional Air Partnership (WRAP) | Comprehensive Data Management of WRAP Emissions Data. 2009. | This is a data management plan for emissions data that could be used as guidance for the creation of an ORD data management plan policy. (Note: The Western Governors' Association and the National Tribal Environmental Council receive funding from the U.S. EPA to administer and support the WRAP.) | * | http://www.epa.gov/ttn/chief/conference/ei18/session1/hoek.pdf |
| | OEI | EPA Enterprise Architecture Target Data Architecture. 2009. | In the framework presented in Section 4.3.1, EPA Program Offices that oversee Agency-wide business lines will ensure that quality-related activities associated with each phase of the EPA Data Lifecycle Framework (Figure 16) are documented. See also Appendix A. | *** | Email communication with Kevin Kirby. 7/14/09. |
| | OEI | Guidance for Geospatial Data Quality Assurance Project Plans. March 2003. | This guidance document describes the type of information that would be included in a QA Project Plan by anyone developing a geospatial project or using geospatial data for EPA. | ** | Email communication with Lynne Petterson, 6/10/09. |
| | OEI | Data Standards Policy. 6/28/2007. | This document states that: "All Agency information systems that exchange information shall implement applicable data standards in the most current version at the appropriate phase in the development life cycle but no later than the required implementation date specified in the standard unless a waiver has been obtained. When a new version of a standard is issued the old version is given a retirement date and should not be used after that date. Implementation of data standards or the appropriate waiver shall be described in the lifecycle and solution architecture documentation for each applicable EPA system and documented in the Registry of EPA Applications and Databases (READ) in conformance with the READ record maintenance schedule." | * | http://www.epa.gov/oamhpod1/adm_placement/ITS_BISS/datastd.pdf |
| | OSWER | System Life Cycle Management Guidance Part 3 Practice Paper: Configuration Management. January 1989. | Exhibit 2-1 provides an overview of configuration management throughout a system life cycle. This is more for a "system" life cycle than "documentation" life cycle, but might still have some relevance. | ** | http://www.epa.gov/oswer/docs/oswerlcm/00000019.pdf |
| | OSWER | System Life Cycle Management Guidance Part 3 Practice Paper: Data Management During the Life Cycle. January 1989. | In the data management paper, Chapter 3 provides a high level review of the recommended approach for each step of the system life cycle. | *** | http://www.epa.gov/oswer/oswerlcm.htm |

| Table 4. Develop a scientific data management plan that covers the full data life cycle: EPA documents and resources (Policy Area #2) | | | | | |
|---|---|---|---|---|---|
| **Level** | **Agency name/office** | **Document title/date** | **Description/pertinent aspects** | **Applicability[a]** | **Link (or reference)** |
| | OSWER | OSWER Life Cycle Management Guide. 1989 | This document provides suggestions for both the initiation and concept phase of information management by discussing initiation phase objectives, concept phase objects, decisions, activities, roles/responsibilities and the decision paper.<br><br>Pages 22 and 23 discuss the creation of a data management plan and what should be included. Page 26 discusses the data management plan and what should be included in the definition stage. Chapter 4 discusses the expansion of the data dictionary and data management plan. Chapter 10 details how all life cycle stages work together and/or overlap. | *** | http://www.epa.gov/oswer/oswerlcm.htm |
| **Specific guidance (e.g., how to interpret and use policies)** | OEI | Guidance on Systematic Planning Using the Data Quality Objectives Process. February 2006. | EPA has established a policy that states that before information or data are collected on Agency-funded or regulated environmental programs and projects, a systematic planning process must occur during which performance or acceptance criteria are developed for the collection, evaluation, or use of these data. This document provides specific guidance at each step of using the data quality objectives process. | ** | http://www.epa.gov/quality/qs-docs/g4-final.pdf |
| | OSWER | System Life Cycle Management Guidance Part 3 Practice Paper: Data Management During the Life Cycle. Janurary1989. | This paper describes data management during the system life cycle and provides guidance concerning major topics that should be addressed by project teams. Data management begins during the concept phase, proceeds as requirements are defined and software is implemented, and continues until the application system is terminated or replaced. The chapters include the following: Selecting a data management approach, overview of data management topics, data modeling activities, data design activities, data stewardship, data documentation activities, and terms/reference manual. This document provides a useful synopsis of many of the *System Life Cycle* chapters 1-10. | ** | http://www.epa.gov/oswer/oswerlcm.htm |
| | OSWER | System Life Cycle Reviews and Approvals. 1989. | This document provides all the steps and information necessary to review and approval all stages of the system life cycle. | ** | http://www.epa.gov/oswer/docs/oswerlcm/00000018.pdf |
| **Other** | OEI<br><br>Office of Technology Operations and Planning (OTOP) | IT Policy Mega-Matrix. 2009. | The IT Policy Mega-Matrix is a master list of all the EPA IT policy documents (e.g., Policies, Procedures, Standards, and Guidance) that OTOP maintains. Page 12 contains SLC documents. | * | This document is located on the EPA intranet at: http://intranet.epa.gov/otop/itpolicy/IT_Policy_Mega-Matrix_Feb2009_external.pdf |

a. The applicability rating is shown as one, two, or three stars. A one-star rating (*) means that the information is related to ORD's SDM policy framework, but is expected to be of limited value in developing its policies and guidance. A two-star (**) rating means that the information is somewhat relevant to ORD's policy framework and/or is presented in limited detail. A three-star rating (***) means that the information provided is directly applicable to ORD's SDM policy/guidance goals and could serve as a model for ORD.

## 3.2 Other Federal Agency Policy Information

Table 5 summarizes more than 20 documents and resources developed by non-EPA federal agencies on developing a SDM plan. Five agencies have policies or guidance documents that stress the importance of developing SDM plans that cover the full data life cycle. Three of the documents have three-star ratings because they can serve as models for ORD – a NOAA document, *Environmental Data Management at NOAA*, which defines and describes principles that recognize the importance of SDM planning, and two NSF documents that provide SDM policies related to data submission and archiving. The remaining documents are more general and less relevant to ORD, but do provide some useful information about SDM planning.

# 4. Identify Scientific Data with Metadata to Enable Needed Business Operations (Policy Area #3)

To gain maximum value from scientific data, it must be easily accessed and understood by those who use it. The information that provides this understanding is "data about scientific data" – metadata. Metadata can provide "provenance" or data lineage (e.g., by linking to information products such as the data management plan or final report) and can also enable data discovery, retrieval, and appropriate reuse. Metadata is essential for identifying, searching for, locating, storing and retrieving scientific data. The consistent development and use of metadata enables communication between cooperating agencies and the public users of data, and can be used to identify appropriate data users and help provide access control.

Metadata is a complicated and expansive topic that may extend beyond ORD's needs to manage scientific data. Policies and guidance developed under this policy area can include ensuring that appropriate metadata is selected to help manage data creation and retention, developing guidance on how to use approved standards and tools to create metadata, developing guidance on the minimum essential metadata required, and reviewing information on who should be involved in metadata development.

## 4.1 EPA Policy Information

Table 6 identifies 16 individual documents and resources that provide principles, policies, and guidance on developing metadata to support SDM. The majority of these resources are general and specific guidance documents. The most relevant resources include an OSWER principle that describes the importance of keeping accurate metadata when managing data, OEI's *Enterprise Architecture Target Data Architecture* report, which addresses metadata management in the context of data management, and the OSWER *Life Cycle Management Guide*, which provides guidance on developing metadata during the design phase of data management.

| Table 5. Develop a scientific data management plan that covers the full data life cycle: Other federal agency documents and resources (Policy Area #2) | | | | | |
|---|---|---|---|---|---|
| **Level** | **Agency name/office** | **Document title/date** | **Description/pertinent aspects** | **Applicability[a]** | **Link (or reference)** |
| **Goals, vision statements** | NOAA | NOAA Report to Congress on Data and Information Management. 2005. | According to the report, NOAA is in the initial stages of developing and implementing an integrated data management system. | * | http://www.ngdc.noaa.gov/noaa_pubs/pdf/NOAA_Congress2005.pdf |
| | DOE | The State of Data Management in the DOE Research and Development Complex. 7/14-15/2004. | The report states that DOE needs a department-wide policy that recognizes life-cycle data management. It recommends an umbrella policy for data generators, collectors, curators, and users. | ** | http://www.osti.gov/publications/2007/datameetingreport.pdf |
| **Principles** | NOAA | Environmental Data Management at NOAA: Archiving, Stewardship, and Access. 2007. | Principle #7 states that effective data management requires a formal, ongoing planning process. NOAA should establish and codify an enterprise-wide data management plan (elements of plan listed on p 87-88).<br><br>Principle #2 states that data-generating activities should include adequate resources to support end-to-end data management. | *** | National Research Council, 2007. |
| | Climate Change Science Program | Strategic Plan for the Climate Change Science Program Final Report: Chapter 13. Data Management and Information. July 2003. | The report states that data managers must be able to understand, communicate, and work closely with scientists and others to ensure proper stewardship for the data archive and its distribution. | ** | http://www.climatescience.gov/Library/stratplan2003/final/ccspstratplan2003-chap13.htm |
| | National Science and Technology Council (Office of Science and Technology Policy) | Harnessing the Power of Digital Data for Science and Society. 2009. | The report discusses the importance of cooperation among industry, academia, nongovernmental organizations (NGOs), and international agencies (p. 19). | ** | http://www.nitrd.gov/about/harnessing_power_web.pdf |
| **Recommendations for policies** | | | | | |
| **Policies** | NIH – Division of Acquired Immunodeficiency Syndrome (DAIDS): Clinical Research Policies and Standard Procedures Documents | Requirements for Data Management and Statistics for DAIDS Funded and/or Sponsored Clinical Trials. 2007. | This policy states that clinical trial data need to be managed in such a way as to ensure the authenticity and integrity of the data elements collected and to comply with applicable regulations. | ** | http://www3.niaid.nih.gov/LabsAndResources/resources/DAIDSClinRsrch/PDF/DataMgt_StatPolicy.htm |
| | NSF – Division of Ocean Sciences (DOS) | Division of Oceans: Data and Sample Policy. 11/3/2003. | The policy states that programs may establish more stringent data submission procedures to meet the needs of these programs. Principal investigators supported by these programs are required to follow these data submission procedures. | *** | http://www.nsf.gov/pubs/2004/nsf04004/nsf04004_1b.htm |
| | NSF – Social, Behavioral and Economic Sciences (SES) | Data Archiving Policy. 7/8/2008. | This policy recognizes that many complexities arise across the range of data collection supported by SES programs, and that unusual circumstances may require modifications or even full exemptions. For example, human subjects protection requires removing identifiers, which may be prohibitively expensive or render the data meaningless in research that relies heavily on extensive in-depth interviews. | *** | http://www.nsf.gov/sbe/ses/common/archive.jsp |
| | NOAA | NOAA Administrative Order: 216-101. Ocean Data Acquisitions. 7/9/1990. | The order defines responsibilities and procedures for all NOAA activities that involve the collection and archiving of ocean data from the open-ocean, Great Lakes, coastal waters, and estuaries. | * | http://www.corporateservices.noaa.gov/~ames/NAOs/Chap_216/naos_216_101.html |
| | NOAA | NOAA Administrative Order: 212-15. Management of Environmental and Geospatial Data and Information. 2008. | The order states that the NOAA CIO must develop a data management plan in coordination with the appropriate data center, specifying the data life cycle and disposition of data and information for each program. | ** | http://www.corporateservices.noaa.gov/~ames/NAOs/Chap_212/naos_212_15.html |
| **General guidance** | NASA – Heliophysics Great Observatory | NASA Heliophysics Science Data Management Policy. 2007. | The policy document provides a blueprint for a data management plan, tracing the data lifecycle from measurements to final archives, and provides examples of information appropriate for each data provider to include in a data management plan (p. 23). | ** | http://hpde.gsfc.nasa.gov/Heliophysics_Data_Policy_2007June25.pdf |
| | NASA – Consultative Committee for Space Data Systems (CCSDS) | Reference Model for an Open Archival Information System (OAIS). 2002. | The document shows a data flow diagram that represents the operational OAIS archive external data flows. The diagram shows the flow of information among producers, consumers, and the OAIS (but does not include flows that involve management). | ** | http://public.ccsds.org/publications/archive/650x0b1.pdf |

| Table 5. Develop a scientific data management plan that covers the full data life cycle: Other federal agency documents and resources (Policy Area #2) | | | | | |
|---|---|---|---|---|---|
| **Level** | **Agency name/office** | **Document title/date** | **Description/pertinent aspects** | **Applicability[a]** | **Link (or reference)** |
| | DOE – ORNL | Developing Data Management Policy and Guidance Documents for your NARSTO Program or Project. 2005. | The guidelines provides a data flow chart before, during, and after a field campaign. It mentions that a clear statement of the importance of the data collection and the flow of the data in the broadest possible context is needed. In addition, advanced planning for archiving project data furthers efforts to identify, collect, and report consistent data and metadata and to facilitate timely data analysis, sharing, integration, and synthesis. | ** | http://cdiac.ornl.gov/programs/NARSTO/DM_develop_guide.pdf |
| | DOE – ORNL | Guidelines for Archiving Data in the NARSTO Permanent Data Archive. May2, 2006. | The document provides characteristics of a project data management plan that will result in successful data archiving. | ** | http://cdiac.ornl.gov/programs/NARSTO/Guidelines_for_Archiving_NARSTO_Data.pdf |
| | NSF | Long-Lived Digital Data Collections: Enabling Research and Education in the 21st Century. 2005. | The contents of the data management plan should include: the types of data to be authored; the standards that would be applied for format, metadata content, etc.; provisions for archiving and preservation; access policies and provisions; and plans for eventual transition or termination of the data collection in the long-term future. | ** | http://www.nsf.gov/pubs/2005/nsb0540/ |
| | NSF – Division of Earth Sciences (EAR) | Implementation of the NSF Data Sharing Policy. April 2002. | The policy mentions that compliance with stated data management guidelines will be considered in the Program Officer's overall evaluation during the proposal review process. | * | http://www.nsf.gov/geo/ear/EAR_data_policy_204.pdf |
| | NOAA | NOAA Administrative Order: 212-15. Management of Environmental and Geospatial Data and Information. 2008. | The document states that NOAA data management planning will include end-to-end data stewardship. | ** | http://www.corporateservices.noaa.gov/~ames/NAOs/Chap_212/naos_212_15.html |
| | NIH – Office of Extramural Research | NIH Data Sharing Policy and Implementation Guide. 3.5.2003. | The policy states that the content and level of detail included in a data-sharing plan depends on several factors, such as whether or not the investigator is planning to share data, and the size and complexity of the dataset. | * | http://grants.nih.gov/grants/policy/data_sharing/ |
| | NIH – National Cancer Institute (NCI) | National Cancer Institute, Division of Cancer Prevention (DCP), Data Management Requirements. October 2003. | The document states that a data management plan is prepared by the Consortium Principal Investigator and approved by the NCI and DCP. | * | http://prevention.cancer.gov/clinicaltrials/management/consortia/step-2/data |
| | NIH – DAIDS: Clinical Research Policies and Standard Procedures Documents | Requirements for Data Management and Statistics for DAIDS Funded and/or Sponsored Clinical Trials. 2007. | The document describes the processes and methods that data collection sites and central data management facilities must develop to manage their data, including: data management operations, overall data management system, data storage, database closure and archiving, and data audits. | ** | http://www3.niaid.nih.gov/LabsAndResources/resources/DAIDSClinRsrch/DataManagement.htm |
| **Specific guidance (e.g., how to interpret and use policies)** | NASA – Office of Space Science and Applications | Guidelines for Development of a Project Data Management Plan. 1993. | The guidelines mention that project data flow should be provided, including an overall functional data flow diagram. The diagram should identify those facilities performing various functions as the project progresses through its various mission phases. | * | http://nssdc.gsfc.nasa.gov/nssdc/pdmp_guidelines_march93.rtf |
| | NASA – National Space Science Data Center | White Paper on NASA Science Data Retention. 2007. | The paper states that policies must ensure the continuing preservation, accessibility, and usability of the data in their care. Plans for doing so should be spelled out in Archives' Operating Plans. | * | http://nssdc.gsfc.nasa.gov/nssdc/data_retention.html |
| | DOE – ORNL | NARSTO Quality Systems Management Plan. 9/30/1999. | The document provides a project plan and data archival process flow chart. | ** | http://cdiac.ornl.gov/programs/NARSTO/pdf/qsmp_current_version.PD |
| | NIH – National Institute on Aging | Guidelines for Developing a Manual of Operations and Procedures (MOP). 2007. | Guidelines for program investigators of multi-site clinical trials to follow when preparing MOPs. MOPs are intended to facilitate consistency in protocol implementation and data collection, and are prepared before the study begins. The guidelines most relevant to ORD include data flow (e.g., data flow, data entry, data correction), data retention, data management, study completion and closeout procedures, and confidentiality. | ** | http://www.nia.nih.gov/NR/rdonlyres/AEC5CE46-96E1-43D9-BA77-BAE8BF0D6CDC/0/ManualofProceduresMOPFinal1.doc |
| **Other** | NASA – Jet Propulsion Laboratory | Cassini/Huygens Program Archive Plan for Science Data. 2004. | The document states that archive policies, guidelines, and requirements have been developed to ensure data products meet standards and support collaborative studies among Cassini Orbiter and Huygens Probe data. | * | http://trs-new.jpl.nasa.gov/dspace/bitstream/2014/14261/1/00-0674.pdf |
| | National Science and Technology Council (Office of Science and Technology Policy) | Harnessing the Power of Digital Data for Science and Society. 2009. | The document provides a full description of the data life cycle, which includes creation, ingestion or acquisition, documentation, organization, migration, protection, access, and disposition. | ** | http://www.nitrd.gov/about/harnessing_power_web.pdf |

a. The applicability rating is shown as one, two, or three stars. A one-star rating (*) means that the information is related to ORD's SDM policy framework, but is expected to be of limited value in developing its policies and guidance. A two-star (**) rating means that the information is somewhat relevant to ORD's policy framework and/or is presented in limited detail. A three-star rating (***) means that the information provided is directly applicable to ORD's SDM policy/guidance goals and could serve as a model for ORD.

| Table 6. Identify scientific data with metadata to enable needed business operations: EPA documents and resources (Policy Area #3) | | | | | |
|---|---|---|---|---|---|
| **Level** | **Agency name/office** | **Document title/date** | **Description/pertinent aspects** | **Applicability[a]** | **Link (or reference)** |
| **Goals, vision statements** | | | | | |
| **Principles** | Office of Science Advisor | Assessment Factors. June 2003. | Section 2.2.3 e of this document asks: Is the complete data set accessible, including metadata, data-dictionaries and embedded definitions (e.g., codes for missing values, data quality flags and questionnaire responses)? Are there confidentiality issues that may limit accessibility to the complete data set? | * | http://www.epa.gov/OSA/spc/pdfs/assess2.pdf |
| | OSWER | System Life Cycle Management Guidance Part 3 Practice Paper: Data Management During the Life Cycle. January 1989. | This document states the principle that accurate information about data is essential. Effective management of data collected by OSWER requires that accurate information about data (i.e., metadata) be kept. | *** | http://www.epa.gov/oswer/docs/oswerlcm/00000021.pdf |
| **Recommendations for policies** | | | | | Email communication with Lynne Petterson, 6/10/09. |
| **Policies** | OEI | Data Standards Policy. 6/28/2007. | This Data Standards Policy establishes principles, responsibilities, and requirements for the development, maintenance, and implementation of data standards within the jurisdiction of the U.S. Environmental Protection Agency. This policy discusses the use of common terminology and data elements for consistency and data sharing; the use of centralized registries of data elements, XML schema and code sets, based on approved data standards, and related roles and responsibilities. | ** | http://www.epa.gov/oamhpod1/adm_placement/ITS_BISS/datastd.pdf |
| **General guidance** | Office of Air and Radiation<br><br>Emissions, Monitoring, and Analysis Division | Annual Air Quality Data Certifications for PM and Ozone Design Values. 6/12/2002. | This memo requires states and Tribes to document their annual air quality data sets so EPA can accurately interpret the reported data. States and Tribes must certify that prior year data are entered and the summary report is accurate. | * | http://www.epa.gov/ttn/amtic/files/ambient/pm25/datamang/designmem.pdf |
| | Office of Air and Radiation<br><br>Emission Inventory Improvement Program (EIIP) Data Management Committee (DMC) | EIIP Phase I Data Model. 1999. | This document describes four views of the EIIP Data Model that provide common formats so data can be shared. It also provides a thorough data element dictionary, a list of entities and their attributes, and data model codes. | * | http://www.epa.gov/ttn/chief/eiip/techreport/volume07/vii01.pdf |
| | OEI | EPA Enterprise Architecture Target Data Architecture. 6/23/2009. | Section 4 examines the various components of data management that are critical at the enterprise level and must be addressed for enterprise architecture. Topics in this section address data quality, enterprise data security, metadata and master data management and data governance. | *** | Email communication with Kevin Kirby. 7/14/09. |
| | OEI | EPA Enterprise Architecture Target Data Architecture. 6/23/2009. | Enterprise Metadata Architecture, Section 4.6-4.6. The enterprise metadata architecture proposed for EPA is a cross-cutting framework of policy, standards, communication, implementation, and continual evaluation required for enabling a consistent metadata capability. This document includes information on Metadata Standards and Policy Development (Section 4.6.1), Governance for Data and Metadata (Section 4.6.2), Communication and Outreach (Section 4.6.3), Implementation Assistance (Section 4.6.4), Lessons Learned and Performance Measures (Section 4.6.5). See also Appendix C. | *** | Email communication with Kevin Kirby. 7/14/09. |

| Table 6. Identify scientific data with metadata to enable needed business operations: EPA documents and resources (Policy Area #3) | | | | | |
|---|---|---|---|---|---|
| **Level** | **Agency name/office** | **Document title/date** | **Description/pertinent aspects** | **Applicability[a]** | **Link (or reference)** |
| | OEI | Metadata Standards for the Enterprise Content Management Program. Last updated 7/2/09. | The purpose of these standards is to define a consistent set of required metadata elements for all applications participating in the enterprise content management program. These standards cover unstructured information, which includes but is not limited to documents and records, and applies to all EPA Programs, Regions, Labs and Offices. Specifically, these standards underscore the importance of a consistent, yet somewhat flexible, set of metadata elements for the effective and accurate classification, retrieval, management and use of unstructured information. This document provides examples of baseline metadata standards and some associated roles and responsibilities. | *** | Email communication with Lynne Petterson, 6/10/09. |
| | Office of Research and Development (ORD) | Implementing the National Geospatial Data Policy: Lessons Learned. 2009. | This document provides lessons learned on data management policy implementation. It identifies weaknesses related to the National Geospatial Data Policy, including: metadata, infrastructure (e.g., network and systems interoperability regarding metadata and data load), and data management [e.g., need to develop a process through which project data will be cataloged and disseminated through Environmental Information Management System (EIMS)]. | ** | http://intranet.epa.gov/ospintra/Science%20Council/Related%20Docs/ORDNGDPPILOTS.pdf |
| | OSWER | OSWER Life Cycle Management Guide. 1989. | In Chapter 3 of the OSWER Life Cycle Management Guide, Page 21, Exhibit 3-10 discusses the Requirements Data Dictionary and how it serves as a repository for metadata.<br><br>Chapter 4 describes that, in the design phase, it is up to the user to enter metadata in the design data dictionary to document the physical design of each data base or data file. | *** | http://www.epa.gov/oswer/oswerlcm.htm |
| | EPA Region 9<br>Tribal Water Protection | National Tribal WQX/STORET Data Management. 2008. | This is a presentation on how to apply metadata to data for sharing purposes, emphasizing consistency. The examples given are for the Water Quality Exchange (WQX) and STORET and may not carry over to other projects. | * | http://www.epa.gov/region09/water/tribal/storet-training/pdf/WQXTemplate.pdf |
| **Specific guidance (e.g., how to interpret and use policies)** | Great Lakes National Program Office | Lake Michigan Mass Balance Metadata. 3/9/2006. | The Metadata link offers some guidance on metadata reporting formats and sample naming. | ** | http://www.epa.gov/greatlakes/lmmb/metadata.html |
| | OEI | National Geospatial Data Policy. Procedure for Geospatial Metadata Management. 10/25/2007. | Geospatial Data Stewards must create or update the metadata record for each acquired data set so that it meets the minimum requirements of the EPA Metadata Technical Specification. During the data storage and access phase, stewards must refer to the technical specification for data storage and access requirements. Maintenance responsibility for geospatial and metadata falls to the data owner or data steward of the program office or division. | ** | http://www.epa.gov/geospatial/docs/2131.pdf |
| | OEI | Data Standards Implementation. 6/28/2007. | This document contains procedures establishing the key steps to follow for implementation of EPA data standards. It discusses procedures for the following areas: development of implementation guidance for a data standard, review/approval of implementation guidance for a data standard, conformance assistance, and conformance measurement. | ** | http://www.epa.gov/irmpoli8/policies/2133p3.pdf |
| | OEI | Data Standards Maintenance. 6/28/2007. | This document contains procedures establishing the key steps to follow for maintenance and revision of EPA data standards and implementation guidance. It discusses procedures for the following areas: proposal for data standard and/or implementation guidance revision, development of minor and major data standard revisions, data standards review and approval procedures for major revisions. | ** | http://www.epa.gov/irmpoli8/policies/2133p2.pdf |

| Table 6. Identify scientific data with metadata to enable needed business operations: EPA documents and resources (Policy Area #3) | | | | | |
|---|---|---|---|---|---|
| **Level** | **Agency name/office** | **Document title/date** | **Description/pertinent aspects** | **Applicability[a]** | **Link (or reference)** |
| | OEI | Requesting Data Standards Conformance Waiver. 6/28/2007. | This document contains procedures establishing the key steps to follow for requesting a data standard conformance waiver from EPA data standards. It discusses procedures for the following areas: types of waivers; determination of need; and submission, disposition and posting of a waiver. | ** | http://www.epa.gov/irmpoli8/policies/2133p4.pdf |
| | OEI | Data Standards Development. 6/28/2007. | These procedures establish the key steps to follow for development and approval of EPA data standards. This document provides procedures for the following: Data standard proposal, development, and approval and draft data standards review. | * | http://www.epa.gov/irmpoli8/policies/2133p1.pdf |
| | OSWER | System Life Cycle Management Guidance Part 3 Practice Paper: Data Modeling. May 1992. | This is a detailed document that includes topics such as: What are data models, creating data entities, data relationships and creating relationships between data entities, creating data elements, and changing the model. "This paper (1) introduces data modeling techniques; (2) defines specific data standards for logical data modeling to follow during the SLC; and (3) offers some "how to" guidance throughout the data modeling process." | *** | http://www.epa.gov/oswer/docs/oswerlcm/00000022.pdf |
| **Other** | | | | | |

a. The applicability rating is shown as one, two, or three stars. A one-star rating (*) means that the information is related to ORD's SDM policy framework, but is expected to be of limited value in developing its policies and guidance. A two-star (**) rating means that the information is somewhat relevant to ORD's policy framework and/or is presented in limited detail. A three-star rating (***) means that the information provided is directly applicable to ORD's SDM policy/guidance goals and could serve as a model for ORD.

## 4.2 Other Federal Agency Policy Information

A significant amount of information on the development of metadata was found among non-EPA federal agencies. As Table 7 shows, more than 20 principles, policy recommendations, policies, guidance, and other documents related to metadata and SDM were identified. Three agencies – DOE, the NSF, and NOAA – have published a total of four documents that contain principles regarding identifying scientific data with metadata. All five federal agencies have developed policy recommendations (six documents), policies (six documents) and general guidance (eight documents) related to metadata, including three three-star-rated documents (the *NIH Data Sharing Policy and Implementation Guide*, the NSF's *Division of Ocean Sciences: Data and Sample Policy*, and the National Science and Technology Council's *Harnessing the Power of Digital Data for Science and Society*). Five specific guidance documents were identified, two of which are particularly relevant to ORD (i.e., NASA's *Heliophysics Science Data Management Policy* and the *Strategic Plan for the Climate Change Science Program Final Report*).

# 5. Manage Scientific Data for Appropriate Control (Policy Area #4)

Scientific data can be developed under agreements such as contracts, grants, and partnerships. In these cases, rights to the data and its reuse may be specified in the agreements, and data must be managed to comply with these provisions. It is also important to provide credit to the data creators. Often referred to as IP, this important policy area describes intangible assets that allow the rights of the data provider to be recognized in order to comply with legal obligations. The development of this policy area will create specific metadata and data management requirements, which may be documented in the data management plan and procedures.

Development of this policy area will include, among other things, guidance on understanding data rights and circumstances (e.g., proprietary data) that create different types of data rights, policies to establish and maintain an identification process for IP, and guidance on establishing levels of control and how to select the appropriate level of control for a data set given specific data rights. This policy area is related to other policy areas – for example, access controls may be part of the SDM planning process in Section 3, *Develop a Scientific Data Management Plan that Covers the Full Data Life Cycle*, and will create metadata requirements (from Section 4, *Identify Scientific Data with Metadata to Enable Needed Business Operations*). This area may also affect "embargoed" data that cannot be released outside the project until the final project deliverables are released (see Section 6, *Maintain Version and Change Control on Data Sets*).

| Table 7. Identify scientific data with metadata to enable needed business operations: Other federal agency documents and resources (Policy Area #3) | | | | | |
|---|---|---|---|---|---|
| **Level** | **Agency name/office** | **Document title/date** | **Description/pertinent aspects** | **Applicability[a]** | **Link (or reference)** |
| **Goals, vision statements** | | | | | |
| **Principles** | NSF – Office of Polar Programs (OPP) | Guidelines and Award Conditions for Scientific Data. 1998. | The OPP considers the documentation of data sets (metadata) as vital to the exchange of information on polar research and to a data set's accessibility and longevity for reuse. | * | http://www.nsf.gov/pubs/1999/opp991/opp991.doc |
| | DOE – OSTI | The State of Data Management in the DOE Research and Development Complex. 7/14-15/2004. | The report states that metadata must be optimized for future retrieval, assimilation, and re-use. A professional staff of scientists is needed to manage data. | ** | http://www.osti.gov/publications/2007/datameetingreport.pdf |
| | DOE – ORNL | Developing Data Management Policy and Guidance Documents for your NARSTO Program or Project. 2005. | The guidance states that there must be a decision on whether investigators have an obligation to make data easy to use by others. | ** | http://cdiac.ornl.gov/programs/NARSTO/DM_develop_guide.pdf |
| | NOAA | Environmental Data Management at NOAA: Archiving, Stewardship, and Access. 2007. | Principle #5 states that metadata are essential for scientific data management.<br><br>Principle #6 states that scientific data stewardship, with assigned organizational responsibility, should be applied to all environmental data sets and their associated metadata to ensure that this information is preserved, remains continually accessible and can be improved as future discoveries build understanding and knowledge.<br><br>Principle #8 states that an effective data archive should provide for discovery, access, and integration. | ** | National Research Council, 2007. |
| **Recommendations for policies** | NIH – Office of Extramural Research | NIH Data Sharing Policy and Implementation Guide. 3/5/2003. | The policy mentions that regardless of the mechanism used to share data, each dataset will require documentation.<br><br>The policy also states that data sharing promotes many goals of the NIH research endeavor. This is particularly important for unique data that cannot be readily replicated. | *** | http://grants.nih.gov/grants/policy/data_sharing/ |
| | NSF – Office of Polar Programs | Guidelines and Award Conditions for Scientific Data. 1998. | The guidelines recommend that data archives of OPP-supported projects should include easily accessible information about the data holdings, including quality assessments, supporting ancillary information, and guidance for locating and obtaining the data. | ** | http://www.nsf.gov/pubs/1999/opp991/opp991.doc |
| | NSF – Social, Behavioral and Economic Sciences | Data Archiving Policy. 7/8/2008. | The policy recommends that if it is appropriate for other researchers to have access to data, the investigators should specify a time at which they will be made generally available, in an appropriate form and at a reasonable cost. | ** | http://www.nsf.gov/sbe/ses/common/archive.jsp |
| | NOAA | Environmental Data Management at NOAA: Archiving, Stewardship, and Access. 2007. | This book states that:<br><br>(1) Guidelines are needed on stewardship and the need for systematic, ongoing assessment and improvement of data. Stewardship plans should be consistent but flexible so improvements in data and metadata are captured. (2) Guidelines are needed on making data available to users in a timely manner and accessible with as few barriers as possible (administrative, technological, and systematic barriers are described). (3) Environmental data should be easily discoverable by a broad range of users. Data discovery should not require any specific knowledge about the data or how they are managed. (4) A distributed data access structure can support improved data discovery and seamless integration. (5) Metadata that adequately documents and describes each archived data set should be created and preserved to ensure the enhancement of knowledge. (6) Search tools and other discovery-enhancing features could be improved at many environmental data access points by the use of expanded metadata (detailed list provided on pp. 75-76).<br><br>It further recommends that: (1) NOAA policies establish and maintain data and metadata migration plans for all current and future long-term archive systems to adapt to information technology evolution. (2) NOAA and partners should continue to expand usage of standards and reference models. | ** | National Research Council, 2007. |

| Table 7. Identify scientific data with metadata to enable needed business operations: Other federal agency documents and resources (Policy Area #3) | | | | | |
|---|---|---|---|---|---|
| **Level** | **Agency name/office** | **Document title/date** | **Description/pertinent aspects** | **Applicability[a]** | **Link (or reference)** |
| | Government Accountability Office (GAO) | Climate Change Research: Agencies Have Data-Sharing Policies but Could Do More to Enhance the Availability of Data from Federally Funded Research. 2007. | The guidelines recommend that NOAA evaluate whether additional strategies are warranted to facilitate the permanent archiving of relevant data, which may include: leveraging existing resources and devoting a greater portion of data collection funds to archiving activities. | * | http://www.gao.gov/new.items/d071172.pdf |
| **Policies** | NIH – NHBLI | Policy for Distribution of Data. Undated. | The policy states that documentation for data sets must be comprehensive and sufficiently clear to enable investigators who are not familiar with a data set to use it. The documentation must include data collection forms, study procedures and protocols, descriptions of all variable recoding performed, and a list of major study publications. | ** | http://www.nhlbi.nih.gov/resources/deca/policy_new.htm |
| | NIH – DAIDS: Clinical Research Policies and Standard Procedures Documents | Requirements for Data Management and Statistics for DAIDS Funded and/or Sponsored Clinical Trials. 2007. | The requirements recommend that clinical trial data need to be managed in such a way as to ensure the authenticity and integrity of the data elements collected and to comply with applicable regulations. | ** | http://www3.niaid.nih.gov/LabsAndResources/resources/DAIDSClinRsrch/DataManagement.htm |
| | DOE – Atmospheric Radiation Measurement (ARM) | Atmospheric Radiation Measurement Data Sharing and Distribution Policy. 2006. | All data sets acquired during an Intensive Operational Period (IOP) or campaign will be made available to the ARM External Data Center for dissemination to users and forwarding to the ARM Archive. | * | http://www.arm.gov/data/docs/policy |
| | NSF – Division of Ocean Sciences | Division of Ocean Sciences: Data and Sample Policy. 11/3/2003. | The policy recommends that annual reports, required for all projects, should address progress on data and research product sharing.<br><br>The policy also states that where no data or sample repository exists for the collected data or samples, metadata must be prepared and made available. The principal investigator is required to address alternative strategies for complying with the general philosophy of sharing research products and data. | *** | http://www.nsf.gov/pubs/2004/nsf04004/nsf04004_1b.htm |
| | NOAA | NOAA Report to Congress on Data and Information Management 2005. | The report recommends that integration and interoperability should be achieved through common protocols, hardware, and software, as well as the use of data and metadata standards. NOAA has begun this process by adopting a common enterprise-wide IT architecture. | ** | http://www.ngdc.noaa.gov/noaa_pubs/pdf/NOAA_Congress2005.pdf |
| | NOAA | NOAA Administrative Order: 216-101: Ocean Data Acquisitions. 1990. | The order states that NOAA managers of programs that conduct ocean data collection activities are responsible for assuring that data and related information with high utility for other users are available in a timely manner at national processing centers and national data centers, and are documented and archived in designated national data management centers. | * | http://www.corporateservices.noaa.gov/~ames/NAOs/Chap_216/naos_216_101.html |
| **General guidance** | NASA – Office of Space Science and Applications | Guidelines for Development of a Project Data Management Plan (PDMP). 1993. | The guidelines state that a section of the PDMP should identify and describe all data sets expected to be generated. This includes the science data itself, associated ancillary data, and orbit/attitude data of the spacecraft. | ** | http://nssdc.gsfc.nasa.gov/nssdc/pdmp_guidelines_march93.rtf |
| | NASA – National Space Science Data Center | White Paper on NASA Science Data Retention. 2007. | Projects must create and certify optimally standards-adherent definitive data sets, and accompanying material (documentation, ancillary data, software, etc.) as needed to make the data independently usable. | ** | http://nssdc.gsfc.nasa.gov/nssdc/data_retention.html |
| | NASA – Jet Propulsion Laboratory | Cassini/Huygens Program Archive Plan for Science Data. 2004. | The policy describes that labels and index files provide searchable keys and describe characteristics of the products. Index files are used to populate the search catalog.<br><br>The document states that the Planetary Data System (PDS) Discipline Node assigned to an instrument team coordinates and leads a peer review of a sample volume. Members will be asked to participate in peer reviews as well as members of the science community outside the program. The peer review is used to ensure that the archive contains all the components needed to perform science analysis, and is prepared as documented in the Software Interface Specification. | * | http://trs-new.jpl.nasa.gov/dspace/bitstream/2014/14261/1/00-0674.pdf |

| Table 7. Identify scientific data with metadata to enable needed business operations: Other federal agency documents and resources (Policy Area #3) | | | | | |
|---|---|---|---|---|---|
| **Level** | **Agency name/office** | **Document title/date** | **Description/pertinent aspects** | **Applicability[a]** | **Link (or reference)** |
| | NIH – Office of Extramural Research | NIH Data Sharing Policy and Implementation Guide. 3/5/2003. | According to the policy, final research data are recorded factual material commonly accepted in the scientific community as necessary to document, support, and validate research findings. This does not mean summary statistics or tables; rather, it means the data on which summary statistics and tables are based. | ** | http://grants.nih.gov/grants/policy/data_sharing/ |
| | NIH – NCI | National Cancer Institute (NCI), Division of Cancer Prevention (DCP) Data Management Requirements. October 2003. | The data management plan should document the rules for handling data ranges, data types, and coding of missing data. | ** | ftp://narsto.esd.ornl.gov/pub/DES_metadata/var_names_web_sources/NARSTO_template_atmospheric_measurements.xls |
| | National Science and Technology Council | Harnessing the Power of Digital Data for Science and Society. 2009. | The report provides examples of data management mechanisms that include: continued improvement in interoperability across all layers (from software to hardware to networks and resources); comprehensive, global, and transparent search, query, and retrieval capabilities; development, continuing evolution, broad adoption, and regular use of appropriate, community based, cost-effective standards designed to allow efficient information use in innovative ways and in complex combinations; and promotion of ready access to appropriate documentation and metadata. | *** | http://www.nitrd.gov/about/harnessing_power_web.pdf |
| | DOE – ORNL | Developing Data Management Policy and Guidance Documents for your NARSTO Program or Project. 2005. | The document states that metadata should clearly state the source of data, and whether data are preliminary and for use only among the project or suitable for widespread dissemination and citation requirements. | ** | http://cdiac.ornl.gov/programs/NARSTO/DM_develop_guide.pdf |
| | NOAA | NOAA Administrative Order: 216-101: Ocean Data Acquisitions. 7/9/1990. | The order states that data submitted to the national data management centers are to be submitted via computer-compatible digital media when possible rather than as printed reports. Documentation must include information sufficient to fully describe the physical recording technique, data format, recording mode, blocking factor, and other pertinent items. | * | http://www.corporateservices.noaa.gov/~ames/NAOs/Chap_216/naos_216_101.html |
| **Specific guidance (e.g., how to interpret and use policies)** | NASA Heliophysics Great Observatory | NASA Heliophysics Science Data Management Policy. 2007. | The document states that the Heliophysics Data Environment (HPDE) will benefit greatly from more conventional standards, but experience has shown that if these are imposed by bodies without community input they tend to be ignored. | *** | http://hpde.gsfc.nasa.gov/Heliophysics_Data_Policy_2007June25.pdf |
| | DOE – ORNL | Developing Data Management Policy and Guidance Documents for your NARSTO Program or Project. 2005. | ORNL uses a web-based inventory of project data using the existing ORNL metadata search and data retrieval system called Mercury. | * | http://cdiac.ornl.gov/programs/NARSTO/DM_develop_guide.pdf |
| | DOE – ORNL | The NARSTO Atmospheric Measurements Template. 4/29/2005. | NARSTO provides a Data Exchange Standard (DES) template that is designed to help data originators create DES files. The worksheet titled "Detailed Metadata" contains a possible layout and content of a companion detailed metadata document. | ** | ftp://narsto.esd.ornl.gov/pub/DES_metadata/var_names_web_sources/NARSTO_template_atmospheric_measurements.xls |
| | Climate Change Science Program | Strategic Plan for the Climate Change Science Program Final Report. 2003. | The report states that the CCSP will provide additional specific community-based guidelines for scientific metadata content where and as appropriate. One approach will be to adopt the ISO 19115 /TC211 Geographic Information/Geomatics standard, which is built on the Federal Geospatial Data Clearinghouse (FGDC) core standards. | *** | http://www.climatescience.gov/Library/stratplan2003/final/ccspstratplan2003-chap13.htm |
| | NIH Enterprise Architecture | Active Directory (AD) Attribute Data Content and Management: Best Community Practice v1.3. 2008 | The document lists user attribute data content management rules. | ** | http://enterprisearchitecture.nih.gov/NR/rdonlyres/8B8AFA60-68A1-4155-A08F-03163B610E39/0/NIHRFC0008ActiveDirectoryAttributeDataContentandManagement.pdf |
| **Other** | NOAA | NOAA Report to Congress on Data and Information Management. 2005. | The report states that NOAA faces a major challenge in enabling interoperability between legacy systems and emerging data systems. This lack of system interoperability, across NOAA and across agencies, hampers the collaborations enabled by technological gains. | * | http://www.ngdc.noaa.gov/noaa_pubs/pdf/NOAA_Congress2005.pdf |

a. The applicability rating is shown as one, two, or three stars. A one-star rating (*) means that the information is related to ORD's SDM policy framework, but is expected to be of limited value in developing its policies and guidance. A two-star (**) rating means that the information is somewhat relevant to ORD's policy framework and/or is presented in limited detail. A three-star rating (***) means that the information provided is directly applicable to ORD's SDM policy/guidance goals and could serve as a model for ORD.

## 5.1 EPA Policy Information

Table 8 presents the findings for ten EPA documents and resources on managing scientific data for appropriate control. Three OEI policies and a National Health and Environmental Effects Research Laboratory (NHEERL) policy were identified related to this topic. A significant amount of supporting documentation was found for the "General Guidance" and "Specific Guidance" policy areas. No supporting documentation was found for "Goals, Vision Statements," "Principles," and "Recommendations for Policies." Two resources are directly relevant to EPA/ORD and six documents are considered to be at least somewhat relevant.

## 5.2 Other Federal Agency Policy Information

As Table 9 shows, a significant amount of information related to IP, data rights, and other issues involving the control of SDM has been developed by other federal agencies. Close to 20 documents and resources were identified, at all levels. A wide range of federal agencies and offices, including DOE, NASA, NIH, NOAA, NSF, the Climate Change Science Program, the National Science and Technology Council, and GAO, developed these documents. Eight resources are categorized as "Recommendations for Polices" and "Policies," half of which are considered to be of direct use to ORD. These "three-star" documents include DOE's *ARM Data Sharing and Distribution Policy*, which provides several policies that may be of direct use to ORD. Six documents provide information on general guidance for managing scientific data and two documents offer specific guidance. All but one of the guidance documents are "two-star" documents, considered to be potentially useful for ORD.

# 6. Maintain Version and Change Control on Data Sets (Policy Area #5)

Control of scientific data is needed to ensure the integrity of the data and the final product. Data within a project undergoes a continued development phase, from working data to mature, released, submitted, and archived data. This includes, for example, developing naming conventions and other approaches to maintain version and change control. Not all data require the same level of control, depending on customer-imposed requirements, and agency requirements. One factor to be considered is the maturity of the data set. For example, putting the data under control too early in its life cycle becomes burdensome and yields little business value. Control of data within a project might be considered as important as the control of the final product.

| Table 8. Manage scientific data for appropriate control: EPA documents and resources (Policy Area #4) | | | | | |
|---|---|---|---|---|---|
| **Level** | **Agency name/office** | **Document title/date** | **Description/pertinent aspects** | **Applicability[a]** | **Link (or reference)** |
| **Goals, vision statements** | | | | | |
| **Principles** | | | | | |
| **Recommendations for policies** | | | | | |
| **Policies** | OEI | Agency-wide Quality System Documents. 12/30/2009. | This site provides links to potentially helpful documents, such as Overview of the EPA Quality System for Environmental Data and Technology, Guidance for Developing Quality Systems for Environmental Programs, Guidance on Systematic Planning using the Data Quality Objectives Process, Guidance for Preparing Standard Operating Procedures, Guidance on Environmental Data Verification and Data Validation, and Data Quality Assessment: A Reviewer's Guide. | * | Email communication with Lynne Petterson, 6/10/09. |
| | OEI | EPA Quality Manual for Environmental Programs. 5/5/2000. | This document discusses Requirements for Reporting Environmental Data. Section 2.5 covers requirements for reporting technical data; 2.6 covers QA and QC requirements and guidance (mandatory and advisory). The document states: "The primary goal of the Agency-wide Quality System is to ensure that environmental programs and decisions are supported by data of the type and quality needed and expected for their intended use, and that decisions involving the design, construction, and operation of environmental technology are supported by appropriate quality assured engineering standards and practices. The scope of this Manual includes applicable environmental programs involving: the collection, evaluation, and use of environmental data by and for the Agency, and the design, construction, and operation of environmental technology by the Agency." | ** | http://www.epa.gov/irmpoli8/ciopolicy/2105-P-01-0.pdf |
| | National Health and Environmental Effects Research Laboratory (NHEERL) | NHEERL Data Management Policy and Practices: Genomics and Related High Throughput Data. | The document provides data sharing guidelines. For example, it states that data access to scientists not directly involved in the original project team will be initially restricted. The data available to outside investigators upon data upload will be limited to a brief description of the experiment sufficient to determine the utility of the underlying data for other purposes. In addition, data should be available to all members of a collaborative unit, irrespective of the composition of that unit, as soon as it is generated and reviewed for accuracy. | ** | EPA, Undated. |
| | OEI | National Geospatial Data Policy. CIO Policy Transmittal 05-002. 8/24/2008. | Geospatial data that are acquired by EPA (including contractors, grantees and vendors) must comply with all procedures and standards applicable to those data as if they were collected by EPA. | *** | http://www.epa.gov/esd/ggc/pdf/epa_natl_geo_data_policy.pdf |
| **General guidance** | OSWER | OSWER Life Cycle Management Guide. 1989 | Chapters 2-9 of the Life Cycle Management Guide provide suggestions as to how to properly manage information through the following phases: Definition, Design, Development, Implementation, Production, Evaluation and Archive. | *** | http://www.epa.gov/oswer/oswerlcm.htm |
| | OSWER | System Life Cycle Management Guidance Part 3 Practice Paper: Data Management During the Life Cycle. January1989. | Chapter 2 discusses how to select the right data model based on the level of impact it will have in relation to data sharing, organizational impact and cost. | ** | http://www.epa.gov/oswer/docs/oswerlcm/00000021.pdf |
| **Specific guidance (e.g., how to interpret and use policies)** | OEI | Procedures for Preparing Privacy Act Statements. 2009. | These procedures provide instructions for developing Privacy Act Statements (PAS) that must be provided to individuals when a federal agency requests personal information about them that is to be maintained in a system of records retrieved by name or personal identifier (5 U.S.C. 552a (e)(3)). These procedures list what to put in the PAS and provides a sample. This document lists the process the PAS goes through. | ** | Email communication with Lynne Petterson, 6/10/09. |
| | EPA Privacy Policy | Procedures for Preparing Privacy Impact Assessments. 2008 | These procedures provide instructions for determining if Personal Identifiable Information is collected in systems and ensuring adequate controls are put in place. The Privacy Impact Assessment (PIA) is the tool required by the OMB for addressing privacy issues with electronic systems. No specific guidance is provided for completing the PIA, just the process for submitting them and having them reviewed and accepted. | ** | http://intranet.epa.gov/oei/imitpolicy/qic/ciopolicy/2151-p-04.pdf |

| Table 8. Manage scientific data for appropriate control: EPA documents and resources (Policy Area #4) | | | | | |
|---|---|---|---|---|---|
| **Level** | **Agency name/office** | **Document title/date** | **Description/pertinent aspects** | **Applicability[a]** | **Link (or reference)** |
| | EPA Privacy Policy | Procedures for Preparing and Publishing Privacy Act Systems of Records Notices. 2008 | These procedures provide the instructions for preparing a System of Records Notice (SORN). These procedures apply whenever information is retrieved by a name or personal identifier from records under the control of the Agency, regardless of format or location (i.e., systems, applications, databases, Web sites, filing cabinets). These procedures must be followed before collecting personal information on an individual and retrieving it by one of those elements. | ** | http://intranet.epa.gov/oei/imitpolicy/qic/ciopolicy/2151-p-03.pdf |
| | Office of Water<br><br>Office of Wetlands, Oceans and Watersheds | Volunteer Stream Monitoring: A Methods Manual Chapter 6 Managing and Presenting Monitoring Data. 11/30/2006. | This chapter emphasizes the need to establish a method for data management and handling. It doesn't offer much guidance, however. | * | http://www.epa.gov/volunteer/stream/vms60.html |
| **Other** | | | | | |

a. The applicability rating is shown as one, two, or three stars. A one-star rating (*) means that the information is related to ORD's SDM policy framework, but is expected to be of limited value in developing its policies and guidance. A two-star (**) rating means that the information is somewhat relevant to ORD's policy framework and/or is presented in limited detail. A three-star rating (***) means that the information provided is directly applicable to ORD's SDM policy/guidance goals and could serve as a model for ORD.

| Table 9. Manage scientific data for appropriate control: Other federal agency documents and resources (Policy Area #4) | | | | | |
|---|---|---|---|---|---|
| **Level** | **Agency name/office** | **Document title/date** | **Description/pertinent aspects** | **Applicability[a]** | **Link (or reference)** |
| **Goals, vision statements** | Climate Change Science Program | Strategic Plan for the Climate Change Science Program Final Report. 2003. | The report states that full and open sharing of the full suite of global data sets for all global change researchers is a fundamental objective. | ** | http://www.climatescience.gov/Library/stratplan2003/final/ccspstratplan2003-chap13.htm |
| | DOE – ORNL | Developing Data Management Policy and Guidance Documents for your NARSTO Program or Project. 2005. | The document states that copyrights are a difficult issue. A plan must allow the instrument operator to reap the rewards of his/her efforts, but the common good is served by sharing. | ** | http://cdiac.ornl.gov/programs/NARSTO/DM_develop_guide.pdf |
| **Principles** | NSF | Long-Lived Digital Data Collections: Enabling Research and education in the 21st Century. 2005. | The reports states that NSF expects significant findings from the research and education activities it supports to be promptly submitted for publication, with authorship that accurately reflects the contributions of those involved. | ** | http://www.nsf.gov/pubs/2005/nsb0540/ |
| | DOE – ORNL | Developing Data Management Policy and Guidance Documents for your NARSTO Program or Project. 2005. | A policy must address how the project will ensure intellectual property rights are protected and co-authorship or credit is given to originators and investigators. | ** | http://cdiac.ornl.gov/programs/NARSTO/DM_develop_guide.pdf |
| **Recommendations for policies** | National Science and Technology Council | Harnessing the Power of Digital Data for Science and Society. 2009. | The report describes examples of mechanisms that include: reliable protection of security, privacy, confidentiality; and intellectual property rights in complex data environments. | *** | http://www.nitrd.gov/about/harnessing_power_web.pdf |
| | Climate Change Science Program | Strategic Plan for the Climate Change Science Program Final Report. 2003. | The report recommends improved access to data by expanding the Global Change Master Directory (GCMD) to facilitate access to data.<br><br>The CCSP will develop and implement guidelines for when and under what conditions data will be made available to users other than those who collected them. | ** | http://www.climatescience.gov/Library/stratplan2003/final/ccspstratplan2003-chap13.htm |
| | NIH – Office of Extramural Research | NIH Data Sharing Policy and Implementation Guide. 3/5/2003. | According to the policy, investigators sharing under their own auspices should consider using a data-sharing agreement to impose appropriate limitations on users.<br><br>The policy mentions that regardless of the mechanism used to share data, each dataset will require documentation.<br><br>The policy states that data sharing promotes many goals of the NIH research endeavor. It is particularly important for unique data that cannot be readily replicated. | ** | http://grants.nih.gov/grants/policy/data_sharing/ |
| | Government Accountability Office | Climate Change Research: Agencies Have Data-Sharing Policies but Could Do More to Enhance the Availability of Data from Federally Funded Research. 2007. | The report recommends NOAA develop mechanisms for agencies to be systematically notified when data have been submitted to archives, so that agency officials have current information about the extent of data availability in order to adjust data-sharing policies over time to best meet the needs of researchers and the communities that use their data. | *** | http://www.gao.gov/new.items/d071172.pdf |
| **Policies** | DOE – ARM | ARM Data Sharing and Distribution Policy. 2006. | ARM data are available to all participants on a free and open basis and are publishable upon receipt with acknowledgment of ARM as the source.<br><br>The policy states that researchers and participants may release their own preliminary data to whomever they wish and the preliminary data of other investigators with consent from the data's originator.<br><br>The automatic inclusion of a data originator as a co-author is not insisted upon in the ARM Program, but the source of any data should be clearly recognized either as a co-author or through an appropriate acknowledgment. | *** | http://www.arm.gov/data/docs/policy |
| | DOE – OSTI | NARSTO Quality Systems Data Center: Developing Data Management and Guidance Documents. 2006. | The guidelines state that a policy must provide standard names to identify the project, data files, and data sets. | ** | http://cdiac.ornl.gov/programs/NARSTO/DM_develop_guide.pdf |
| | NIH – Office of Extramural Research | NIH Data Sharing Policy and Implementation Guide. 3/5/2003. | The policy states that recognizing that the value of data often depends on their timeliness, and data sharing should occur in a timely fashion. NIH expects the timely release and sharing of data to be no later than the acceptance for publication of the main findings from the final dataset. | ** | http://grants.nih.gov/grants/policy/data_sharing/ |

| Table 9. Manage scientific data for appropriate control: Other federal agency documents and resources (Policy Area #4) | | | | | |
|---|---|---|---|---|---|
| **Level** | **Agency name/office** | **Document title/date** | **Description/pertinent aspects** | **Applicability[a]** | **Link (or reference)** |
| | NSF – Social, Behavioral and Economic Sciences | Data Archiving Policy. 7/8/2008. | The policy states that intellectual property rights may be at risk in some forms of data collection. The policy is intended to be flexible enough to accommodate the variety of scientific enterprises that constitute SES programs. No comprehensive set of rules is possible. | *** | http://www.nsf.gov/sbe/ses/common/archive.jsp |
| **General guidance** | NIH – Office of Extramural Research | NIH Data Sharing Policy and Implementation Guide. 3/5/2003. | The policy says that it is appropriate for scientific authors to acknowledge the source of data upon which their manuscript is based. Many investigators include this information in the methods and/or reference sections of their manuscripts. | ** | http://grants.nih.gov/grants/policy/data_sharing/ |
| | NIH – National Institute on Aging | Guidelines for Developing a Manual of Operations and Procedures (MOP). 2007. | The guidelines discuss the safeguards put in place to ensure participant confidentiality and data security. A list of safeguards is provided on p. 22. | * | http://www.nia.nih.gov/NR/rdonlyres/AEC5CE46-96E1-43D9-BA77-BAE8BF0D6CDC/0/ManualofProceduresMOPFinal1.doc |
| | NOAA | NOAA Administrative Order: 216-101: Ocean Data Acquisitions. 7/9/1990. | The order states that managers will work with their principal investigators to assure that other data, which may not be appropriate for archival at national centers, are documented and archived within the established period of time at the principal investigator's or an associated institution so that these data will be available for other uses upon request. | ** | http://www.corporateservices.noaa.gov/~ames/NAOs/Chap_216/naos_216_101.html |
| | NOAA | NOAA Administrative Order: 212-15. Management of Environmental and Geospatial Data and Information. 12/2/2008. | The order states that managers should maintain a list of applicable reference materials and will provide access to their electronic editions on its web site. | ** | http://www.corporateservices.noaa.gov/~ames/NAOs/Chap_212/naos_212_15.html |
| | NSF | Long-Lived Digital Data Collections: Enabling Research and education in the 21st Century. 2005. | The report identifies and describes the roles of key actors in digital data collections, and the key contents of a data management plan. | ** | http://www.nsf.gov/pubs/2005/nsb0540/ |
| | NSF – EAR | Implementation of the NSF Data Sharing Policy. April 2002. | The policy states that data may be made available for secondary use through submission to a national data center, publication in a widely available scientific journal, book or web site, through the institutional archives that are standard for a particular discipline, or through other EAR-specified repositories. | ** | http://www.nsf.gov/geo/ear/EAR_data_policy_204.pdf |
| **Specific guidance (e.g., how to interpret and use policies)** | NSF – Office of Polar Programs | Guidelines and Award Conditions for Scientific Data. 1998. | The guidelines recommend that principal investigators should make their data available to all reasonable requests and should submit the data collected to designated data centers as soon as possible, but no later than two years after the data are collected. | ** | http://www.nsf.gov/pubs/1999/opp991/opp991.doc |
| | NASA – Consultative Committee for Space Data Systems | Reference Model for an Open Archival Information System. 2002. | The document states that some projects have one-year proprietary periods before data are released to the science community. The policy is to avoid receipt of any proprietary data sets during the proprietary period.<br><br>The document states that the word processing format is proprietary, and it can't be acquired even to the level of simply viewing the document. It may be necessary to migrate the document to a non-proprietary format to ensure its long-term preservation. | ** | http://public.ccsds.org/publications/archive/650x0b1.pdf |
| **Other** | DOE – OSTI | The State of Data Management in the DOE Research and Development Complex. 7/14-15/2004 | According to the report, issues such as data ownership and DOE rights of re-use compound the problem of how to manage resulting data. | ** | http://www.osti.gov/publications/2007/datameetingreport.pdf |
| | NIH – Office of Extramural Research | NIH Data Sharing Policy and Implementation Guide. 3/5/2003. | The policy states that the rights and privacy of human subjects who participate in NIH-sponsored research must be protected at all times. It is the responsibility of the investigators, their Institutional Review Board (IRB), and their institution to protect the rights of subjects and the confidentiality of the data. | * | http://grants.nih.gov/grants/policy/data_sharing/ |

a. The applicability rating is shown as one, two, or three stars. A one-star rating (*) means that the information is related to ORD's SDM policy framework, but is expected to be of limited value in developing its policies and guidance. A two-star (**) rating means that the information is somewhat relevant to ORD's policy framework and/or is presented in limited detail. A three-star rating (***) means that the information provided is directly applicable to ORD's SDM policy/guidance goals and could serve as a model for ORD.

Development of this policy area might include a specification of the different levels of control to be used in ORD. In many cases, because of the nature of scientific data life cycle, etc., there may be unique "stages" that need to be identified before change control guidance may be developed. In many instances, a unique change control number is assigned to each request and entered into a change status accounting tracking system. Also included are a specification of the business rules for each level of control and the development of the roles and responsibilities for different levels of control. This may include a requirement to coordinate this activity with the development of the governance structure for SDM (see Section 2). It should be noted that this is not a comprehensive list of issues that may need to be addressed in this policy area.

## 6.1 EPA Policy Information

Table 10 presents the EPA documents and resources regarding maintaining version and change control on data sets. Only three individual documents were found. The OSWER document on *System Life Cycle Management Guidance* provides an example of both general guidance and specific guidance. The paper's information may be considered relevant for ORD (i.e., a two-star rating). In addition, two policy recommendations were identified, including ORD's *Scientific Data Management Strategy*, which suggests that EPA should establish standards, policies, and procedures for scientific data quality cleanup, change control, and audits. No supporting documentation was found for "Goals, Vision Statements," "Principles," and "Policies."

## 6.2 Other Federal Agency Policy Information

As shown in Table 11, a total of 15 other federal agency documents and resources were identified that address the issue of maintaining version and change control. These include four policy recommendations, four policies, four general guidance documents, and five specific guidance documents (numbers do not total 15 because some documents provide information at more than one level). A wide range of federal agencies and offices, including DOE, NASA, NIH, NOAA, NSF, and the Climate Change Science Program, developed these documents. The "Policies and "General Guidance" documents are all rated with two- or three-stars. Most notably, NASA's *Guidelines for Development of a Project Data Management Plan* provides information on policies for change control that could be very useful to ORD. The policy recommendations and specific guidance documents all have one- or two-star ratings.

| Table 10. Maintain version and change control on data sets: EPA documents and resources (Policy Area #5) | | | | | |
|---|---|---|---|---|---|
| **Level** | **Agency name/office** | **Document title/date** | **Description/pertinent aspects** | **Applicability[a]** | **Link (or reference)** |
| **Goals, vision statements** | | | | | |
| **Principles** | | | | | |
| **Recommendations for policies** | OEI | QIC Steering Committee – CIO Policy Consolidated Comments Form. 2009. | This is a steering committee form with reviewer comments regarding possible changes/clarifications to the following documents: Enterprise Content Management Policy, Metadata Standards for the Enterprise Content Management Program, and E-mail Records Procedures. | * | Email communication with Lynne Petterson, 6/10/2009. |
| | ORD | Scientific Data Management Strategy. 2007. | The strategy suggests that EPA should establish standards, policies, and procedures for scientific data quality cleanup, change control, and audits. For example, if problems or issues arise with the quality of scientific data, there must be a defined set of guidelines to determine what actions to take. | *** | Email communication with Lynne Petterson, 6/10/09. |
| **Policies** | | | | | |
| **General guidance** | OSWER | System Life Cycle Management Guidance Part 3 Practice Paper: Configuration Management. 1989. | This document provides guidance regarding the implementation of configuration management (CM), defined as systematically identifying the characteristics of a system and formally controlling any changes or additions to these items. The guidance describes specific activities associated with CM; describes project organization structures to accomplish CM; describes the documentation of project-specific CM activities in a CM plan.<br><br>Chapter 2 of this paper discusses the establishment of configuration item identification, which acts as "labels" for the characteristics described in the documentation. This chapter also discusses change request impact analysis. | ** | http://www.epa.gov/oswer/docs/oswerlcm/00000019.pdf |
| **Specific guidance (e.g., how to interpret and use policies)** | OSWER | System Life Cycle Management Guidance Part 3 Practice Paper: Configuration Management. January 1989. | Chapter 3 of this paper provides steps to implementing CM into an organization. | ** | http://www.epa.gov/oswer/docs/oswerlcm/00000019.pdf |
| **Other** | | | | | |

a. The applicability rating is shown as one, two, or three stars. A one-star rating (*) means that the information is related to ORD's SDM policy framework, but is expected to be of limited value in developing its policies and guidance. A two-star (**) rating means that the information is somewhat relevant to ORD's policy framework and/or is presented in limited detail. A three-star rating (***) means that the information provided is directly applicable to ORD's SDM policy/guidance goals and could serve as a model for ORD.

| Table 11. Maintain version and change control on data sets: Other federal agency documents and resources (Policy Area #5) | | | | | |
|---|---|---|---|---|---|
| **Level** | **Agency name/office** | **Document title/date** | **Description/pertinent aspects** | **Applicability[a]** | **Link (or reference)** |
| **Goals, vision statements** | | | | | |
| **Principles** | | | | | |
| **Recommendations for policies** | NOAA | NOAA Administrative Order: 212-15. Management of Environmental and Geospatial Data and Information. 12/2/2008. | The order recommends managers be alert to and mitigate the risks caused by changes of instruments, platforms, locations, and methods for observing or processing data. | * | http://www.corporateservices.noaa.gov/~ames/NAOs/Chap_212/naos_212_15.html |
| | NOAA | Environmental Data Management at NOAA: Archiving, Stewardship, and Access. 2007. | The book recommends that NOAA policies establish and maintain data and metadata migration plans for all current and future long-term archive systems to adapt to information technology evolution. | ** | National Research Council, 2007. |
| | DOE – ORNL | Developing Data Management Policy and Guidance Documents for your NARSTO Program or Project. 2005. | The guidance recommends that policies adopt, adapt, or refine model documents as appropriate with input from managers, investigators, modelers, and data coordinators. They must also address data validation and assigning quality levels. | * | http://cdiac.ornl.gov/programs/NARSTO/DM_develop_guide.pdf |
| | NIH – Office of Extramural Research | NIH Data Sharing Policy and Implementation Guide. 3/5/2003. | According to the policy, investigators sharing under their own auspices should consider using a data-sharing agreement to impose appropriate limitations on users. | ** | http://grants.nih.gov/grants/policy/data_sharing/ |
| **Policies** | NIH – DAIDS | Requirements for Data Management and Statistics for DAIDS Funded and/or Sponsored Clinical Trials. 2007. | This policy states that change control procedures should ensure quality control in changes made to the data collection tools. It includes how changes are requested, how the impact of changes is assessed, who is responsible for authorizing the changes, how the changes are tested and released, and how the changes are documented. | *** | http://www3.niaid.nih.gov/LabsAndResources/resources/DAIDSClinRsrch/DataManagement.htm |
| | NASA – Office of Space Science and Applications | Guidelines for Development of a Project Data Management Plan (PDMP). 1993. | The guidelines state that policies should illustrate the plans for modifications and updates to this document over time, and how those changes will be controlled.<br><br>The guidelines state that each PDMP should have a glossary of terms relevant to that project. Each PDMP should have an acronym list of terms relevant to that project. | *** | http://nssdc.gsfc.nasa.gov/nssdc/pdmp_guidelines_march93.rtf |
| | DOE – ORNL | Developing Data Management Policy and Guidance Documents for your NARSTO Program or Project. 2005. | The guidelines state that a policy must provide standard names to identify the project, data files, and data sets. | ** | http://cdiac.ornl.gov/programs/NARSTO/DM_develop_guide.pdf |
| | NSF – EAR | Implementation of the NSF Data Sharing Policy. April 2002. | The document states that data inventories should be published or entered into a public database periodically and whenever there is a significant change in type, location, or frequency of such observations. | ** | http://www.nsf.gov/geo/ear/EAR_data_policy_204.pdf |
| **General guidance** | NASA – National Space Science Data Center | White Paper on NASA Science Data Retention. 2007. | The paper suggests that ensuring continuing data integrity and usability requires periodic data renewal cycles. Some of these cycles will involve only bit migration from old to new media. | *** | http://nssdc.gsfc.nasa.gov/nssdc/data_retention.html |
| | NASA – Consultative Committee for Space Data Systems | Reference Model for an Open Archival Information System. 2002. | The document addresses the migration of digital information to new media and forms, the data models used to represent the information, the role of software in information preservation, and the exchange of digital information among archives. | ** | http://public.ccsds.org/publications/archive/650x0b1.pdf |
| | NSF – Social, Behavioral and Economic Sciences | Data Archiving Policy. 7/8/2008. | According to the policy, the kinds of qualitative information collected in research projects supported by SES can range from microfilms and other copies of very old documents to oral interviews and video tapes about historical events in science or about contemporary technological controversies. They can consist of hand- written records of open-ended interviews. Investigators should consider whether and how they can develop special arrangements to keep or store these materials so that others can use them. | ** | http://www.nsf.gov/sbe/ses/common/archive.jsp |
| | NIH – National Institute on Aging | Guidelines for Developing a Manual of Operations and Procedures (MOP). 2007. | The guidelines state that when updating, staff must correct data and maintain an audit trail of all data changes. | *** | http://www.nia.nih.gov/NR/rdonlyres/AEC5CE46-96E1-43D9-BA77-BAE8BF0D6CDC/0/ManualofProceduresMOPFinal1.doc |

| Table 11. Maintain version and change control on data sets: Other federal agency documents and resources (Policy Area #5) | | | | | |
|---|---|---|---|---|---|
| **Level** | **Agency name/office** | **Document title/date** | **Description/pertinent aspects** | **Applicability[a]** | **Link (or reference)** |
| **Specific guidance (e.g., how to interpret and use policies)** | NASA – Jet Propulsion Lab | Cassini/Huygens Program Archive Plan for Science Data. 2004. | The policy states that filenames will adhere to International Organization for Standardization (ISO) 9660 level 2 specifications that allow the total filename length of 31 characters. | * | http://trs-new.jpl.nasa.gov/dspace/bitstream/2014/14261/1/00-0674.pdf |
| | DOE – ORNL | Developing Data Management Policy and Guidance Documents for your NARSTO Program or Project. 2005. | The document mentions that data from NARSTO projects are formatted in the NARSTO Data Exchange Standard (a spreadsheet-compatible layout, which uses standardized and consistent metadata values). | * | http://cdiac.ornl.gov/programs/NARSTO/DM_develop_guide.pdf |
| | DOE – ARM | ARM Data Sharing and Distribution Policy. 2006. | The ARM External Data Center and Archive will track data versions and ensure latest data versions are made available to data recipients. | * | http://www.arm.gov/data/docs/policy |
| | DOE – ORNL | The NARSTO Atmospheric Measurements Template. 4/29/2005. | Every regular measurement needs to have an associated NARSTO standard flag. "Dimensional" variables indicate the setting for measurements, such as site, date, time, and altitude. | * | ftp://narsto.esd.ornl.gov/pub/DES_metadata/var_names_web_sources/NARSTO_template_atmospheric_measurements.xls |
| | DOE – OSTI – LLNL | Management of OSTI-LLNL Electronic Data. 2005. | According to the document, electronic files may be converted from one software to another. Staff should include an entry in their scientific notebook indicating that a verification of file conversion has been conducted. It also includes detailed steps on data transfer. | ** | https://eed.llnl.gov/ymp/pdf/IM-317550-2.pdf |
| **Other** | | | | | |

a. The applicability rating is shown as one, two, or three stars. A one-star rating (*) means that the information is related to ORD's SDM policy framework, but is expected to be of limited value in developing its policies and guidance. A two-star (**) rating means that the information is somewhat relevant to ORD's policy framework and/or is presented in limited detail. A three-star rating (***) means that the information provided is directly applicable to ORD's SDM policy/guidance goals and could serve as a model for ORD.

# 7. Retain Data Commensurate with Its Value (Policy Area #6)

Data should only be retained as long as it has value to current or future users. There must be a method in place to ensure adequate retention and preservation of data that have value to the agency and how to dispose of data that no longer have value. Data can be retained in many ways, at differing costs (e.g., on-line, near-on-line, archives). Determining the probability and value of future use and the appropriate retention mechanism and timing requires cost-effectiveness assessment and the participation of all stakeholders, including those who represent potential future users of the data (e.g., librarians).

The development of this policy area will include guidance on the factors to be considered in making data retention decisions. Data retention includes preservation, maintenance and control of data for future use, while data refresh/migration refers to the periodic transfer of data to new hardware/software configurations to ensure data can still be used. In addition, disposal of data includes specific instructions for elimination of data from different types of media, to different levels of assurance of destruction. Guidance on making retention decisions might include developing a flowchart or decision tree for such decisions. There may also be a development of guidance regarding business rules that different retention decisions trigger. For example, this may include the maintenance of associated indices of those data. In many cases, there may be case studies since most of these decisions are both decentralized and unique to each agency or project.

## 7.1 EPA Policy Information

EPA information on retaining data commensurate with its value is shown in Table 12. Ten documents were found that might offer insight into Policy Area #6. These documents cover the "Principles," "Recommendations for Policies," "Policies," "General Guidance," "Specific Guidance," and "Other" areas. Three documents are very applicable to ORD (i.e., a three-star rating), including OEI's *Enterprise Content Management Policy*, which provides information on data storage and records management. No supporting documentation was found for "Goals, Vision Statements."

| Table 12. Retain data commensurate with its value: EPA documents and resources (Policy Area #6) | | | | | |
|---|---|---|---|---|---|
| **Level** | **Agency name/office** | **Document title/date** | **Description/pertinent aspects** | **Applicability[a]** | **Link (or reference)** |
| **Goals, vision statements** | | | | | |
| **Principles** | Great Lakes National Program Office | Great Lakes Environmental Database. 2009. | The Great Lakes Environmental Database pages state: "Long after the studies are completed, the data remain and must be managed." | ** | http://www.epa.gov/glnpo/monitoring/data_proj/glenda/index.html |
| **Recommendations for policies** | ORD | Scientific Data Management Strategy. 2007. | The paper states that it is necessary to maintain scientific records for historical research and regulatory purposes. It notes that there are many conflicting data formats, making it difficult to retrieve and re-use the information they contain. Therefore, a policy should develop overall standards and guidelines for acceptable formats for long-term retention. The paper also suggests a records retention schedule to ensure that records are kept only as long as legally and operationally required and that obsolete records are retired or disposed of in a controlled manner. This strategy paper also discusses the need for a disaster recovery plan. | *** | EPA, 2007. |
| **Policies** | OEI | Enterprise Content Management Policy. July 8, 2009. | This policy establishes the EPA Enterprise Content Management Program. The Program advises EPA staff on how best to store data, how to apply established data and metadata resources and how to manage records in accordance with all federal and Agency records management statues, regulations, policies, procedures and standards. | *** | Email communication with Lynne Petterson, 6/10/09. |
| | National Health and Environmental Effects Research Laboratory (NHEERL) | NHEERL Data Management Policy and Practices: Genomics and Related High Throughput Data. | The policy states that storage should be completed within three months after completion of primary data generation to allow for sufficient quality assurance of the raw data. If additional time is needed for QA of the raw data, the length should be determined following discussion with the project lead and the appropriate AD. | ** | EPA, Undated. |
| | OEI | National Geospatial Data Policy. CIO Policy Transmittal 05-002. 8/24/2008. | The program office or project sponsoring the original collection effort is responsible for spatial data maintenance and decisions regarding ultimate retention and disposal. Data disposition for archiving must also comply with the records retention requirements of the program under which the data were collected. | *** | http://www.epa.gov/esd/ggc/pdf/epa_natl_geo_data_policy.pdf |
| **General guidance** | Great Lakes National Program Office | Introduction to Lake Michigan Mass Balance Data. 3/9/2006. | The database was developed under the following guidelines: develop a system having cross-program and project utility, document the quality of all data populating the system, ensure that the resulting data base has long-term value, and avoid duplicating effort with other data systems. | ** | http://www.epa.gov/greatlakes/lmmb/database.html |
| | Office of Water<br><br>Office of Wetlands, Oceans and Watersheds | Volunteer Stream Monitoring: A Methods Manual: Chapter 6 Managing and Presenting Monitoring Data. 11/30/2006. | This document stresses checking with data users to ascertain both how the data will be used, and processes/presentation formats. References STORET as best repository for data sharing. | ** | http://www.epa.gov/volunteer/stream/vms60.html |
| **Specific guidance (e.g., how to interpret and use policies)** | EPA Records Schedule | EPA Records Schedule, Data Standards and Registry Service. 7/31/2009. | This schedule authorizes the disposition of the record copy in any media (media neutral), excluding any records already in electronic form. Records designated for permanent retention must be transferred to the National Archives in accordance with National Archives and Records Administration (NARA) standards at the time of transfer. (N1-412-08-15). This document provides guidance on what type of disposition is required for each type of media. | * | http://www.epa.gov/records/policy/schedule/sched/096.htm |
| | OEI | E-mail Records Procedures. September 25, 2009. | These procedures state: "E-mail is a significant means of conducting Agency business. As such, some e-mail messages qualify as Agency records and must be managed appropriately to successfully carry out the mission of EPA. Proper e-mail records management enables the Agency to meet its business needs and legal obligations, including responding to Freedom of Information Act (FOIA), litigation and other production requests. This document provides specific steps to maintain email records via ECMS or a paper recordkeeping system." | * | Email communication with Lynne Petterson, 6/10/09. |

| Table 12. Retain data commensurate with its value: EPA documents and resources (Policy Area #6) | | | | | |
|---|---|---|---|---|---|
| **Level** | **Agency name/office** | **Document title/date** | **Description/pertinent aspects** | **Applicability[a]** | **Link (or reference)** |
| **Other** | Office of Technology Operations and Planning | IT Policy Mega-Matrix. 2009. | The IT Policy Mega-Matrix is a master list of the IT policy documents (e.g., Policies, Procedures, Standards, and Guidance) that OTOP maintains. Pages 22-25 contain archive documents. | * | This document is located on the EPA intranet at: http://intranet.epa.gov/otop/itpolicy/IT_Policy_Mega-Matrix_Feb2009_external.pdf |

a. The applicability rating is shown as one, two, or three stars. A one-star rating (*) means that the information is related to ORD's SDM policy framework, but is expected to be of limited value in developing its policies and guidance. A two-star (**) rating means that the information is somewhat relevant to ORD's policy framework and/or is presented in limited detail. A three-star rating (***) means that the information provided is directly applicable to ORD's SDM policy/guidance goals and could serve as a model for ORD.

## 7.2 Other Federal Agency Policy Information

Table 13 shows close to 20 federal agency resources on retaining data commensurate with its value, developed by DOE, NASA, NIH, NOAA, NSF, and the Climate Change Science Program. These documents cover all levels except for "Goals, Vision Statements" and "Other." Three resources are of particular relevance to ORD. *Environmental Data Management at NOAA: Archiving, Stewardship, and Access* provides policy recommendations and specific guidance on developing infrastructure that ensures long-term access and preservation of data assets. The Climate Change Science Program's *Strategic Plan for the Climate Change Science Program* presents policies and general guidance data acquisition, retention, and purging. Another key document is DOE's *Guidelines for Archiving Data in the NARSTO Permanent Archive*. Most of the other documents in this policy area are rated with two stars, and provide additional valuable information for ORD in developing policies and guidance on data retention.

# 8. Ensure that Scientific Data Management Processes Are Integrated with Knowledge Management Initiative (Policy Area #7)

Data management and KM are interdependent. The foundation provided by SDM can enable both knowledge sharing (through discovery and retrieval of scientific data, for example) and knowledge retention by supporting knowledge harvesting when a principal investigator retires or leaves EPA. In all SDM activities, one must remain knowledgeable about KM initiatives.

Development of this policy area may include a determination of any KM initiatives that will impact ORD during the SDM implementation horizon, and of any interdependence with the SDM initiative. Also included may be a review of KM tools (processes, approaches to change management, or analytical tools) that can be adapted for the SDM initiative. It is important to note that both SDM and KM face cultural hurdles, since both are intimately involved in professionals' daily work habits and may benefit from shared approaches. Some data mining tools developed for the "unstructured" data could prove useful to SDM (for example, analyzing collections of images, text, and briefings could lead to discovery of projects that have useful associated scientific data).

## 8.1 EPA Policy Information

There was no supporting documentation found for this policy area.

| Table 13. Retain data commensurate with its value: Other federal agency documents and resources (Policy Area #6) | | | | | |
|---|---|---|---|---|---|
| **Level** | **Agency name/office** | **Document title/date** | **Description/pertinent aspects** | **Applicability[a]** | **Link (or reference)** |
| **Goals, vision statements** | | | | | |
| **Principles** | NASA – Jet Propulsion Laboratory | Cassini/Huygens Program Archive Plan for Science Data. 2004. | The policy should ensure the long-term preservation of data. | * | http://trs-new.jpl.nasa.gov/dspace/bitstream/2014/14261/1/00-0674.pdf |
| | DOE – OSTI | The State of Data Management in the DOE Research and Development Complex. 7/14-15/2004. | According to the report, a data management plan should describe how data should be preserved, documentation needed to assure validation and future use, and funding/infrastructure needs to ensure longevity. It states that turning over data files is not mandatory. | ** | http://www.osti.gov/publications/2007/datameetingreport.pdf |
| | NOAA | Environmental Data Management at NOAA: Archiving, Stewardship, and Access. 2007. | Principle #1. Environmental data should be archived and made easily accessible to researchers and consumers.<br><br>Principle #9. A formal, ongoing process with broad community input is needed to decide what data to archive and what to dispose. | ** | National Research Council, 2007. |
| **Recommendations for policies** | NIH – National Cancer Institute | National Cancer Institute (NCI), Division of Cancer Prevention (DCP), Data Management Requirements. October 2003. | The document recommends that policies indicate how long the records will be retained and when the process begins. | ** | http://prevention.cancer.gov/clinicaltrials/management/consortia/step-2/data |
| | DOE – ORNL | Developing Data Management Policy and Guidance Documents for your NARSTO Program or Project. 2005. | The guidance recommends that managers ask about value of data: short-term (3-5 years), mid-term (10 years), or longer (20 years). In addition, scientists are encouraged to document their data at a level sufficient to satisfy the "20-year test." Someone 20 years from now, not familiar with the data or how they were obtained, should be able to find data of interest and then fully understand and use the data solely with the aid of the documentation archived with the data. | ** | http://cdiac.ornl.gov/programs/NARSTO/DM_develop_guide.pdf |
| | NOAA | Environmental Data Management at NOAA: Archiving, Stewardship, and Access. 2007. | The book states that archiving and access decisions are closely related. When resources are limited, access to older or less commonly used data should be scaled back, rather than removing data from the archive. | ** | National Research Council, 2007. |
| **Policies** | NIH – Office of Extramural Research | NIH Data Sharing Policy and Implementation Guide. 3/5/2003. | The policy states that recognizing that the value of data often depends on their timeliness, and data sharing should occur in a timely fashion. NIH expects the timely release and sharing of data to be no later than the acceptance for publication of the main findings from the final dataset. | ** | http://grants.nih.gov/grants/policy/data_sharing/ |
| | NIH – DAIDS | Requirements for Data Management and Statistics for DAIDS Funded and/or Sponsored Clinical Trials. 2007. | The requirements state that policies must develop a plan for record retention, both electronic and hard copy. It must include when record retention begins, the length of time the records are retained, where the records are retained, the security of the storage space, who has access to the storage space, and who is responsible for approving access. | ** | http://www3.niaid.nih.gov/LabsAndResources/resources/DAIDSClinRsrch/DataManagement.htm |
| | Climate Change Science Program | Strategic Plan for the Climate Change Science Program Final Report. 2003. | The report states that procedures and criteria for setting priorities for data acquisition, retention, and purging should be developed by participating agencies, both nationally and internationally. A clearinghouse process should be established to prevent the purging and loss of important data sets. | *** | http://www.climatescience.gov/Library/stratplan2003/final/ccspstratplan2003-chap13.htm |
| | NASA – Heliophysics Great Observatory | NASA Heliophysics Science Data Management Policy. 2007. | The National Space Science Data Center (NSSDC) policy ensures the maintenance of the permanent archive. The physical arrangements for such storage will be made in whatever manner is most economical, secure, and accessible.<br><br>NASA archives must have user advisory committees to advise on the likely future use and value of datasets candidate for resource-intensive renewal cycles. | * | http://hpde.gsfc.nasa.gov/Heliophysics_Data_Policy_2007June25.pdf |

| Table 13. Retain data commensurate with its value: Other federal agency documents and resources (Policy Area #6) | | | | | |
|---|---|---|---|---|---|
| **Level** | **Agency name/office** | **Document title/date** | **Description/pertinent aspects** | **Applicability[a]** | **Link (or reference)** |
| | NSF – EAR | Implementation of the NSF Data Sharing Policy. April 2002. | The policy recommends the preservation of all data, samples, physical collections, and other supporting materials needed for long-term earth science research. Education is required of all EAR-supported researchers. | * | http://www.nsf.gov/geo/ear/EAR_data_policy_204.pdf |
| **General guidance** | NASA – Office of Space Science and Applications | Guidelines for Development of a Project Data Management Plan (PDMP). 1993. | The guidelines state that project data repositories are project specific, providing temporary storage for active data as they are being processed and analyzed. This section of a PDMP should address the requirements placed on the project data repositories. Once archived, data sets and supporting information shall be periodically reviewed to assess their value for continued retention by NASA.<br><br>The guidelines state that plans should address how data will transition from project to permanent discipline archives. Table 8 provides a format for summarizing storage requirements by data set. | ** | http://nssdc.gsfc.nasa.gov/nssdc/pdmp_guidelines_march93.rtf |
| | NSF | Long-Lived Digital Data Collections: Enabling Research and education in the 21st Century. 2005. | The report states that the vast majority of NSF support carries with it no long-term commitment. Principal investigator grants have a duration of several years. Centers are typically funded for five years with a potential for an additional five years of funding. Long-lived digital data collections raise a new issue. It is timely for NSF to consider whether it should make very long-term commitments to a digital collection. | ** | http://www.nsf.gov/pubs/2005/nsb0540/ |
| | DOE – ORNL | Guidelines for Archiving Data in the NARSTO Permanent Data Archive. 5/2/2006. | According to the document, NARSTO encourages scientists to document their data at a level sufficient to satisfy the “20-year test."<br><br>The document includes guidance that stresses characteristics of projects and data that are worthy or able to be well-archived. | *** | http://cdiac.ornl.gov/programs/NARSTO/Guidelines_for_Archiving_NARSTO_Data.pdf |
| | Climate Change Science Program | Strategic Plan for the Climate Change Science Program Final Report. 2003. | It mentions that lessons learned from NASA's efforts in handling its current holdings (more than 2,500 terabytes) must be used by the community.<br><br>Many important heritage data sets face a growing risk of loss due to deterioration of paper records, obsolescence of electronic media and associated hardware and software, and the gradual loss of experienced personnel. | *** | http://www.climatescience.gov/Library/stratplan2003/final/ccspstratplan2003-chap13.htm |
| | NASA – Consultative Committee for Space Data Systems | Reference Model for an Open Archival Information System. 2002. | The document states that a long-term timeframe is long enough to be concerned with the impacts of changing technologies, including support for new media and data formats, or with a changing user community. | ** | http://public.ccsds.org/publications/archive/650x0b1.pdf |
| | NIH – National Cancer Institute | National Cancer Institute (NCI), Division of Cancer Prevention (DCP), Data Management Requirements. October 2003. | The document states that the data management plan should include a description of the security plan and should delineate responsibilities and expected behavior of all individuals who have access to study data and systems. It also indicates how long the records will be retained and when the process begins. | ** | http://prevention.cancer.gov/clinicaltrials/management/consortia/step-2/data |
| | NOAA | NOAA Administrative Order: 216-101: Ocean Data Acquisitions. 1990. | This order recognizes that data are used weeks to decades after the initial data acquisition. These archived data sets usually have more stringent quality requirements than do real-time data. | * | http://www.corporateservices.noaa.gov/~ames/NAOs/Chap_216/naos_216_101.html |

| Table 13. Retain data commensurate with its value: Other federal agency documents and resources (Policy Area #6) | | | | | |
|---|---|---|---|---|---|
| **Level** | **Agency name/office** | **Document title/date** | **Description/pertinent aspects** | **Applicability[a]** | **Link (or reference)** |
| **Specific guidance (e.g., how to interpret and use policies)** | NASA – National Space Science Data Center | White Paper on NASA Science Data Retention. 2007. | The paper states that data sets leading up to the production of the definitive dataset should be retained only to a point six months past the creation and certification of the definitive dataset.<br><br>The paper also states that derived datasets should be retained as long as they remain scientifically viable (i.e., algorithms or coefficients used in their derivation remain credible) and the cost of regenerating them (for some anticipated request level) outweighs the cost of their retention and maintenance.<br><br>The paper recommends that NASA archives should have user advisory committees to advise on (among other things) the likely future use and value of data sets candidate for resource-intensive renewal cycles. | ** | http://nssdc.gsfc.nasa.gov/nssdc/data_retention.html |
| | NOAA | Environmental Data Management at NOAA: Archiving, Stewardship, and Access. 2007. | The book recommends that NOAA develop and maintain scalable and reliable infrastructure that ensures long-term access and preservation of data assets.<br><br>The book suggests that it may be cost-effective to regenerate certain kinds of environmental data on demand. | *** | National Research Council, 2007. |
| | DOE – ORNL | NARSTO Quality Systems Management Plan. 9/30/1999. | The document provides a project plan and data archival process flow chart. | * | http://cdiac.ornl.gov/programs/NARSTO/pdf/qsmp_current_version.PDF |
| | NSF – Division of Ocean Sciences | Division of Ocean Sciences: Data and Sample Policy. 11/3/2003. | According to the policy, principal investigators are required to submit all environmental data collected to the designated National Data Centers as soon as possible, but no later than two years after the data are collected. | * | http://www.nsf.gov/pubs/2004/nsf04004/nsf04004_1b.htm |
| | NSF – EAR | Implementation of the NSF Data Sharing Policy. 2002. | The paper states that for those programs in which selected principal investigators have initial periods of exclusive data use, data should be made openly available as soon as possible, but no later than two years after the data were collected. This period may be extended under exceptional circumstances, but only by agreement between the principal investigator and the NSF. For continuing observations or for long-term (multi-year) projects, data are to be made public annually. | ** | http://www.nsf.gov/geo/ear/EAR_data_policy_204.pdf |
| | NIH – National Institute on Aging | Guidelines for Developing a Manual of Operations and Procedures (MOP). 2007. | The guidelines state that the length of time study files are to be maintained should be specified in the MOP. NIH policy requires that studies conducted under a grant retain participant forms for three years, while studies conducted under contract must retain participant forms for seven years. Individual Institutional Review Boards (IRBs), institutions, states, and countries may have different requirements for record retention. Investigators should adhere to the most rigorous requirements. | * | http://www.nia.nih.gov/NR/rdonlyres/AEC5CE46-96E1-43D9-BA77-BAE8BF0D6CDC/0/ManualofProceduresMOPFinal1.doc |
| **Other** | | | | | |

a. The applicability rating is shown as one, two, or three stars. A one-star rating (*) means that the information is related to ORD's SDM policy framework, but is expected to be of limited value in developing its policies and guidance. A two-star (**) rating means that the information is somewhat relevant to ORD's policy framework and/or is presented in limited detail. A three-star rating (***) means that the information provided is directly applicable to ORD's SDM policy/guidance goals and could serve as a model for ORD.

## 8.2 Other Federal Agency Policy Information

There was no supporting documentation found for this policy area.

# 9. Conclusions

As described above, a wide variety of documents and resources about SDM-related goals, policies, and guidance developed by EPA and other federal agencies were identified. This review demonstrates that, in general, federal agencies have yet not developed comprehensive policies and approaches for managing the burgeoning amount of scientific data that they create. Nevertheless, this compilation of resources provides a solid base of information for beginning to develop a set of ORD SDM policies and guidance.

Table 14 presents a summary of the SDM documents and resources by policy area, level, and applicability rating. The following sections briefly summarize these resources by policy area, key resources, information gaps, and next steps.

## 9.1 Resources by Policy Area and Agency

A total of 78 individual documents and other SDM resources were identified. The text box at right, *Number of individual documents and resources by agency*, shows how these resources are distributed by agency (see Appendix C for a complete list of resources).

**Number of individual documents and resources by agency**

EPA – 40
DOE – 7
NASA – 5
NIH – 7
NOAA – 6
NSF – 5
Other – 8

Total – 78

Many of these resources apply to more than one policy area. Consequently, as shown in Table 14, when allowing for double counting, the 78 EPA and non-EPA federal agency documents and resources provided 189 references to SDM goals and visions, principles, policy recommendations, policies, general guidance, specific guidance, and related information.

The greatest number of EPA SDM resources identified during this task relate to Policy Area #2, developing a SDM plan that covers the full data life cycle (16 double-counted resources), and Policy Area #3, identifying scientific data with metadata to enable needed business operations (17 double-counted resources). More limited information is available regarding all the other policy areas except for Policy Area #7, ensuring SDM integrates with KM.

In terms of non-EPA federal agencies, a large number of resources (ranging from 17–29 double-counted items per policy area) are available relating to Policy Areas #2, #3, #4, #5, and #6. Eight (double-counted) resources (which received two- and three-star applicability ratings) were found for Policy Area #1 and no resources were found for Policy Area #7.

**Table 14. Number of references to SDM documents and resources by policy area, applicability rating, agency type, and level**

| | Number of references[a] | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Policy #1: Manage SDM as asset or liability | | | Policy #2: Develop a SDM plan | | | Policy #3: Identify SDM with metadata | | | Policy #4: Manage SDM for appropriate control | | | Policy #5: Maintain version and change control | | | Policy #6: Retain data commensurate with its value | | | Policy #7: Ensure SDM integrates with KM | | | All policy areas | | | |
| | * | ** | *** | * | ** | *** | * | ** | *** | * | ** | *** | * | ** | *** | * | ** | *** | * | ** | *** | * | ** | *** | All |
| **EPA** | | | | | | | | | | | | | | | | | | | | | | | | | |
| Goal, vision statements | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Principles | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 2 | 4 |
| Recommendations for policies | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 4 | 0 | 5 |
| Policies | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 2 | 1 | 0 | 0 | 0 | 0 | 1 | 2 | 0 | 0 | 0 | 2 | 10 | 1 | 13 |
| General guidance | 0 | 0 | 2 | 3 | 2 | 3 | 3 | 1 | 3 | 0 | 1 | 1 | 0 | 1 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 6 | 7 | 9 | 22 |
| Specific guidance | 0 | 0 | 0 | 0 | 3 | 0 | 1 | 5 | 1 | 1 | 3 | 0 | 0 | 1 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 4 | 12 | 1 | 17 |
| Other | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 2 |
| Total | 0 | 1 | 5 | 5 | 6 | 5 | 5 | 7 | 5 | 2 | 6 | 2 | 1 | 2 | 1 | 3 | 4 | 3 | 0 | 0 | 0 | 16 | 26 | 21 | 63 |
| **Federal agencies** | | | | | | | | | | | | | | | | | | | | | | | | | |
| Goals, vision statements | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 3 | 0 | 4 |
| Principles | 0 | 1 | 1 | 0 | 2 | 1 | 1 | 3 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 1 | 2 | 0 | 0 | 0 | 0 | 2 | 10 | 2 | 14 |
| Recommendations for policies | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 3 | 1 | 0 | 2 | 2 | 2 | 2 | 0 | 0 | 3 | 0 | 0 | 0 | 0 | 3 | 10 | 3 | 16 |
| Policies | 0 | 3 | 0 | 1 | 2 | 2 | 2 | 3 | 1 | 0 | 2 | 2 | 0 | 2 | 2 | 2 | 2 | 1 | 0 | 0 | 0 | 5 | 14 | 8 | 27 |
| General guidance | 0 | 0 | 0 | 3 | 7 | 0 | 2 | 5 | 1 | 1 | 5 | 0 | 0 | 2 | 2 | 1 | 4 | 2 | 0 | 0 | 0 | 7 | 23 | 5 | 35 |
| Specific guidance | 0 | 1 | 1 | 2 | 2 | 0 | 1 | 2 | 2 | 0 | 2 | 0 | 4 | 1 | 0 | 3 | 2 | 1 | 0 | 0 | 0 | 10 | 10 | 4 | 24 |
| Other | 0 | 1 | 0 | 1 | 1 | 0 | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 3 | 3 | 0 | 6 |
| Total | 0 | 6 | 2 | 8 | 15 | 3 | 8 | 16 | 5 | 2 | 16 | 4 | 6 | 7 | 4 | 7 | 13 | 4 | 0 | 0 | 0 | 31 | 73 | 22 | 126 |
| **Total by policy and rating** | **0** | **7** | **7** | **13** | **21** | **8** | **13** | **23** | **10** | **4** | **22** | **6** | **7** | **9** | **5** | **10** | **17** | **7** | **0** | **0** | **0** | **47** | **99** | **43** | **189** |
| **Grand totals** | **14** | | | **42** | | | **46** | | | **32** | | | **21** | | | **34** | | | **0** | | | **189** | | | **189** |

Key to applicability ratings:

* = Related topic that ORD needs to be aware of, but doesn't offer much useful information.

** = Covers a few/some of the subject areas within the policy area – but in limited detail.

*** = A model for re-use, very applicable to ORD.

a. Double counting of documents occurs since some documents refer to more than one policy area and/or level.

## 9.2 Key Resources

As shown in Table 14, the greatest number of resources (99 when double counted) received a two-star applicability rating, followed by 47 (double-counted) resources with a one-star rating, and 43 (double-counted) resources with a three-star rating. These three-star resources represent a total of 22 individual resources. Table 15 lists the three-star resources and shows the policy areas that each resource addresses.

## 9.3 Information Gaps

Several gaps in the information compiled for this report are apparent, suggesting areas for additional research. These gaps include:

- *Limited or no resources were found for certain policy areas.* No resources were found for Policy Area #7, ensure SDM integrates with KM. Consequently, if ORD decides to consider policies and guidance related to KM, (e.g., ways to ensure that knowledge about scientific projects and data is retained when ORD scientists retire or leave EPA), then additional research will be required to identify KM best practices currently used by other agencies or organizations. Limited resources (14 of the 189 references) were found for Policy Area #1, manage scientific data as an enterprise asset or liability, and Policy Area #5, maintain version and change control on data sets (21 of the 189 references).

- *Several of the resources are more than ten years old.* For example, OSWER's Life Cycle Management Guidance and System Life Cycle Management Guidance documents provide a large amount of information directly related to Policy Area #1: manage scientific data as an enterprise asset or liability, Policy Area #3: identify scientific data with metadata to enable needed business operations, and Policy Area #4: manage scientific data for appropriate control. However, these documents were written in 1989 and 1992, respectively, so the information may be outdated.

- *Many of the resources provide only general information or information that is otherwise not explicitly relevant to ORD.* As indicated by the applicability ratings, the majority of references were rated as one-star or two-star documents, with only 22 individual documents being identified as providing highly relevant information.

- *The resources were identified from secondary sources.* The non-EPA documents and resources gathered for this report were discovered based on Internet research. EPA documents were found through searching both the Internet and the EPA Intranet. Consequently, to obtain a larger universe of SDM materials, it will be important to contact EPA and other federal agency representatives to identify documents that may have missed and any on-going projects related to SDM.

**Table 15. SDM documents and resources with three-star ratings by agency and policy area**

| Document/resource title and date | Policy #1 | Policy #2 | Policy #3 | Policy #4 | Policy #5 | Policy #6 | Policy #7 |
|---|---|---|---|---|---|---|---|
| **EPA** | | | | | | | |
| EPA Enterprise Architecture Target Data Architecture. 2009. | ● | ● | ● | | | ● | |
| Metadata Standards for the Enterprise Content Management Program. 2009. | | | ● | | | | |
| OEI. National Geospatial Data Policy. CIO Policy Transmittal 05-002. 2005. | ● | ● | | ● | | ● | |
| ORD. Scientific Data Management Strategy. 2007. | ● | ● | | | ● | ● | |
| OSWER Life Cycle Management Guidance. 1989 | ● | | ● | ● | | | |
| OSWER System Life Cycle Management Guidance. Part 3 Practice Paper: Data Modeling. 1992. | | | ● | | | | |
| **DOE** | | | | | | | |
| ARM Data Sharing and Distribution Policy. 2006. | | | | ● | | | |
| Guidelines for Archiving Data in the NARSTO Permanent Data Archive. 2006. | | | | | | ● | |
| Developing Data Management Policy and Guidance Documents for your NARSTO Program or Project. 2005 | ● | | | | | | |
| **NASA** | | | | | | | |
| Guidelines for Development of a Project Data Management Plan (PDMP). 1993. | ● | | | | ● | | |
| Heliophysics Science Data Management Policy. 2007. | | | ● | | | | |
| White Paper on NASA Science Data Retention. 2007. | | | | | ● | | |
| **NIH** | | | | | | | |
| NIH Data Sharing Policy and Implementation Guidance. Undated. | | | ● | | | | |
| Requirements for Data Management and Statistics for DAIDS Funded and/or Sponsored Clinical Trials. 2007. | | | | | ● | | |
| Guidelines for Developing a Manual of Operations and Procedures (MOP). 2007. | | | | | ● | | |
| **NOAA** | | | | | | | |
| Environmental Data Management at NOAA: Archiving, Stewardship, and Access. 2007. | | ● | | | | ● | |

**Table 15. SDM documents and resources with three-star ratings by agency and policy area (cont.)**

| Document/resource title and date | Policy #1 | Policy #2 | Policy #3 | Policy #4 | Policy #5 | Policy #6 | Policy #7 |
|---|---|---|---|---|---|---|---|
| **NSF** | | | | | | | |
| Division of Oceans: Data and Sample Policy. 2004. | | • | • | | | | |
| Data Archiving Policy. Undated. | | • | | • | | | |
| **Other** | | | | | | | |
| Interagency Working Group on Digital Data to the Committee on Science of the National Science and Technology Council. Harnessing the Power of Digital Data for Science and Society. 2009. | | | • | • | | | |
| Climate Change Science Program. Strategic Plan for Climate Change Science Final Report: Chapter 13. Data Management and Information. 2003. | | | • | | | • | |
| GAO. Climate Change Research: Agencies Have Data-Sharing Policies but Could Do More to Enhance the Availability of Data from Federally Funded Research. 2007. | | | | • | | | |
| U.S. Climate Change Science Program. Strategic Plan for Climate Change Science Final Report: Chapter 13. Data Management and Information. 2003. | | | • | | | • | |

## 9.4 Next Steps

The introduction to this report laid out a general, long-term approach for two broad goals: (1) developing a SDM policy framework and (2) developing policies, guidance, and tools that fit within this framework. Based on the findings of this report, the following steps are recommended (which can be conducted in the order below, a different order, or in parallel) for accomplishing the first goal of developing a SDM policy framework. It will be important to coordinate this work with the ongoing Strategic Action Plan being developed by OSIM and with OSIM's 2007 SDM Strategy.

- *Identify SDM policy areas to pursue*. This report presents resources related to seven different types of SDM policies. ORD can review the information on each policy area presented in this report to assess which are most relevant to its goals and objectives, and develop a schedule – or prioritization – for pursuing additional information on some or all of these policy areas.

- *Assess selected policy areas based on the resources compiled for this report.* Once ORD decides which of the policy areas it wishes to include in its policy framework and the relative schedule for assessing each area, the relevant documents and other resources assembled in this report can be reviewed and short "issue outlines" developed for each policy area. These outlines will focus on the two- and three-star resources identified in this report and additional research will be conducted as needed to try to fill in some of the gaps. The issue outlines could summarize, for example, examples of specific policies included under each policy area, issues and concerns related to each policy, and sample approaches for stating goals and principles and for developing guidance documents. This information could provide input for further data gathering and a series of SDM workshops, described below.

- *Conduct conversations with EPA staff.* Key ORD managers, scientists, and other researchers will be identified to ascertain their needs related to managing their scientific data and to obtain their input on key topics of discussion for the series of interagency workshops (see below). It will also be important to coordinate with EPA offices that are involved with scientific data issues (e.g., individual branches within OSIM) and developing data tools, including, for example, the proposed data set registry, the Science Inventory (SI), and the Registry for EPA Applications and Databases (READ).

- *Convene a series of workshops.* EPA/ORD could jointly sponsor a series of workshops with CENDI, the Interagency Working Group on Digital Data (IWGDD), and possibly other entities such as the Committee on Data for Science and Technology (CODATA) to bring EPA, other federal agency officials, and SDM experts together to discuss SDM requirements and best practices. These workshops could be organized by policy area and

would provide a forum for sharing information on managing scientific data. Workshop goals could also include (1) outlining a comprehensive framework of SDM policies and guidance that meets the needs of researchers, science managers, policy makers, and that general public, and conforms to current federal information and science policies, and (2) laying the foundation for the development of EPA digital scientific data policy. It will be important for the workshop agenda to include discussion of how to develop SDM policies that are consistent with current federal mandates for scientific data (e.g., data.gov and science.gov).

- *Develop SDM policy framework report(s).* This report or series of policy-related reports would be based on the findings and conclusions derived from the analysis of SDM resources, workshops, and conversations with EPA staff. It would present an initial SDM framework, based on best practices gleaned from EPA and the other science and technology agencies that participate in the SDM workshops. This framework would outline, by policy area, best practices related to defining goals and visions, principles, specific policies, and types of guidance and tools needed to convey SDM policies to the appropriate audiences. The report would also describe the issues that will require further analysis and illustrate the complexity of SDM and its potential role in supporting integrated, multidisciplinary, collaborative science.

The SDM policy framework report will suggest potential next steps for (1) future inter-agency collaborations regarding SDM and (2) EPA's development of its own SDM policies, guidance, and tools.

# A. Summary of Findings by Office and Policy Area – EPA

## U.S. Environmental Protection Agency (EPA)

| Office | Policy Areas | Level | Document Title/Date | Description | Applicability[a] | Link |
|---|---|---|---|---|---|---|
| EPA Privacy Policy | Manage scientific data for appropriate control | Specific Guidance | Procedures for Preparing Privacy Impact Assessments, 2009 | These procedures provide instructions for determining if Personally Identifiable Information is collected in systems and ensuring adequate controls are put in place. The Privacy Impact Assessment (PIA) is the tool required by the OMB for addressing privacy issues with electronic systems. No specific guidance is provided for completing the PIA, just the process for submitting them and having them reviewed and accepted. | ** | http://intranet.epa.gov/oei/imitpolicy/qic/ciopolicy/2151-p-04.pdf |
| EPA Privacy Policy | Manage scientific data for appropriate control | Specific Guidance | Procedures for Preparing and Publishing Privacy Act Systems of Records Notices, 2009 | These procedures provide the instructions for preparing a System of Records Notice (SORN). These procedures apply whenever information is retrieved by a name or personal identifier from records under the control of the Agency, regardless of format or location (i.e., systems, applications, databases, Web sites, filing cabinets). These procedures must be followed before collecting personal information on an individual and retrieving it by one of those elements. | ** | http://intranet.epa.gov/oei/imitpolicy/qic/ciopolicy/2151-p-03.pdf |
| EPA Records Schedule | Retain data commensurate with its value | Specific Guidance | EPA Records Schedule, Data Standards and Registry Service, 7/31/2009 | This schedule authorizes the disposition of the record copy in any media (media neutral), excluding any records already in electronic form. Records designated for permanent retention must be transferred to the National Archives in accordance with National Archives and Records Administration (NARA) standards at the time of transfer. (N1-412-08-15). This document provides guidance on what type of disposition is required for each type of media. | * | http://www.epa.gov/records/policy/schedule/sched/096.htm |
| EPA Region 9<br>Tribal Water Protection | Identify scientific data with metadata to enable needed business operations | General Guidance | National Tribal WQX/STORET Data Management, 11/18/2008 | This is a presentation on how to apply metadata to data for sharing purposes, emphasizing consistency. The examples given are for the Water Quality Exchange (WQX) and STORET and may not carry over to other projects. | * | http://www.epa.gov/region09/water/tribal/storet-training/pdf/WQXTemplate.pdf |
| Great Lakes National Program Office | Retain data commensurate with its value | Principles | Great Lakes Environmental Database, 6/17/2008 | The Great Lakes Environmental Database pages state: "Long after the studies are completed, the data remain and must be managed." | ** | http://www.epa.gov/glnpo/monitoring/data_proj/glenda/index.html |
| Great Lakes National Program Office | Identify scientific data with metadata to enable needed business operations | Specific Guidance | Lake Michigan Mass Balance Metadata, 3/9/2006 | The Metadata link offers some guidance on metadata reporting formats and sample naming. | ** | http://www.epa.gov/greatlakes/lmmb/metadata.html |
| Great Lakes National Program Office | Retain data commensurate with its value | General Guidance | Introduction to<br>Lake Michigan Mass Balance Data, 3/9/2006 | The database was developed under the following guidelines: develop a system having cross-program and -project utility, document the quality of all data populating the system, ensure that the resulting data base has long-term value, and avoid duplicating effort with other data systems. | ** | http://www.epa.gov/greatlakes/lmmb/database.html |
| National Health and Environmental Effects Research Laboratory (NHEERL) | Scientific data are enterprise assets or liabilities | Policies | NHEERL Data Management Policy and Practices: Genomics and Related High Throughput Data. | The policy states that data collected from human subjects presents a challenge in that sharing of data can only be done if the confidentiality of the subjects has been assured. Assurance must be obtained from the NHEERL Human Subjects Research Official prior to the entry of such data into a centralized data base. | ** | EPA, Undated |
| National Health and Environmental Effects Research Laboratory (NHEERL) | Manage scientific data for appropriate control | Policies | NHEERL Data Management Policy and Practices: Genomics and Related High Throughput Data. | The document provides data sharing guidelines. For example, it states that data access to scientists not directly involved in the original project team will be initially restricted. The data available to outside investigators upon data upload will be limited to a brief description of the experiment sufficient to determine the utility of the underlying data for other purposes. In addition, data should be available to all members of a collaborative unit, irrespective of the composition of that unit, as soon as it is generated and reviewed for accuracy. | ** | EPA, Undated |

## U.S. Environmental Protection Agency (EPA)

| Office | Policy Areas | Level | Document Title/Date | Description | Applicability[a] | Link |
|---|---|---|---|---|---|---|
| National Health and Environmental Effects Research Laboratory (NHEERL) | Retain data commensurate with its value | Policies | NHEERL Data Management Policy and Practices: Genomics and Related High Throughput Data. | The policy states that storage should be completed within 3 months after completion of primary data generation to allow for sufficient quality assurance of the raw data. If additional time is needed for quality assurance (QA) of the raw data, the length should be determined following discussion with the project lead and the appropriate administrator. | ** | EPA, Undated |
| Office of Air and Radiation<br><br>Emission Inventory Improvement Program (EIIP) Data Management Committee | Identify scientific data with metadata to enable needed business operations | General Guidance | EIIP Phase I Data Model, January 1999 | This document describes four views of the EIIP Data Model that provide common formats so data can be shared. It also provides a thorough data element dictionary, a list of entities and their attributes, and data model codes. | * | http://www.epa.gov/ttn/chief/eiip/techreport/volume07/vii01.pdf |
| Office of Air and Radiation<br><br>Emissions, Monitoring, and Analysis Division | Identify scientific data with metadata to enable needed business operations | General Guidance | Annual Air Quality Data Certifications for PM and Ozone Design Values, 6/12/2002 | This memorandum requires states and Tribes to document their annual air quality data sets so EPA can accurately interpret the reported data. States and Tribes must certify that prior year data are entered and the summary report is accurate. | * | http://www.epa.gov/ttn/amtic/files/ambient/pm25/datamang/designmem.pdf |
| Office of Environmental Information | Develop a scientific data management plan that covers the full data life cycle | General Guidance | EPA Enterprise Architecture Target Data Architecture, 6/23/2009 | In the framework presented in Section 4.3.1, EPA Program Offices that oversee Agency-wide business lines will ensure that quality-related activities associated with each phase of the EPA Data Lifecycle Framework (Figure 16) are documented. See also Appendix A. | *** | Email communication with Kevin Kirby. 7/14/09. |
| Office of Environmental Information | Develop a scientific data management plan that covers the full data life cycle | General Guidance | Guidance for Geospatial Data Quality Assurance Project Plans, March 2003 | This guidance document describes the type of information that would be included in a Quality Assurance (QA) Project Plan by anyone developing a geospatial project or using geospatial data for EPA. | ** | http://www.epa.gov/quality/qs-docs/g5g-final.pdf |
| Office of Environmental Information | Develop a scientific data management plan that covers the full data life cycle | Policies | Information Resources Management (IRM) Policy, Chapter 19 Information and Data Management, 2001 | Section 5, Policies, of this document lists EPA policies on information and data management. Note that this document has expired and has not yet been updated. | ** | http://www.epa.gov/irmpoli8/expiredpolicies/Chaptr19.PDF |
| Office of Environmental Information | Develop a scientific data management plan that covers the full data life cycle | Specific Guidance | Guidance on Systematic Planning Using the Data Quality Objectives Process, February 2006 | EPA has established a policy that states that before information or data are collected on Agency-funded or regulated environmental programs and projects, a systematic planning process must occur during which performance or acceptance criteria are developed for the collection, evaluation, or use of these data. This document provides specific guidance at each step of using the data quality objectives process. | ** | http://www.epa.gov/quality/qs-docs/g4-final.pdf |
| Office of Environmental Information | Identify scientific data with metadata to enable needed business operations | General Guidance | EPA Enterprise Architecture Target Data Architecture, 6/23/2009 | Section 4 examines the various components of data management that are critical at the enterprise level and must be addressed for enterprise architecture. Topics in this section address data quality, enterprise data security, metadata and master data management and data governance. | *** | Email communication with Kevin Kirby. 7/14/09. |
| Office of Environmental Information | Identify scientific data with metadata to enable needed business operations | General Guidance | EPA Enterprise Architecture Target Data Architecture, 6/23/2009 | Enterprise Metadata Architecture, Section 4.6. The enterprise metadata architecture proposed for EPA is a cross-cutting framework of policy, standards, communication, implementation, and continual evaluation required for enabling a consistent metadata capability. This document includes information on Metadata Standards and Policy Development (Section 4.6.1), Governance for Data and Metadata (Section 4.6.2), Communication and Outreach (Section 4.6.3), Implementation Assistance (Section 4.6.4), Lessons Learned and Performance Measures (Section 4.6.5). See also Appendix C. | *** | Email communication with Kevin Kirby. 7/14/09. |

## U.S. Environmental Protection Agency (EPA)

| Office | Policy Areas | Level | Document Title/Date | Description | Applicability[a] | Link |
|---|---|---|---|---|---|---|
| Office of Environmental Information | Manage scientific data for appropriate control | Policies | Agency-wide Quality System Documents, 12/30/2009 | This site provides links to potentially helpful documents, such as Overview of the EPA Quality System for Environmental Data and Technology, Guidance for Developing Quality Systems for Environmental Programs, Guidance on Systematic Planning using the Data Quality Objectives Process, Guidance for Preparing Standard Operating Procedures, Guidance on Environmental Data Verification and Data Validation, and Data Quality Assessment: A Reviewer's Guide. | * | http://www.epa.gov/quality/qa_docs.html |
| Office of Environmental Information | Manage scientific data for appropriate control | Policies | EPA Quality Manual for Environmental Programs, 5/5/2000 | This document discusses Requirements for Reporting Environmental Data. Section 2.5 covers requirements for reporting technical data; 2.6 covers QA and quality control (QC) requirements and guidance (mandatory and advisory). The document states: "The primary goal of the Agency-wide Quality System is to ensure that environmental programs and decisions are supported by data of the type and quality needed and expected for their intended use, and that decisions involving the design, construction, and operation of environmental technology are supported by appropriate quality assured engineering standards and practices. The scope of this Manual includes applicable environmental programs involving: the collection, evaluation, and use of environmental data by and for the Agency, and the design, construction, and operation of environmental technology by the Agency." | ** | http://www.epa.gov/irmpoli8/ciopolicy/2105-P-01-0.pdf |
| Office of Environmental Information | Scientific data are enterprise assets or liabilities | General Guidance | EPA Enterprise Architecture Target Data Architecture, 6/23/2009 | The successful management of information and data as an enterprise asset is of critical importance. To achieve the vision of maximizing the value of enterprise data assets, EPA will establish an Enterprise Data Architecture (EDA) Program to create a proactive, enterprise service organization focusing specifically on critical data management issues and challenges faced by EPA programs and their partners. | *** | Email communication with Kevin Kirby. 7/14/09. |
| Office of Environmental Information | Develop a scientific data management plan that covers the full data life cycle | General Guidance | Data Standards Policy, 6/28/2007 | This document states that: "All Agency information systems that exchange information shall implement applicable data standards in the most current version at the appropriate phase in the development life cycle but no later than the required implementation date specified in the standard unless a waiver has been obtained. When a new version of a standard is issued the old version is given a retirement date and should not be used after that date. Implementation of data standards or the appropriate waiver shall be described in the lifecycle and solution architecture documentation for each applicable EPA system and documented in the Registry of EPA Applications and Databases (READ) in conformance with the READ record maintenance schedule." | * | http://www.epa.gov/irmpoli8/ciopolicy/2133.0.pdf |
| Office of Environmental Information | Develop a scientific data management plan that covers the full data life cycle | Policies | Records Management, 12/11/2009 | This policy states: "The Records Management Policy establishes principles, responsibilities, and requirements for managing EPA's records to ensure that the agency is in compliance with federal laws and regulations, EPA policies, and best practices for managing records. This Agency-wide policy provides the framework for specific guidance and detailed operating procedures governing records management organization and implementation." | * | http://www.epa.gov/records/policy/index.htm |

## U.S. Environmental Protection Agency (EPA)

| Office | Policy Areas | Level | Document Title/Date | Description | Applicability[a] | Link |
|---|---|---|---|---|---|---|
| Office of Environmental Information | Identify scientific data with metadata to enable needed business operations | General Guidance | Metadata Standards for the Enterprise Content Management Program, last updated 7/2/09 | The purpose of these standards is to define a consistent set of required metadata elements for all applications participating in the enterprise content management program. These standards cover unstructured information, which includes but is not limited to documents and records, and applies to all EPA Programs, Regions, Labs and Offices. Specifically, these standards underscore the importance of a consistent, yet somewhat flexible, set of metadata elements for the effective and accurate classification, retrieval, management and use of unstructured information. This document provides examples of baseline metadata standards and some associated roles and responsibilities. | *** | Email communication with Lynne Petterson. 6/10/09. |
| Office of Environmental Information | Identify scientific data with metadata to enable needed business operations | Policies | Data Standards Policy, 6/28/2007 | This Data Standards Policy establishes principles, responsibilities, and requirements for the development, maintenance, and implementation of data standards within the jurisdiction of EPA. This policy discusses the use of common terminology and data elements for consistency and data sharing; the use of centralized registries of data elements, XML schema and code sets, based on approved data standards, and related roles and responsibilities. | ** | http://www.epa.gov/irmpoli8/ciopolicy/2133.0.pdf |
| Office of Environmental Information | Identify scientific data with metadata to enable needed business operations | Specific Guidance | Data Standards Implementation, 6/28/2007 | This document contains procedures establishing the key steps to follow for implementation of EPA data standards. It discusses procedures for the following areas: development of implementation guidance for a data standard, review/approval of implementation guidance for a data standard, conformance assistance, and conformance measurement. | ** | http://www.epa.gov/irmpoli8/ciopolicy/2133-p-3.pdf |
| Office of Environmental Information | Identify scientific data with metadata to enable needed business operations | Specific Guidance | Data Standards Maintenance, 6/28/2007 | This document contains procedures establishing the key steps to follow for maintenance and revision of EPA data standards and implementation guidance. It discusses procedures for the following areas: proposal for data standard and/or implementation guidance revision, development of minor and major data standard revisions, data standards review and approval procedures for major revisions. | ** | http://www.epa.gov/irmpoli8/ciopolicy/2133-p-2.pdf |
| Office of Environmental Information | Identify scientific data with metadata to enable needed business operations | Specific Guidance | Requesting Data Standards Conformance Waiver 6/28/ 2007 | This document contains procedures establishing the key steps to follow for requesting a data standard conformance waiver from EPA data standards. It discusses procedures for the following areas: types of waivers; determination of need; and submission, disposition and posting of a waiver. | ** | http://www.epa.gov/irmpoli8/ciopolicy/2133-p-4.pdf |
| Office of Environmental Information | Identify scientific data with metadata to enable needed business operations | Specific Guidance | Data Standards Development, 6/28/2007 | These procedures establish the key steps to follow for development and approval of EPA data standards. This document provides procedures for the following: Data standard proposal, development, and approval and draft data standards review. | * | http://www.epa.gov/irmpoli8/ciopolicy/2133-p-1.pdf |
| Office of Environmental Information | Retain data commensurate with its value | Policies | Enterprise Content Management Policy, 6/10/2009 | This policy establishes the EPA Enterprise Content Management Program. The Program advises EPA staff on how best to store data, how to apply established data and metadata resources and how to manage records in accordance with all federal and Agency records management statues, regulations, policies, procedures and standards. | *** | Email communication with Lynne Petterson. 6/10/09. |
| Office of Environmental Information | Develop a scientific data management plan that covers the full data life cycle | Other | IT Policy Mega-Matrix | The IT Policy Mega-Matrix is a master list of all the EPA IT policy documents (e.g., Policies, Procedures, Standards, and Guidance) that the Office of Technology Operations and Planning (OTOP) maintains. Page 12 contains SLC documents. | * | http://intranet.epa.gov/otop/itpolicy/IT_Policy_Mega-Matrix_Feb2009_external.pdf |
| Office of Environmental Information | Retain data commensurate with its value | Other | IT Policy Mega-Matrix | The IT Policy Mega-Matrix is a master list of the IT policy documents (e.g., Policies, Procedures, Standards, and Guidance) that OTOP maintains. Pages 22 -25 contain archived documents. | * | http://intranet.epa.gov/otop/itpolicy/IT_Policy_Mega-Matrix_Feb2009_external.pdf |

## U.S. Environmental Protection Agency (EPA)

| Office | Policy Areas | Level | Document Title/Date | Description | Applicability[a] | Link |
|---|---|---|---|---|---|---|
| Office of Environmental Information | Scientific data are enterprise assets or liabilities | Policies | National Geospatial Data Policy. CIO Policy Transmittal 05-002. 8/24/2005. | The policy requires that all EPA investment in geospatial data should be leveraged for enterprise use and managed through enterprise architecture guidance. | *** | http://www.epa.gov/esd/ggc/pdf/epa_natl_geo_data_policy.pdf |
| Office of Environmental Information | Develop a scientific data management plan that covers the full data life cycle | Policies | National Geospatial Data Policy. CIO Policy Transmittal 05-002. 8/24/2005. | The policy establishes specific requirements under which all EPA program offices and labs will adhere in the planning, collecting, acquiring, processing, documenting, storing, accessing, maintaining, and retiring of geospatial data. | *** | http://www.epa.gov/esd/ggc/pdf/epa_natl_geo_data_policy.pdf |
| Office of Environmental Information | Identify scientific data with metadata to enable needed business operations | Specific Guidance | National Geospatial Data Policy. Procedure for Geospatial Metadata Management. 10/25/2007. | According to the policy, Geospatial Data Stewards must create or update the metadata record for each acquired data set so that it meets the minimum requirements of the EPA Metadata Technical Specification. During the data storage and access phase, stewards must refer to the technical specification for data storage and access requirements. Maintenance responsibility for geospatial and metadata falls to the data owner or data steward of the program office or division. | ** | http://www.epa.gov/geospatial/docs/2131.pdf |
| Office of Environmental Information | Manage scientific data for appropriate control | Policies | National Geospatial Data Policy. CIO Policy Transmittal 05-002. 8/24/2005. | The policy states that geospatial data that is acquired by EPA (including contractors, grantees and vendors) must comply with all procedures and standards applicable to those data as if they were collected by EPA. | *** | http://www.epa.gov/esd/ggc/pdf/epa_natl_geo_data_policy.pdf |
| Office of Environmental Information | Retain data commensurate with its value | Policies | National Geospatial Data Policy. CIO Policy Transmittal 05-002. 8/24/2005. | The program office or project sponsoring the original collection effort is responsible for spatial data maintenance and decisions regarding ultimate retention and disposal. Data disposition for archiving must also comply with the records retention requirements of the program under which the data was collected. | *** | http://www.epa.gov/esd/ggc/pdf/epa_natl_geo_data_policy.pdf |
| Office of Environmental Information | Maintain version and change control on data sets | Recommendations for Policies | QIC Steering Committee – CIO Policy Consolidated Comments Form, 6/10/2009 | This is a steering committee form with reviewer comments regarding possible changes/clarifications to the following documents: Enterprise Content Management Policy, Metadata Standards for the Enterprise Content Management Program, and E-mail Records Procedures. | * | Email communication with Lynne Petterson. 6/10/09. |
| Office of Environmental Information | Manage scientific data for appropriate control | Specific Guidance | Procedures for Preparing Privacy Act Statements, 2009 | These procedures provide instructions for developing Privacy Act Statements (PAS) that must be provided to individuals when a federal agency requests personal information about them that is to be maintained in a system of records retrieved by name or personal identifier (5 U.S.C. 552a (e)(3)). These procedures list what to put in the PAS and provides a sample. This document lists the process the PAS goes through. | ** | http://intranet.epa.gov/oei/imitpolicy/qic/ciopolicy/2151-p-05.pdf |
| Office of Environmental Information | Retain data commensurate with its value | Specific Guidance | E-mail Records Procedures, 9/25/2009 | These procedures state: "E-mail is a significant means of conducting Agency business. As such, some e-mail messages qualify as Agency records and must be managed appropriately to successfully carry out the mission of EPA. Proper e-mail records management enables the Agency to meet its business needs and legal obligations, including responding to Freedom of Information Act (FOIA), litigation and other production requests. This document provides specific steps to maintain email records via ECMS or a paper recordkeeping system." | * | Email Communication. |
| Office of Research and Development (ORD) | Identify scientific data with metadata to enable needed business operations | General Guidance | Implementing the National Geospatial Data Policy: Lessons Learned | This document provides valuable lessons learned on data management policy implementation. Weaknesses noted include: metadata, infrastructure (e.g., network and systems interoperability re: metadata and data load), data management ("...apparent that WED needs to develop a process through which project data will be cataloged and disseminated through Environmental Information Management System (EIMS) (which has now been integrated with the Science Inventory). This issue is probably not unique to WED, and processes will need to be refined for implementation across ORD..." | ** | http://intranet.epa.gov/ospintra/Science%20Council/Related%20Docs/ORDNGDPPILOTS.pdf |

## U.S. Environmental Protection Agency (EPA)

| Office | Policy Areas | Level | Document Title/Date | Description | Applicability[a] | Link |
|---|---|---|---|---|---|---|
| Office of Research and Development (ORD) | Scientific data are enterprise assets or liabilities | Recommendations for Policies | Scientific Data Management Strategy. 2007. | This strategy states an objective to identify and prioritize SDM projects by determining where there are “hidden” data management projects, some of which add significant value to the agency. It identifies others as "pet projects" which add no valued. | *** | EPA, 2007 |
| Office of Research and Development (ORD) | Develop a scientific data management plan that covers the full data life cycle | Recommendations for Policies | Scientific Data Management Strategy. 2007. | The strategy states an objective to define an SDM organizational structure. The structure needs to be "tuned" to the specific needs of each L/C/O. | *** | EPA, 2007 |
| Office of Research and Development (ORD) | Maintain version and change control on data sets | Recommendations for Policies | Scientific Data Management Strategy. 2007. | The strategy suggests that EPA should establish standards, policies, and procedures for scientific data quality cleanup, change control, and audits. For example, if problems or issues arise with the quality of scientific data, there must be a defined set of guidelines to determine what actions to take. | *** | EPA, 2007 |
| Office of Research and Development (ORD) | Retain data commensurate with its value | Recommendations for Policies | Scientific Data Management Strategy. 2007. | The paper states that it is necessary to maintain scientific records for historical research and regulatory purposes. It notes that there are many conflicting data formats, making it difficult to retrieve and re-use the information they contain. Therefore, a policy should develop overall standards and guidelines for acceptable formats for long-term retention. The paper also suggests a records retention schedule to ensure that records are kept only as long as legally and operationally required and that obsolete records are retired or disposed of in a controlled manner. This strategy paper also discusses the need for a disaster recovery plan. | *** | EPA, 2007 |
| Office of Science Advisor | Identify scientific data with metadata to enable needed business operations | Principles | Assessment Factors, June 2003 | Section 2.2.3 e of this document asks: Is the complete data set accessible, including metadata, data-dictionaries and embedded definitions (e.g., codes for missing values, data quality flags and questionnaire responses)? Are there confidentiality issues that may limit accessibility to the complete data set? | * | http://www.epa.gov/OSA/spc/pdfs/assess2.pdf |
| Office of Solid Waste and Emergency Response | Develop a scientific data management plan that covers the full data life cycle | General Guidance | OSWER Life Cycle Management Guide, 1989 | This document provides suggestions for both the Initiation and Concept phase of information management by discussing Initiation Phase objectives, Concept Phase objects, decisions, activities, roles/responsibilities and the decision paper.<br><br>Pages 22 and 23 discuss the creation of a data management plan and what should be included. Page 26 discusses the data management plan and what should be included in the definition stage. Chapter 4 discusses the expansion of the data dictionary and data management plan. Chapter 10 details how all life cycle stages work together and/or overlap. | *** | http://www.epa.gov/oswer/oswerlcm.htm |
| Office of Solid Waste and Emergency Response | Develop a scientific data management plan that covers the full data life cycle | General Guidance | System Life Cycle Management Guidance Part 3 Practice Paper: Data Management During the Life Cycle, January 1989 | In the data management paper, Chapter 3 provides a high level review of the recommended approach for each step of the SLC. | *** | http://www.epa.gov/oswer/oswerlcm.htm |
| Office of Solid Waste and Emergency Response | Develop a scientific data management plan that covers the full data life cycle | Specific Guidance | System Life Cycle Management Guidance Part 3 Practice Paper: Data Management During the Life Cycle, 1989 | This paper describes data management during the SLC and provides guidance concerning major topics that should be addressed by project teams. Data management begins during the concept phase, proceeds as requirements are defined and software is implemented, and continues until the application system is terminated or replaced. The chapters include the following: Selecting a data management approach, overview of data management topics, data modeling activities, data design activities, data stewardship, data documentation activities, and terms/reference manual. This document provides a useful synopsis of much of the SLC Chapters 1-10. | ** | http://www.epa.gov/oswer/oswerlcm.htm |
| Office of Solid Waste and Emergency Response | Develop a scientific data management plan that covers the full data life cycle | Specific Guidance | System Life Cycle Reviews and Approvals, January 1989 | This document provides all the steps and information necessary to review and approve all stages of the SLC. | ** | http://www.epa.gov/oswer/docs/oswerlcm/00000018.pdf |

## U.S. Environmental Protection Agency (EPA)

| Office | Policy Areas | Level | Document Title/Date | Description | Applicability[a] | Link |
|---|---|---|---|---|---|---|
| Office of Solid Waste and Emergency Response | Identify scientific data with metadata to enable needed business operations | General Guidance | OSWER Life Cycle Management Guide, 1989 | In Chapter 3 of the OSWER Life Cycle Management Guide, Page 21, Exhibit 3-10 discusses the Requirements Data Dictionary and how it serves as a repository for metadata.<br><br>In Chapter 4, it is discussed that in the design phase, you should enter metadata in the design data dictionary documenting the physical design of each data base or data file. | *** | http://www.epa.gov/oswer/oswerlcm.htm |
| Office of Solid Waste and Emergency Response | Identify scientific data with metadata to enable needed business operations | Principles | System Life Cycle Management Guidance Part 3 Practice Paper: Data Management During the Life Cycle, 1989 | Accurate information about data is essential. Effective management of data collected by OSWER requires that accurate information about data (i.e., metadata) be kept. | *** | http://www.epa.gov/oswer/docs/oswerlcm/00000021.pdf |
| Office of Solid Waste and Emergency Response | Identify scientific data with metadata to enable needed business operations | Specific Guidance | System Life Cycle Management Guidance Part 3 Practice Paper: Data Modeling, May 1992 | This is a detailed document that includes topics such as: What are data models, creating data entities, data relationships and creating relationships between data entities, creating data elements, and changing the model. "This paper (1) introduces data modeling techniques; (2) defines specific data standards for logical data modeling to follow during the SLC; and (3) offers some "how to" guidance throughout the data modeling process." | *** | http://www.epa.gov/oswer/docs/oswerlcm/00000022.pdf |
| Office of Solid Waste and Emergency Response | Maintain version and change control on data sets | General Guidance | System Life Cycle Management Guidance Part 3 Practice Paper: Configuration Management, January 1989 | This document provides guidance regarding the implementation of CM, defined as systematically identifying the characteristics of a system and formally controlling any changes or additions to these items. The guidance describes specific activities associated with CM; describes project organization structures to accomplish CM; describes the documentation of project-specific CM activities in a CM plan. | ** | http://www.epa.gov/oswer/docs/oswerlcm/00000019.pdf |
| Office of Solid Waste and Emergency Response | Maintain version and change control on data sets | General Guidance | System Life Cycle Management Guidance Part 3 Practice Paper: Configuration Management, January 1990 | Chapter 2 of this paper discusses the establishment of configuration item identification, which acts as "labels" for the characteristics described in the documentation. This chapter also discusses change request impact analysis. | ** | http://www.epa.gov/oswer/docs/oswerlcm/00000019.pdf |
| Office of Solid Waste and Emergency Response | Maintain version and change control on data sets | Specific Guidance | System Life Cycle Management Guidance Part 3 Practice Paper: Configuration Management, January 1991 | Chapter 3 of this paper provides steps to implementing CM into an organization. | ** | http://www.epa.gov/oswer/docs/oswerlcm/00000019.pdf |
| Office of Solid Waste and Emergency Response | Manage scientific data for appropriate control | General Guidance | OSWER Life Cycle Management Guide Chapter 2, 1989 | Chapters 2-9 of the Life Cycle Management Guide provide suggestions as to how to properly manage information through the following phases: Definition, Design, Development, Implementation, Production, Evaluation and Archive. | *** | http://www.epa.gov/oswer/oswerlcm.htm |
| Office of Solid Waste and Emergency Response | Manage scientific data for appropriate control | General Guidance | System Life Cycle Management Guidance Part 3 Practice Paper: Data Management During the Life Cycle, 1989 | Chapter 2 discusses how to select the right data model based on the level of impact it will have in relation to data sharing, organizational impact and cost. | ** | http://www.epa.gov/oswer/oswerlcm.htm |
| Office of Solid Waste and Emergency Response | Scientific data are enterprise assets or liabilities | General Guidance | System Life Cycle Management Guidance Part 3 Practice Paper: Data Management During the Life Cycle, 1989 | This document states: "If you choose an approach that doesn't address data dictionary issues as part of a large, high impact project, you will increase the risk of time and cost overruns for your project." | *** | http://www.epa.gov/oswer/oswerlcm.htm |

## U.S. Environmental Protection Agency (EPA)

| Office | Policy Areas | Level | Document Title/Date | Description | Applicability[a] | Link |
|---|---|---|---|---|---|---|
| Office of Solid Waste and Emergency Response | Scientific data are enterprise assets or liabilities | Principles | System Life Cycle Management Guidance Part 3 Practice Paper: Data Management During the Life Cycle, 1989 | This document states: "Data is a valuable resource. Data is collected, stored, and used to support critical OSWER program activities and decisions, making accurate and timely data an important OSWER resource." OSWER data "is used to make decisions affecting public health and safety, environmental quality, and the use of public funds. Without this information OSWER could not perform its mission. The data collected, stored, processed and disseminated by OSWER systems are used to create the information OSWER needs to operate." | *** | http://www.epa.gov/oswer/oswerlcm.htm |
| Office of Solid Waste and Emergency Response | Scientific data are enterprise assets or liabilities | Principles | System Life Cycle Management Guidance Part 3 Practice Paper: Data Management During the Life Cycle, 1989 | Page 4 lists several benefits for increasing the focus on data management. | *** | http://www.epa.gov/oswer/oswerlcm.htm |
| Office of Solid Waste and Emergency Response (OSWER) | Develop a scientific data management plan that covers the full data life cycle | General Guidance | System Life Cycle Management Guidance Part 3 Practice Paper: Configuration Management, January 1989 | Exhibit 2-1 provides an overview of configuration management (CM) throughout a system life cycle (SLC). This is more for a "system" life cycle than "documentation" life cycle, but might still have some relevance. | ** | http://www.epa.gov/oswer/docs/oswerlcm/000000019.pdf |
| Office of Solid Waste and Emergency<br><br>Brownfields and Land Revitalization Technology Support Center | Develop a scientific data management plan that covers the full data life cycle | General Guidance | Management and Interpretation of Data Under a Triad Approach – Technology Bulletin, May 2007 | The triad approach produces flexible, but rigorous project plans; DM is key to rapid collection and analysis of data gathered. "A successful data management strategy depends on input not only from data management specialists but also from those who will be generating and using the data, including vendors, geoscientists, chemists, and other technical specialists. The data management plan must address how data from different sources will be integrated to support decisions." | * | http://www.brownfieldstsc.org/pdfs/Management_and_Interpretation_of_Data.pdf |
| Office of Water<br><br>Office of Wetlands, Oceans and Watersheds | Retain data commensurate with its value | General Guidance | Volunteer Stream Monitoring: A Methods Manual<br><br>Chapter 6 Managing and Presenting Monitoring Data, 11/30/2006 | This document stresses checking with data users to ascertain both how the data will be used, and processes/presentation formats. References STORET as best repository for data sharing. | ** | http://www.epa.gov/volunteer/stream/vms60.html |
| Office of Water<br><br>Office of Wetlands, Oceans and Watersheds | Manage scientific data for appropriate control | Specific Guidance | Volunteer Stream Monitoring: A Methods Manual<br><br>Chapter 6 Managing and Presenting Monitoring Data, 11/30/2006 | This chapter emphasizes the need to establish a method for data management and handling. It doesn't offer much guidance, however. | * | http://www.epa.gov/volunteer/stream/vms60.html |
| Western Regional Air Partnership (WRAP) | Develop a scientific data management plan that covers the full data life cycle | General Guidance | Comprehensive Data Management of WRAP Emissions Data, 2009 | This is a data management plan for emissions data that could be used as guidance for the creation of an ORD data management plan policy.<br><br>(Note: The Western Governors' Association and the National Tribal Environmental Council receive funding from EPA to administer and support the WRAP.) | * | http://www.epa.gov/ttn/chief/conference/ei18/session1/hoek.pdf |

a. The applicability rating is shown as one, two, or three stars (***). A one-star rating means that the information is related to ORD's scientific data management (SDM) policy framework, but is expected to be of limited value in developing its policies and guidance. A two-star rating means that the information is somewhat relevant to ORD's policy framework and/or is presented in limited detail. A three-star rating means that the information provided is directly applicable to ORD's SDM policy/guidance goals and could serve as a model for ORD.

## B. Summary of Findings by Office and Policy Area – Other Federal Agencies

## Department of Energy (DOE)

| Office/project | Title | Description | Enterprise assets and liabilities | Scientific data management plans (full life-cycle) | Identify scientific data with metadata | Manage data for control (Intellectual property) | Maintain version and change control | Data retention and data valuation | Knowledge management capture | Contacts | URL |
|---|---|---|---|---|---|---|---|---|---|---|---|
| ARM | ARM Data Sharing and Distribution Policy | The policy sets expectations and establishes procedures for sharing data acquired in the course of the Atmospheric Radiation Measurement (ARM) Program. | | | From the USGCRP data policy: Full and open sharing of the full suite of global data sets for all global change researchers.<br><br>All data sets acquired during an IOP or campaign will be made available to the ARM External Data Center for dissemination to users and forwarding to the ARM Archive. | ARM data are available to all participants on a free and open basis and are publishable upon receipt with acknowledgment of ARM as the source.<br><br>The policy states that researchers and participants may release their own preliminary data to whomever they wish and the preliminary data of other investigators with consent from the data's originator.<br><br>The automatic inclusion of a data originator as a co-author is not insisted upon in the ARM Program, but the source of any data should be clearly recognized either as a co-author or through an appropriate acknowledgment. | The ARM External Data Center and Archive will track data versions and ensure latest data versions are made available to data recipients. | | | | http://www.arm.gov/data/docs/policy |
| OSTI-LLNL | The State of Data Management in the DOE Research and Development Complex, 7/14-15/2004 | The report suggests guidelines that data-intensive programs and facilities may adopt to assure that data generated are effectively managed and made available. It recommends collecting and retaining data that might otherwise be lost to future scientists. | The report discusses issues such as data ownership and DOE rights of re-use compound the problem of how to manage resulting data. | The report states that DOE needs a department-wide policy that recognizes life-cycle data management. It recommends an umbrella policy for data generators, collectors, curators, and users. | The report states that metadata must be optimized for future retrieval, assimilation and re-use and professional staff of scientists are needed to manage data. | According to the report, issues such as data ownership and DOE rights of re-use compound the problem of how to manage resulting data. | | According to the report, a data management plan (DMP) would describe how data should be preserved, documentation needed to assure validation and future use, and funding/infrastructure needs to ensure longevity. It states that turning over data files is not mandatory by DOE (more guidelines on page 7, also recommends to look to DOE data centers on retention policies). | | Sharon Jordan (865) 576-1194 jordans@osti.gov | http://www.osti.gov/publications/2007/datameetingreport.pdf |
| OSTI-LLNL | Management of OSTI-LLNL Electronic Data | The document describes the controls for managing electronic data produced for OSTI-LLNL. It focuses on meeting their quality assurance plan (QAP) and covers (1) data accuracy, completeness and integrity, (2) data transfer, (3) data storage and maintenance, (4) equipment access and backup, (5) data security, and (6) submittal of data to technical data coordinator. | | | | | According to the document, electronic files may be converted from one software to another. Staff should include entry in scientific notebook indicating that a verification of file conversion has been conducted. It also includes detailed steps on data transfer. | | | | https://eed.llnl.gov/ymp/pdf/IM-317550-2.pdf |
| ORNL | Developing Data Management Policy and Guidance Documents for your NARSTO Program or Project | The document provides guidance for archiving data, data file format template, and guidance for data policies and plans. A compilation of data management policy and guidance documents for program and project use in developing data management plans are available at http://cdiac.ornl.gov/programs/NARSTO/about.html | The document states that at some point there is a legal obligation for data collected with government funds to be freely available. | The guidelines provides a data flow chart before, during, and after, field campaign. It mentions that a clear statement of the importance of the data collection and the flow of the data in the broadest possible context is needed.<br><br>In addition, advanced planning for archiving project data furthers efforts to identify, collect, and report consistent data and metadata and to facilitate timely data analysis, sharing, integration, and synthesis. | The guidance states that there must be a decision on whether investigators have an obligation to make data easy to use by others.<br><br>It also states that metadata should clearly state source of data, whether data are preliminary and for use only among the project or suitable for widespread dissemination and citation requirements.<br><br>ORNL uses a web-based inventory of project data using the existing ORNL metadata search and data retrieval system called Mercury. | The document states that copyrights are a difficult issue. Must allow instrument operator to reap rewards of efforts, but common good is served by sharing.<br><br>A policy must address how project will ensure IP rights are protected and co-authorship or credit is given to originators and investigators. | The guidelines state that a policy must provide standard names to identify the project, data files and data sets.<br><br>It mentions that data from NARSTO projects are formatted in the NARSTO Data Exchange Standard (a spreadsheet-compatible layout, which uses standardized and consistent metadata values)<br><br>The guidance recommends that policies adopt, adapt or refine model documents as appropriate with input from managers, investigators, modelers, and data coordinators. They must also addresses data validation and assigning quality levels. | The guidance recommends that managers ask about value of data: short-term (3-5 years) mid-term (10) or longer (20).<br><br>In addition, scientists are encouraged to document their data at a level sufficient to satisfy the well-known "20-year test". Someone 20 years from now, not familiar with the data or how they were obtained, should be able to find data of interest and then fully understand and use the data solely with the aid of the documentation archived with the data (NRC). | | Les Hook, hookla@ornl.gov (865) 241-4846 | http://cdiac.ornl.gov/programs/NARSTO/DM_develop_guide.pdf |
| ORNL | NARSTO Quality Systems Management Plan, 9/30/1999 | The document identifies the NARSTO program quality assurance and data management requirements and guidelines for ensuring NARSTO product credibility, reliability, accessibility and quality. | | The document provides a project plan and data archival process flow chart. | | | | | | | http://cdiac.ornl.gov/programs/NARSTO/pdf/qsmp_current_version.PDF |

## Department of Energy (DOE)

| Office/project | Title | Description | Enterprise assets and liabilities | Scientific data management plans (full life-cycle) | Identify scientific data with metadata | Manage data for control (Intellectual property) | Maintain version and change control | Data retention and data valuation | Knowledge management capture | Contacts | URL |
|---|---|---|---|---|---|---|---|---|---|---|---|
| ORNL | Guidelines for Archiving Data in the NARSTO Permanent Data Archive, 5/2/2006 | The document outlines how data are selected for archiving; identifies ways that Projects can foster archiving; lists items to consider when preparing data for archiving; and describes the archiving process. | | The document provides characteristics of project DMP that will result in successful data archiving. | | | | The document includes guidance that stresses characteristics of projects and data that are worthy or able to be well-archived.<br><br>According to the document, NARSTO encourages scientists to document their data at a level sufficient to satisfy the well-known "20-year test." | | Les Hook, hookla@ornl.gov (865) 241-4846 | http://cdiac.ornl.gov/programs/NARSTO/Guidelines_for_Archiving_NARSTO_Data.pdf |
| ORNL | The NARSTO Atmospheric Measurements Template, 4/29/2005 | The NARSTO Atmospheric Measurements Template replaces the former NARSTO Data Management Handbook, which is no longer available. | | | The Data Exchange Standard (DES) template is designed to help data originators create DES files. The worksheet titled Detailed metadata contains a possible layout and content of a companion detailed metadata document. | | Every regular measurement needs to have an associated NARSTO standard flag. "Dimensional" variables (those that indicate the setting for measurements, such as site, date, time, altitude, etc.)<br><br>The NARSTO QSSC maintains a list of standardized variable names for non-chemical variables, and lists of chemicals that have or do not have CAS numbers. | | | | ftp://narsto.esd.ornl.gov/pub/DES_metadata/var_names_web_sources/NARSTO_template_atmospheric_measurements.xls<br><br>Website link available at http://cdiac.ornl.gov/programs/NARSTO/qadocumentation.html |

## National Aeronautics and Space Administration (NASA)

| Office/project | Title | Description | Enterprise assets and liabilities | Scientific data management plans (full life-cycle) | Identify scientific data with metadata | Manage data for control (Intellectual property) | Maintain version and change control | Data retention and data valuation | Knowledge management capture | Contacts | URL |
|---|---|---|---|---|---|---|---|---|---|---|---|
| NASA's Heliophysics Data Environment (HPDE): Data and Services for the Heliophysics Great Observatory | NASA Heliophysics Science Data Management Policy, 6/25/2007 | The policy provides an overview of the components of the HPDE including: a timeline of significant events in the data lifecycle, guidelines for the preparation of Project Data Management Plans (PDMP), guidelines for the long-term serving and archiving of data, and a plan for keeping the Data Policy updated in light of changing technology and community needs. | The policy states that NASA observational data represent an asset that must be retained in a usable state into the indefinite future. | This policy provides a blueprint for the HPDE, tracing the data lifecycle from measurements to final archives.<br><br>Page 23 provides examples of information that are appropriate for each data provider to include in a PDMP. | According to the policy, the HPDE will benefit greatly from more conventional standards, but experience has shown that if these are imposed by bodies without community input they tend to be ignored. | | | The policy states that the NSSDC will ensure the maintenance of the permanent archive; the physical arrangements for such storage will be made in whatever manner is most economical, secure, and accessible.<br><br>In addition, NASA archives must have user advisory committees to advise on the likely future use and value of datasets candidate for resource-intensive renewal cycles. | | http://nssdc.gsfc.nasa.gov/Heliophysics/pdmp/ | http://hpde.gsfc.nasa.gov/Heliophysics_Data_Policy_2007June25.pdf |
| Office of Space Science and Applications | Guidelines for Development of a Project Data Management Plan (PDMP), March 1993 | The purpose of this document is to provide guidelines/template to assist NASA Project personnel in the preparation of plans for managing the data associated with their project. The document addresses the management of data from space science investigations, from the point of their reaching the ground, to their entry into permanent archives. | The document states that any agreements regarding exclusive rights to data for the PIs should be stated, with summary timelines for when the data will be released to the public. All data sets to be permanently archived should be identified in this section. | The document recommends that the Project Data Flow should be stated in the PDMP, including an overall functional Data Flow Diagram. This diagram should identify those facilities performing various functions as the project progresses through its various mission phases. Example is provided. | A section of the PDMP should identify and describe all data sets expected to be generated. This includes the science data itself, associated ancillary data and orbit/attitude data of the spacecraft.<br><br>Each PDMP should have a glossary of terms relevant to that project.<br><br>Each PDMP should have an acronym list of terms relevant to that project. | | The PDMP should illustrate the plans for modifications and updates to this document over time, and how those changes will be controlled. | Project data repositories are project specific, providing temporary storage for active data as it is being processed and analyzed. This section of the PDMP should address the requirements placed on the project data repositories.<br>The section should address how data will transition from project to permanent discipline archives. Table 8 provides a format for summarizing storage requirements by data set. Once archived, data sets and supporting information shall be periodically reviewed to assess their value for continued retention by NASA. | | | http://nssdc.gsfc.nasa.gov/nssdc/pdmp_guidelines_march93.rtf |
| National Space Science Data Center (NSSDC) | White Paper on NASA Science Data Retention, 8/6/2007 | This brief note addresses which NASA science data should be retained indefinitely, and the conditions under which certain data may and should be released. | | NASA archives must ensure the continuing preservation, accessibility, and usability of the data in their care. Plans for doing so should be spelled out in Archives' Operating Plans. | Projects must create and certify optimally standards-adherent definitive data sets, and accompanying material (documentation, ancillary data, software, etc.) as needed to make the data independently usable, | | Ensuring continuing data integrity and usability requires periodic data renewal cycles. Some such cycles will involve only bit migration from old to new media. | Datasets leading up to the production of the definitive dataset should be retained only to a point six months past the creation and certification of the definitive dataset.<br><br>Derived datasets should be retained as long as they remain scientifically viable (i.e., algorithms or coefficients used in their derivation remain credible) and the cost of regenerating them (for some anticipated request level) outweighs the cost of their retention and maintenance.<br><br>NASA archives must have user advisory committees to advise on (among other things) the likely future use and value of datasets candidate for resource-intensive renewal cycles. | | Ed Grayzeck<br><br>Ed Bell | http://nssdc.gsfc.nasa.gov/nssdc/data_retention.html |
| Consultative Committee for Space Data Systems | Reference Model for an Open Archival Information System, January 2002 | This is a technical recommendation for use in developing a broader consensus on what is required for an archive to provide permanent, or indefinite long-term, preservation of digital information.<br><br>In addition, it establishes a common framework of terms and concepts that comprise an Open Archival Information System (OAIS). | | The purpose of this reference model is to facilitate a much wider understanding of what is required to preserve and access information for the long term.<br><br>It provides a data flow diagram that represents the operational OAIS archive external data flows. This diagram concentrates on the flow of information among producers, consumers and the OAIS and does not include flows that involve Management. | | If the word processing format is proprietary, and it can't be acquired even to the level of simply viewing the document, it may be necessary to migrate the document to a non-proprietary format to ensure its long term preservation.<br><br>Some projects have one-year proprietary periods before data in released to the science community. The Planetary Data System (PDS) policy is to avoid receipt of any proprietary data sets during the proprietary period. | It addresses the migration of digital information to new media and forms, the data models used to represent the information, the role of software in information preservation, and the exchange of digital information among archives. | The information being maintained has been deemed to need long term preservation, even if the OAIS itself is not permanent. Long term is long enough to be concerned with the impacts of changing technologies, including support for new media and data formats, or with a changing user community. | | | http://public.ccsds.org/publications/archive/650x0b1.pdf |

## National Aeronautics and Space Administration (NASA)

| Office/project | Title | Description | Enterprise assets and liabilities | Scientific data management plans (full life-cycle) | Identify scientific data with metadata | Manage data for control (Intellectual property) | Maintain version and change control | Data retention and data valuation | Knowledge management capture | Contacts | URL |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Jet Propulsion Laboratory | NARSTO Quality Systems Management Plan, 9/30/1999 | The document describes the Cassini / Huygens Program plan for generating, validating, and delivering data products to the PDS. Included are the policies, guidelines, and requirements that will be followed by instrument teams in the generation of PDS compliant archives. It provides a high-level description of science and SPICE datasets, data size estimates, and a delivery schedule that can be used by the PDS for planning purposes. | According to the document, PDS archives will be accessible to the public on-line. The PDS on-line system will provide search filters, such as time range or target name, so that a user can retrieve data that meet specific search criteria. NSSDC is responsible for filling large delivery orders to the science community, and making data available to foreign investigators, educators, and the general public. | Archive policies, guidelines and requirements have been developed to ensure data products meet PDS standards and support collaborative studies among Cassini Orbiter and Huygens Probe data. | PDS labels and index files provide searchable keys and describe characteristics of the products. Index files are used to populate the PDS search catalog.<br>There should be a consistent representation of time used in filenames, directory names, labels, and index files.<br><br>The document states that the PDS Discipline Node (DN) assigned to an instrument team coordinates and leads a peer review of a sample volume. Members of the PSG will be asked to participate in peer reviews as well as members of the science community outside the PSG. The peer review is used to ensure the archive contains all the components needed to perform science analysis, and is prepared as documented in the Software Interface Specification (SIS). | | The document states that filenames will adhere to ISO 9660 level 2 specifications that allow the total filename length of 31 characters.<br><br>There is no official reference for the NASA product level descriptions. The source for the NASA product level descriptions found below was taken from a Mars project archive plan. | The NSSDC ensures the long-term preservation of data. | | Other PDMP information found at http://nssdc.gsfc.nasa.gov/archive/pdmp/ | http://trs-new.jpl.nasa.gov/dspace/bitstream/2014/14261/1/00-0674.pdf |

## National Institutes of Health (NIH)

| Office/project | Title | Brief description | Enterprise assets and liabilities | Scientific data management plans (full life-cycle) | Identify scientific data with metadata | Manage data for control (Intellectual property) | Maintain version and change control | Data retention and data valuation | Knowledge management capture | Contacts | URL |
|---|---|---|---|---|---|---|---|---|---|---|---|
| National Institute on Aging | Guidelines for Developing a Manual of Operations and Procedures (MOP), 12/27/2007 | Sets forth guidelines to providing a MOP template for Principal Investigators of multi-site clinical trials. The role of the MOP is to facilitate consistency in protocol implementation and data collection across participants and study sites. | | Guidelines for program investigators of multi-site clinical trials to follow when preparing MOPs. MOPs are intended to facilitate consistency in protocol implementation and data collection, and are prepared before the study begins. The guidelines most relevant to Office of Research and Development (ORD) include data flow (e.g., data flow, data entry, data correction), data retention, data management, study completion and closeout procedures, confidentiality). The guidelines ensure that all forms are complete, intact, and transmitted to the data manager in a single site study or to the Coordinating Center, as appropriate. More recently, in some studies, data are directly entered into an electronic CRF (eCRF).<br><br>In addition, a Users Guide may need to be developed as a separate document to aid the study staff with data management tasks. | | The guidelines discuss the safeguards which have been put in place by the Steering Committee to ensure participant confidentiality and data security.<br><br>P 22: a list of study participant confidentiality safeguards. | The guidelines mention updating which is described as correcting data and maintaining an audit trail of all data changes. | The guidelines state that a MOP must specify the length of time all study files are to be maintained. NIH policy requires that studies conducted under a grant retain participant forms for three years, while studies conducted under contract must retain participant forms for seven years. Individual Institutional Review Boards (IRBs), institutions, states, and countries may have different requirements for record retention. The MOP should also briefly outline the study completion and close-out procedures. | | | http://www.nia.nih.gov/NR/rdonlyres/AEC5CE46-96E1-43D9-BA77-BAE8BF0D6CDC/0/ManualofProceduresMOPFinal1.doc<br><br>Webpage link available at: http://www.nia.nih.gov/ResearchInformation/CTtoolbox/ |
| Division of Acquired Immune Deficiency Syndrome (DAIDS) Clinical Research Policies and Standard Procedures Documents | Requirements for Data Management and Statistics for DAIDS Funded and/or Sponsored Clinical Trials, 2/5/2007 | Provides several documents that briefly discuss data management requirements. | | The requirements state that following clinical trial data management requirements must be met in order to ensure the authenticity and integrity of data.<br><br>It describes the processes and methods that data collection sites and central data management facilities must develop to manage their data, including: data management operations, overall data management system, data storage, database closure and archiving, and data audits. | According to the requirements, clinical trial data need to be managed in such a way as to ensure the authenticity and integrity of the data elements collected and to comply with applicable regulations and International Conference on Harmonization (ICH) Good Clinical Practice (GCP) guidelines. | | The requirements state that policies must develop change control procedures to ensure quality control in changes made to the data collection tools. Include how changes are requested, how the impact of changes is assessed, who is responsible for authorizing the changes, how the changes are tested and released, and how the changes are documented. | The requirements state that policies must develop a plan for record retention, both electronic and hard copy. It must include when record retention begins, the length of time the records are retained, where the records are retained, the security of the storage space, who has access to the storage space, and who is responsible for approving access. | | | http://www3.niaid.nih.gov/LabsAndResources/resources/DAIDSClinRsrch/PDF/DataMgt_StatPolicy.htm<br><br>Webpage link available at: http://www3.niaid.nih.gov/LabsAndResources/resources/DAIDSClinRsrch/DataManagement.htm |
| National Heart, Lung, and Blood Institute (NHLBI) | Policy for Dataset Preparation, 10/1/2005 | This updated policy provides information on data management and responsibility relating to data set requests and procedures for protection of privacy. | According to the policy, their full value of data can only be realized if they are made available, under appropriate terms and conditions consistent with the informed consent provided by individual participants, in a timely manner to the largest possible number of qualified investigators.<br><br>In addition, all investigators seeking access to data from NHLBI-supported studies that are in the possession of the Institute must execute and submit with their requests the appropriate standard Distribution Agreement for each study. | | The policy states that documentation for data sets must be comprehensive and sufficiently clear to enable investigators who are not familiar with a data set to use it. The documentation must include data collection forms, study procedures and protocols, descriptions of all variable recoding performed, and a list of major study publications. | | | | | | https://biolincc.nhlbi.nih.gov/new_data_set_policy/ |
| Office of Chief IT | NIH Enterprise Conceptual Data Model v1.0., January 2007 | The documents provides a specification of the key data entities that support NIH's business processes and provides an overarching framework to organize more detailed data architecture efforts and provide a common taxonomy for describing data assets across NIH. | | | | | | | | John Sharp Demetrios Kostikopoulos: kotsikod@mail.nih.gov | http://enterprisearchitecture.nih.gov/NR/rdonlyres/5D3017EA-22C1-4BCC-8E0F-79EB7B5C797A/0/NRFC0025.pdf |
| NIH Enterprise Architecture | NARSTO Quality Systems Management Plan, 9/30/1999 | The document specifies NIH architecture best practice for the NIH community for AD Attribute Data Content Management and requests suggestions for improvements. | | | Lists user attribute data content management rules. | | | | | | http://enterprisearchitecture.nih.gov/NR/rdonlyres/8B8AFA60-68A1-4155-A08F-03163B610E39/0/NIHRFC0008ActiveDirectoryAttributeDataContentandManagement.pdf |

## National Institutes of Health (NIH)

| Office/project | Title | Brief description | Enterprise assets and liabilities | Scientific data management plans (full life-cycle) | Identify scientific data with metadata | Manage data for control (Intellectual property) | Maintain version and change control | Data retention and data valuation | Knowledge management capture | Contacts | URL |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Office of Extramural Research | NIH Data Sharing Policy and Implementation Guidance, 3/5/2003 | The policy provides guidance on data sharing and additional information on the implementation of the NIH data policy. | | According to the policy, the precise content and level of detail to be included in a data-sharing plan depends on several factors, such as whether or not the investigator is planning to share data, and the size and complexity of the dataset. | The policy states that data sharing promotes many goals of the NIH research endeavor. It is particularly important for unique data that cannot be readily replicated.<br><br>It also mentions that final research data are recorded, factual material that must be commonly accepted in the scientific community as necessary to document, support, and validate research findings. This does not mean summary statistics or tables; rather, it means the data on which summary statistics and tables are based.<br><br>Regardless of the mechanism used to share data, each dataset will require documentation. | The rights and privacy of human subjects who participate in NIH-sponsored research must be protected at all times. It is the responsibility of the investigators, their IRB, and their institution to protect the rights of subjects and the confidentiality of the data.<br><br>It is appropriate for scientific authors to acknowledge the source of data upon which their manuscript is based. Many investigators include this information in the methods and/or reference sections of their manuscripts. | Investigators sharing data under their own auspices should consider using a data-sharing agreement to impose appropriate limitations on users. | Because the value of data often depends on their timeliness, data sharing should occur in a timely fashion. NIH expects the timely release and sharing of data to be no later than the acceptance for publication of the main findings from the final dataset. | | | http://grants.nih.gov/grants/policy/data_sharing/data_sharing_guidance.htm |
| National Cancer Institute (NCI), Division of Cancer Prevention (DCP) | NCI, DCP Data Management Requirements, October 2003 | A short guideline document specific to cancer research. | | The document states that a Data Management Plan (DMP) is a document prepared by the Consortium Principal Investigator and approved by the NCI, DCP. | The DMP should document the rules for handling data ranges, data types and coding of missing data. | | | According to the document, the DMP should include a description of the security plan and should delineate responsibilities and expected behavior of all individuals who have access to study data and systems.<br><br>Also indicates how long the records will be retained and when the process begins. | | | http://prevention.cancer.gov/files/clinical-trials/DataMgmtRqmts.doc<br><br>Webpage available at: http://prevention.cancer.gov/clinicaltrials/management/consortia/step-2/data |

## National Oceanic and Atmospheric Administration (NOAA)

| Office/project | Title | Description | Enterprise assets and liabilities | Scientific data management plans (full life-cycle) | Identify scientific data with metadata | Manage data for control (intellectual property) | Maintain version and change control | Data retention and data valuation | Knowledge management capture | Contacts | URL |
|---|---|---|---|---|---|---|---|---|---|---|---|
| NOAA | NOAA Report to Congress on Data and Information Management, October 2005 | This biannual report was sent to Congress on the state of data management in NOAA.<br><br>The report assembled 12 survey questions focused on a holistic, end-to end observation and data management approach, addressing five data management, archive, and distribution areas identified in Section 106 of the Public Law 102-567, Data and Information Systems. | | According to the report, NOAA is in the initial stages of developing and implementing an integrated data management system, based on common IT architecture and common processes. | The report states that NOAA faces a major challenge in enabling interoperability between legacy systems and emerging data systems. This lack of system interoperability, across NOAA and across agencies, hampers the collaborations enabled by technological gains.<br><br>It adds that integration and interoperability will be achieved through common protocols, hardware, and software, as well as the use of data and metadata standards. NOAA has begun this process by adopting a common enterprise-wide IT architecture. | | | | | | http://www.ngdc.noaa.gov/noaa_pubs/pdf/NOAA_Congress2005.pdf |
| Administrative Management and Executive Secretariat | NOAA Administrative Order: 216-101. Ocean Data Acquisitions, 7/9/1990 | The order establishes policies and procedures to ensure that NOAA ocean data support multiple uses of those data for purposes other than those for which they originally were collected. | The order states that retrospective access to data is required by the research community, climate/global change activities, and cartographic activities through designated national data management centers. | This order defines certain responsibilities of and procedures for all NOAA activities, including reimbursable programs for other agencies and NOAA funded contracts and grants, that involve the collection and archiving of ocean data from the open-ocean, Great Lakes, coastal waters, and estuaries. | The order states that NOAA managers of programs that conduct ocean data collection activities are responsible for assuring that data and related information with high utility for other users are available in a timely manner at national processing centers and national data centers and are documented and archived in designated national data management centers.<br><br>In addition, data submitted to the national data management centers are to be submitted via computer-compatible digital media when possible rather than as printed reports. Documentation must include information sufficient to fully describe the physical recording technique, data format, recording mode, blocking factor, and other pertinent items. | The order suggests that managers should work with their principal investigators to assure that other data, which may not be appropriate for archival at national centers, are documented and archived within the established period of time at the principal investigator's or an associated institution so these data will be available for other uses upon request. | | The order adds that data are used weeks to decades after the initial data acquisition. These archived data sets usually have more stringent quality requirements than real-time data. | | | http://www.corporateservices.noaa.gov/~ames/NAOs/Chap_216/naos_216_101.html |
| Administrative Management and Executive Secretariat | NOAA Administrative Order 212-15. Management of Environmental and Geospatial Data and Information, 12/2/2008 | The order establishes a policy for acquiring, integrating, managing, disseminating, and archiving environmental and geospatial data and information obtained from worldwide sources to support NOAA's mission. | | The order states that NOAA data management planning will include end-to-end data stewardship.<br><br>In addition, the NOAA Chief Information Officer (CIO) must develop a data management plan in coordination with the appropriate data center, specifying the data life cycle and disposition of data and information for each program. | | The order states that managers should maintain a list of applicable reference materials and will provide access to their electronic editions on the define website. | The order states that managers need to be alert to and mitigate the risks caused by changes of instruments, platforms, locations, and methods for observing or processing data. | | | | http://www.corporateservices.noaa.gov/~ames/NAOs/Chap_212/naos_212_15.html |
| NOAA/National Research Council | Environmental Data Management at NOAA: Archiving, Stewardship, and Access. | NOAA asked the National Research Council to help determine which observations, model outputs, and other environmental information should be preserved in perpetuity and made readily accessible and which data have a limited storage lifetime and easier accessibility requirements. This report suggests nine general principles for the effective management of environmental data and specific guidelines and examples illustrating how NOAA could apply these principles. | | Principle #7: Effective data management requires a formal, ongoing planning process. NOAA should establish and codify an enterprise-wide data management plan (elements of plan listed on p. 87-88).<br><br>Principle #2: Data-generating activities should include adequate resources to support end-to-end data management. | Principle #6: Scientific data stewardship, with assigned organizational responsibility, should be applied to all environmental data sets and their associated metadata to ensure that this information is preserved, remains continually accessible and can be improved as future discoveries build understanding and knowledge.<br><br>Guideline: Metadata that adequately document and describe each archived data set should be created and preserved to ensure the enhancement of knowledge.<br><br>Guideline: NOAA and partners should continue to expand use of standards and reference models. | | Guideline: establish and maintain data and metadata migration plans for all current and future long-tem archive systems to adapt to information technology evolution. | Guideline: develop and maintain scalable and reliable infrastructure that ensures long-term access and preservation of data assets.<br><br>Principle #9: A formal, ongoing process with broad community input is needed to decide what data to archive and what to dispose.<br><br>Guideline: It may be cost-effective to regenerate certain kinds of environmental data on demand. | | | http://www.nae.edu/nae/naepcms.nsf/weblinks/MKEZ-79CSA3?OpenDocument |

## National Oceanic and Atmospheric Administration (NOAA)

| Office/project | Title | Description | Enterprise assets and liabilities | Scientific data management plans (full life-cycle) | Identify scientific data with metadata | Manage data for control (intellectual property) | Maintain version and change control | Data retention and data valuation | Knowledge management capture | Contacts | URL |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | Principle #5: Metadata are essential for data management<br><br>Guideline: Stewardship requires systematic, ongoing assessment and improvement of data. Stewardship plans should be consistent but flexible so improvements in data and metadata are captured.<br><br>Guideline: data should be made available to users in a timely manner and should be accessible with as few barriers as possible (administrative, technological, and systematic barriers are described). A distributed data access structure can support improved data discovery and seamless integration. | | | Principle #1: Environmental data should be archived and accessible.<br><br>Guideline: archiving and access decisions are closely related. When resources are limited, access to older or less commonly used data should be scaled back, rather than removing data from archive. | | | |
| | | | | | Principle #8: And effective data archive should provide for discovery, access and integration.<br><br>Guidelines: environmental data should be easily discoverable by a broad range of users. Data discovery should not require any specific knowledge about the data or how they are managed.<br><br>Search tools and other discovery-enhancing features could be improved at many environmental data access points and by the use of expanded metadata (detailed list provided on p 75-76). | | | | | | |
| NOAA/National Environmental Satellite, Data, and Information (NESDIS) | Data Management Systems And Tools | A key recommendation of the 2000 Report of the President's Panel on Ocean Exploration was for NOAA to establish a broad-based data management task force to design and implement an integrated and comprehensive data management system, which would facilitate data sharing across a broad, multidisciplinary community. In October 2002, NESDIS formed an Integrated Product Team in partnership with NOAA's Ocean Exploration staff and other NOAA and non-NOAA partners. | | Office only offers tools for accessing and managing data. There are no guidelines for data management. | | | | | | Phone: 301-713-3578 | http://www.explore.noaa.gov/data-management |
| NWS Telecommunication Operations Center | What Does Data Management Provide to You?, 3/24/2010 | Website provides a variety of tools and resources, such as: Data Management Notices, Data Management Customer Relationship Management System (CRM), and NWS Communication Identifiers. No guidelines. | | | | | | | | | http://www.weather.gov/datamgmt/ |

## National Science Foundation (NSF)

| Office/project | Title | Description | Enterprise assets and liabilities | Scientific data management plans (full life-cycle) | Identify scientific data with metadata | Manage data for control (intellectual property) | Maintain version and change control | Data retention and data valuation | Knowledge management capture | Contacts | URL |
|---|---|---|---|---|---|---|---|---|---|---|---|
| National Science Foundation | Long-Lived Digital Data Collections: Enabling Research and education in the 21st Century, 2005 | The report presents the findings and recommendations from an analysis of the policy issues relevant to long-lived digital data collections. This analysis included a study of data management practices across federal science agencies (done in 2004), which is summarized in this report.<br><br>It also provides summary of current policies at NSF and other agencies on data sharing and archiving. | | The reports describes in-house process (NOAA and NASA) and collections performed by external organizations (NSF and NIH).<br><br>It states that the contents of the data management plan should include: the types of data to be authored; the standards that would be applied for format, metadata content, etc.; provisions for archiving and preservation; access policies and provisions; plans for eventual transition or termination of the data collection in the long term future.<br><br>The report identifies and describes the roles of key actors in digital data collections, key contents of a data management plan. | The report provides data access/release guidelines. | The report states that NSF expects significant findings from research and education activities it supports to be promptly submitted for publication, with authorship that accurately reflects the contributions of those involved. | | According to the report, the vast majority of NSF support carries with it no long-term commitment. Principal investigator grants have a duration of several years. Centers are typically funded for five years with a potential for an additional five years of funding. Long-lived digital data collections raise a new issue. It is timely for NSF to consider whether it should make very long-term commitments to a digital collection. | | | http://www.nsf.gov/pubs/2005/nsb0540/ |
| Division of Ocean Sciences | Division of Ocean Sciences: Data and Sample Policy, 11/3/2003 | The Data and Sample Policy highlights the General Data Policy governing how principal investigators submit and manage their data. | | In addition, focused programs supported by NSF's Division of Ocean Sciences may establish more stringent data submission procedures to meet the needs of such programs. Principal Investigators supported by these programs are required to follow these data submission procedures. | The policy states that annual reports, required for all projects, should address progress on data and research product sharing.<br><br>The policy states that where no data or sample repository exists for the collected data or samples, metadata must be prepared and made available. The principal investigator is required to address alternative strategies for complying with the general philosophy of sharing research products and data as described above. | | | According to the policy, principal investigators are required to submit all environmental data collected to the designated National Data Centers as soon as possible, but no later than two years after the data are collected.<br><br>The encouragement of digital preservation programs explicitly aimed at facilitating sustained access. | | | http://www.nsf.gov/pubs/2004/nsf04004/nsf04004_1b.htm |
| Social, Behavioral and Economic Sciences (SES) | Data Archiving Policy, 11/8/2008 | The web site highlights the Data Archiving Policy, the purpose of which is to advance science by encouraging data sharing among researchers. It provides guidelines for archiving data for several different categories of data, including: quantitative social and economic data, qualitative information, experimental research, and mathematical and computer models. | | This policy explicitly recognizes that many complexities arise across the range of data collection supported by SES programs, and that unusual circumstances may require modifications or even full exemptions. For example, human subjects protection requires removing identifiers, which may be prohibitively expensive or render the data meaningless in research that relies heavily on extensive in-depth interviews. | The policy states that if it is appropriate for other researchers to have access to them, the investigators should specify a time at which they will be made generally available, in an appropriate form and at a reasonable cost. | According to the policy, intellectual property rights may be at risk in some forms of data collection. The policy is intended to be flexible enough to accommodate the variety of scientific enterprises that constitute SES programs. No comprehensive set of rules is possible. | The policy states that the kinds of qualitative information collected in research projects supported by SES can range from microfilms and other copies of very old documents to oral interviews and video tapes about historical events in science or about contemporary technological controversies. They can consist of hand written records of open-ended interviews. Investigators should consider whether and how they can develop special arrangements to keep or store these materials so that others can use them. | | | | http://www.nsf.gov/sbe/ses/common/archive.jsp |
| Division of Earth Sciences (EAR) | Implementation of the NSF Data Sharing Policy, April 2002 | The statement provides guidelines for implementing NSF's data sharing policy. The purpose of the statement is to ensure open access to quality data for Earth Science research and education. | According to the document, it is the responsibility of researchers and organizations to make results, data, derived data products, and collections available to the research community in a timely manner and at a reasonable cost. In the interest of full and open access, data should be provided at the lowest possible cost to researchers and educators. | The document states that within the proposal review process, compliance with data guidelines will be considered in the Program Officer's overall evaluation of a principal investigator's record of prior support. Exceptions to these data guidelines require agreement between the principal investigator and the NSF Program Officer. | | The document states that data may be made available for secondary use through submission to a national data center, publication in a widely available scientific journal, book or website, through the institutional archives that are standard for a particular discipline, or through other EAR-specified repositories. | The document recommends that data inventories are published or entered into a public database periodically and when there is a significant change in type, location or frequency of such observations. | According to the document, preservation of all data, samples, physical collections and other supporting materials needed for long term earth science research and education is required of all EAR-supported researchers.<br><br>In addition, for those programs in which selected principle investigators have initial periods of exclusive data use, data should be made openly available as soon as possible, but no later than two (2) years after the data were collected. This period may be extended under exceptional circumstances, but only by agreement between the principal investigator and the NSF. For continuing observations or for long-term (multi-year) projects, data are to be made public annually. | | | http://www.nsf.gov/geo/ear/EAR_data_policy_204.pdf |
| Office of Polar Programs (OPP) | NARSTO Quality Systems Management Plan, 9/30/1999 | NSF's policy "expects investigators to share with other researchers, at no more than incremental cost and within a reasonable time, the data, derived data products, samples, physical collections and other supported materials gathered or created in the course of the research project." | | | According to the guidelines, OPP considers the documentation of data sets, known as metadata, as vital to the exchange of information on polar research and to a data set's accessibility and longevity for reuse.<br><br>In addition, data archives of OPP-supported projects should include easily accessible information about the data holdings, including quality assessments, supporting ancillary information, and guidance for locating and obtaining the data. | The guidelines state that principal investigators should make their data available to all reasonable requests (as specified in the NSF Grant Proposal Guide, Section VII-H) and where applicable the principal investigators should submit the data collected to designated data centers as soon as possible, but no later than two years after the data are collected. | | | | | http://www.nsf.gov/pubs/1999/opp991/opp991.doc<br><br>Website available at: http://www.nsf.gov/publications/pub_summ.jsp?ods_key=opp991 |

## Other Federal Agencies

| Office/project | Title | Description | Enterprise assets and liabilities | Scientific data management plans (full life-cycle) | Identify scientific data with metadata | Manage data for control (Intellectual property) | Maintain version and change control | Data retention and data valuation | Knowledge management capture | Contacts | URL |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Climate Change Science Program (CCSP) | Strategic Plan for the Climate Change Science Program Final Report: Chapter 13. Data Management and Information, July 2003 | The chapter introduces the objectives for data management to be addressed in the coming decade based upon current knowledge and infrastructure including: collecting and managing data in multiple locations; enabling users to discover and access data and information via the Internet; and preserve data. | | The report states that data managers must be able to understand, communicate, and work closely with scientists and others to ensure proper stewardship for the data archive and its distribution. | The report states that the CCSP will provide additional specific community-based guidelines for scientific metadata content where and as appropriate. One approach will be to adopt the ISO 19115 /TC211 Geographic Information/Geomatics standard, which is built on the Federal Geospatial Data Clearinghouse (FGDC) core standards. | The report states that full and open sharing of the full suite of global data sets for all global change researchers is a fundamental objective.<br><br>It recommends improved access to data by expanding the Global Change Master Directory (GCMD) to facilitate access to data.<br><br>In addition, the CCSP will develop and implement guidelines for when and under what conditions data will be made available to users other than those who collected them. | | The report states that procedures and criteria for setting priorities for data acquisition, retention, and purging should be developed by participating agencies, both nationally and internationally. A clearinghouse process should be established to prevent the purging and loss of important data sets.<br><br>It mentions that lessons learned from NASA's pioneering efforts in handling their current holdings (more than 2,500 terabytes) must be used by the community.<br><br>In addition, many important heritage datasets face a growing risk of loss due to deterioration of paper records, obsolescence of electronic media and associated hardware and software, and the gradual loss of experienced personnel. | | | http://www.climatescience.gov/Library/stratplan2003/final/ccspstratplan2003-chap13.htm |
| General Accountability Office (GAO) | Climate Change Research: Agencies Have Data-Sharing Policies but Could Do More to Enhance the Availability of Data from Federally Funded Research, September 2007 | According to the report, agencies have data-sharing policies but could do more to enhance the availability of data from federally funded research. | | | The report evaluates whether additional strategies are warranted to facilitate the permanent archiving of relevant data. The strategies may include: leveraging existing resources; devoting a greater portion of data collection funds to archiving activities; or working with existing entities such as the National Science and Technology Council's Interagency Working Group on Digital Data, to develop additional data archives. | The report recommends NOAA develop mechanisms for agencies to be systematically notified when data have been submitted to archives, so that agency officials have current information about the extent of data availability in order to adjust data-sharing policies over time to best meet the needs of researchers and the communities that use their data. | | | | | http://www.gao.gov/new.items/d071172.pdf |
| National Science and Technology Council (Office of Science and Technology Policy)-Networking and Information Technology Research and Development (NITRD) Program | Harnessing the Power of Digital Data for Science and Society, January 2009 | The report lays out a strategic vision for a digital scientific data universe in which data creation, collection, documentation, analysis, preservation, and dissemination can be appropriately, reliably, and readily managed, thereby enhancing the return on our nation's research and development investment by ensuring that digital data realize their full potential as catalysts for progress in our global information society. | The report provides examples of mechanisms, including the integration of data from various sources and across projects and disciplines. | The report describes the full data life cycle, which includes creation, ingestion or acquisition, documentation, organization, migration, protection, access, and disposition and has two important features. The cycle is dynamic rather than static and includes ongoing processes of creation, disposition, and use. The steps in the cycle are not independent. Appendix B provides a full description of the data life cycle. | The report provides examples of data rights mechanisms including: (1) continued improvement in interoperability across all layers (from software to hardware to networks and resources); (3) comprehensive, global, and transparent search, query, and retrieval capabilities; (4) development, continuing evolution, broad adoption, and regular use of appropriate, community based, cost-effective standards designed to allow efficient information use in innovative ways and in complex combinations; (5) promotion of ready access to appropriate documentation and metadata. | The report includes examples of mechanisms including: reliable protection of security, privacy, confidentiality, and intellectual property rights in complex data environments.<br><br>The report discusses importance of cooperation among industry, academia, NGOs and international agencies (p. 19). | | Examples of data retention mechanisms include encouragement of digital preservation programs explicitly aimed at facilitating sustained access. | | Chris Greer, Director, NCO; Jeannette Wing, Assistant Director, NSF/CISE; Co-chairs http://www.nitrd.gov/subcommittee/agency-contacts.aspx | http://www.nitrd.gov/about/harnessing_power_web.pdf |
| USDA | Current Data Management Plans. Service Center Data Management, 2010 | This web site provides access to Data Management Policy, standards, procedures, guidance, and descriptive documents developed by the Service Center Modernization Initiative. | | Contains over 40 separate data management plans for USDA related projects. | | | | | | | http://www.itc.nrcs.usda.gov/scdm/current_dmp.htm |
| | | | | | | | Current Standards, Policies, and Guidelines including; File Naming Convention Change Control Policy, Data Naming, Change Control Process. | | | | http://www.itc.nrcs.usda.gov/scdm/current_spg.htm |
| | National Agricultural Library (NAL), 2010 | The NAL manages the world's largest agricultural information collections, designated as a USDA heritage asset, which include more than four million physical items as well as extensive digital information products including databases, digital full-text journals, and digital full-text books and maps. | . | | | | | | | | http://www.ocio.usda.gov/records/tools_records.html |
| | | AGRICOLA (AGRICultural OnLine Access), NAL's online catalog and index to the agricultural literature, serves as the finding tool for these collections and is made available free of charge by NAL at http://agricola.nal.usda.gov and by a number of commercial companies. | | | | | | | | | http://www.nal.usda.gov/ |

| Other Federal Agencies | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| NOAA/National Environmental Satellite, Data, and Information (NESDIS) | Records Management. Maps, Imagery, and Publications, 10/15/2009 | Presents a collection of geospatial data derivatives obtained from many sources including maps, aerial photographs, and remote sensors. Included is a list of public-domain software developed by USGS scientists and partners to support a wide variety of natural science research and mapping activities. | | | | | | | | Kevin Gallagher: Associate Director, Geospatial Information and CIO. | http://www.usgs.gov/pubprod/data.html |

# C. References


**National Aeronautics and Space Administration (NASA)**

NASA. 1993. Guidelines for Development of a Project Data Management Plan (PDMP). March 1993. NASA. Office of Space Science and Applications. Information Systems Branch. Available: http://nssdc.gsfc.nasa.gov/nssdc/pdmp_guidelines_march93.rtf.

NASA. 2002. Reference Model for an Open Archival Information System (OAIS). CCSDS Secretariat. Program Integration Division (Code M-3). National Aeronautics and Space Administration. Washington, DC 20546, USA. Blue Book, Issue 1. January 2002. Available: http://public.ccsds.org/publications/archive/650x0b1.pdf.

NASA. 2007. NASA Heliophysics Science Data Management Policy. June 25, 2007. Available: http://hpde.gsfc.nasa.gov/Heliophysics_Data_Policy_2007June25.pdf.

NASA. 2007. White Paper on NASA Science Data Retention. Available: http://nssdc.gsfc.nasa.gov/nssdc/data_retention.html.

NASA, Jet Propulsion Laboratory, and California Institute of Technology. 2004. Cassini/Huygens Program Archive Plan for Science Data. Version 3. June 2004. California Institute of Technology, Pasadena, CA. Available: http://trs-new.jpl.nasa.gov/dspace/bitstream/2014/14261/1/00-0674.pdf.

**National Institutes of Health (NIH)**

NIH, Division of AIDS. 2007. Requirements for Data Management and Statistics for DAIDS Funded and/or Sponsored Clinical Trials. DAIDS Bethesda, MD. DWD-POL-DM-01.00. Effective Date: February 5, 2007 Available: http://www3.niaid.nih.gov/LabsAndResources/resources/DAIDSClinRsrch/PDF/DataMgt_StatPolicy.htm.

NIH, National Cancer Institute (NCI). 2003. National Cancer Institute, Division of Cancer Prevention (DCP), Data Management Requirements. October, 2003. Available: http://prevention.cancer.gov/files/clinical-trials/DataMgmtRqmts.doc.

NIH, National Heart, Lung, and Blood Institute. Undated. Policy for Distribution of Data. Available: http://www.nhlbi.nih.gov/resources/deca/policy_new.htm.

NIH, National Institute on Aging. 2007. Guidelines for Developing a Manual of Operations and Procedures (MOP). Version 1 – December 27, 2007. Available: http://www.nia.nih.gov/NR/rdonlyres/AEC5CE46-96E1-43D9-BA77-BAE8BF0D6CDC/0/ManualofProceduresMOPFinal1.doc.

NIH, Office of Extramural Research. 2003. NIH Data Sharing Policy and Implementation Guidance. March 5, 2003. Available: http://grants.nih.gov/grants/policy/data_sharing/data_sharing_guidance.htm.

NIH, Office of the Chief IT Architect. 2007. NIH Enterprise Conceptual Data Model v1.0. J. Sharp, D. Kotsikopoulos NRFC0025/ STD0012 Category: Standard OCIO January 2007. Available: http://enterprisearchitecture.nih.gov/NR/rdonlyres/5D3017EA-22C1-4BCC-8E0F-79EB7B5C797A/0/NRFC0025.pdf.

Wampler, V. 2008. AD Attribute Data Content and Management: Best Community Practice v1.3. NIH, Division of Computer System Services. February 2008. Available: http://enterprisearchitecture.nih.gov/NR/rdonlyres/8B8AFA60-68A1-4155-A08F-03163B610E39/0/NIHRFC0008ActiveDirectoryAttributeDataContentandManagement.pdf.

**National Oceanic and Atmospheric Administration (NOAA)**

National Academy of Engineering. 2007. Environmental Data Management at NOAA: Archiving, Stewardship, and Access. The National Academy of Sciences. Available: http://www.nae.edu/nae/naepcms.nsf/weblinks/MKEZ-79CSA3?OpenDocument.

National Weather Service. 2010. What Does Data Management Provide to You? NOAA. March 24, 2010. Available: http://www.weather.gov/datamgmt/.

NOAA. 1990. Administrative Management and Executive Secretariat. NOAA Administrative Order: 216-101. Ocean Data Acquisitions. July 9, 1990. Available: http://www.corporateservices.noaa.gov/~ames/NAOs/Chap_216/naos_216_101.html.

NOAA. 2005. Report to Congress on Data and Information Management 2005. October 2005. Available: http://www.ngdc.noaa.gov/noaa_pubs/pdf/NOAA_Congress2005.pdf.

NOAA. 2008. Administrative Management and Executive Secretariat. NOAA Administrative Order: 212-15. Management of Environmental and Geospatial Data and Information. December 2, 2008. Available: http://www.corporateservices.noaa.gov/~ames/NAOs/Chap_212/naos_212_15.html.

NOAA. 2009. Data Management System and Tools. Ocean Exploration and Research. Available: http://www.explore.noaa.gov/data-management.

**National Science Foundation (NSF)**

National Science Board. 2005. Long-Lived Digital Data Collections Enabling Research and Education in the 21st Century. September 2005. NSB-05-40. Available: http://www.nsf.gov/pubs/2005/nsb0540/.

NSF, Division of Earth Sciences. 2002. Implementation of the NSF Data Sharing Policy. April 2002. Available: http://www.nsf.gov/geo/ear/EAR_data_policy_204.pdf.

NSF, Division of Ocean Sciences. 2003. Division of Ocean Sciences: Data and Sample Policy. November 3, 2003. Available: http://www.nsf.gov/pubs/2004/nsf04004/nsf04004_1b.htm.

NSF, Office of Polar Programs. 1998. Guidelines and Award Conditions for Scientific Data. December 3, 1998. Available: http://www.nsf.gov/pubs/1999/opp991/opp991.doc.

NSF, Social, Behavioral and Economic Sciences. 2008. Data Archiving Policy. July 8, 2008. Available: http://www.nsf.gov/sbe/ses/common/archive.jsp.

**Other**

American National Standards Institute. 2009. ANSI/GEIA 859 Data Management. http://webstore.ansi.org/FindStandards.aspx?SearchString = ANSI%2fGEIA+859-2009&SearchOption = 0&PageNum = 0&SearchTermsArray = null%7cANSI%2fGEIA+859-2009%7cnull.

General Accountability Office. 2007. Climate Change Research: Agencies Have Data-Sharing Policies but Could Do More to Enhance the Availability of Data from Federally Funded Research. GAO-07-1172. September 2007. Available: http://www.gao.gov/new.items/d071172.pdf.

National Science and Technology Council (Office of Science and Technology Policy). Networking and Information Technology Research and Development (NITRD) Program. 2009. Harnessing the Power of Digital Data for Science and Society. Report of the Interagency Working Group on Digital Data to the Committee on Science of the National Science and Technology Council. January 2009. Available: http://www.nitrd.gov/about/harnessing_power_web.pdf.

U.S. Climate Change Science Program. 2003. Strategic Plan for Climate Change Science Final Report: Chapter 13. Data Management and Information. July 2003. Available: http://www.climatescience.gov/Library/stratplan2003/final/ccspstratplan2003-chap13.htm.

USDA. 2010. Current Data Management Plans. Service Center Data Management. Available: http://www.itc.nrcs.usda.gov/scdm/current_dmp.htm.

USDA. 2010. Current Standards, Policies and Guidelines. Service Center Data Management. Available: http://www.itc.nrcs.usda.gov/scdm/current_spg.htm.

USDA. 2010. NAL Catalog (AGRICOLA). National Agricultural Library. Available: http://agricola.nal.usda.gov/.

USDA. 2010. Records Management. Office of the Chief Information Officer. Available: http://www.ocio.usda.gov/records/tools_records.html.

**U.S. Department of Energy (U.S. DOE)**

Barish, V.J. and L.A. Gouveia. 2005. Management of OSTI-LLNL Electronic Data. U.S. Department of Energy. Available: https://eed.llnl.gov/ymp/pdf/IM-317550-2.pdf.

Christensen, S.W., L.A. Hook, B. Vet, and B. Sukloff. 2005. The NARSTO Atmospheric Measurements Template. NARSTO Quality Systems Science Center. April 29, 2005 Available: ftp://narsto.esd.ornl.gov/pub/DES_metadata/var_names_web_sources/NARSTO_template_atmospheric_measurements.xls.

Hook, L.A. and S.W. Christensen. 2005. Developing Data Management Policy and Guidance Documents for your NARSTO Program or Project. NARSTO Quality Systems Science Center. Available: http://cdiac.ornl.gov/programs/NARSTO/DM_develop_guide.pdf.

NARSTO Quality Systems Science Center. 2006. Guidelines for Archiving Data in the NARSTO Permanent Data Archive. May 2, 2006. Available: http://cdiac.ornl.gov/programs/NARSTO/Guidelines_for_Archiving_NARSTO_Data.pdf.

Patterson, R.K., G.F. Momberger, L.A. Hook, M.D. Cheng, and T.A. Boden. 1999. NARSTO Quality Systems Management Plan. NARSTO Quality Systems Science Center. September 30, 1999. Available: http://cdiac.ornl.gov/programs/NARSTO/pdf/qsmp_current_version.PDF.

U.S. DOE. 2004. The State of Data Management in the DOE Research and Development Complex. July 14-15, 2004. Available: http://www.osti.gov/publications/2007/datameetingreport.pdf.

U.S. DOE. 2006. ARM Data Sharing and Distribution Policy. Available: http://www.arm.gov/data/policy.stm.

**U.S. Environmental Protection Agency (U.S. EPA)**

Eastern Research Group. 1999. EIIP Phase I Data Model. Emissions Inventory Improvement Program. U.S. Environmental Protection Agency, Washington, DC. January 1999. Available: http://www.epa.gov/ttn/chief/eiip/techreport/volume07/vii01.pdf.

Moore, T., L. Gribovicz, E.V. Hoek, J. Adlhoch, B. Davis-Noland, D. Randall, and M. Mavko. 2009. Comprehensive Data Management of WRAP Emissions Data. 18th Annual International Emission Inventory Conference Comprehensive Inventories -Leveraging Technology and Resources. Baltimore, MD. Available: http://www.epa.gov/ttn/chief/conference/ei18/session1/hoek.pdf.

U.S. EPA. 1989. OSWER Life Cycle Management Guidance, Chapters 1-4, 10. Available: http://www.epa.gov/oswer/oswerlcm.htm.

U.S. EPA. 1989. System Life Cycle Management Guidance Part 3 Practice Paper: Configuration Management. Available: http://www.epa.gov/oswer/docs/oswerlcm/00000019.pdf.

U.S. EPA. 1989. System Life Cycle Management Guidance Part 3 Practice Paper: Data Management During the Life Cycle. January 1989. Available: http://www.epa.gov/oswer/docs/oswerlcm/00000021.pdf.

U.S. EPA. 1989. System Life Cycle Reviews and Approvals. Available: http://www.epa.gov/oswer/docs/oswerlcm/00000021.pdf.

U.S. EPA. 1992. System Life Cycle Management Guidance Part 3 Practice Paper: Data Modeling. May 1992. Available: http://www.epa.gov/oswer/docs/oswerlcm/00000022.pdf.

U.S. EPA. 2000. EPA Quality Manual for Environmental Programs. May 5, 2000. Available: http://www.epa.gov/irmpoli8/policies/2105P010.pdf.

U.S. EPA. 2001. Information Resources Management (IRM) Policy, Chapter 19 Information and Data Management. Available: http://www.epa.gov/irmpoli8/expiredpolicies/Chaptr19.PDF.

U.S. EPA. 2002. Annual Air Quality Data Certifications for PM and Ozone Design Values. June 12, 2002. Available: http://www.epa.gov/ttn/amtic/files/ambient/pm25/datamang/designmem.pdf.

U.S. EPA. 2003. Assessment Factors. Science Policy Council. June 2003. Available: http://www.epa.gov/OSA/spc/pdfs/assess2.pdf.

U.S. EPA. 2003. Guidance for Geospatial Data Quality Assurance Project Plans. March 2003. Available: http://www.epa.gov/quality/qs-docs/g5g-final.pdf.

U.S. EPA. 2006. Guidance on Systematic Planning Using the Data Quality Objectives Process. February 2006. Available: http://www.epa.gov/quality/qs-docs/g4-final.pdf.

U.S. EPA. 2006. Introduction to Lake Michigan Mass Balance Data. Great Lakes Monitoring. March 9, 2006. Available: http://www.epa.gov/greatlakes/lmmb/database.html.

U.S. EPA. 2006. Lake Michigan Mass Balance Metadata. Great Lakes Monitoring. March 9, 2006. Available: http://www.epa.gov/greatlakes/lmmb/metadata.html.

U.S. EPA. 2006. Volunteer Stream Monitoring: A Methods Manual, Chapter 6 Managing and Presenting Monitoring Data. November 30, 2006. Available: http://www.epa.gov/volunteer/stream/vms60.html.

U.S. EPA. 2007. Data Standards Development. June 28, 2007. Available: http://www.epa.gov/irmpoli8/policies/2133p1.pdf.

U.S. EPA. 2007. Data Standards Implementation. June 28, 2007. Available: http://www.epa.gov/irmpoli8/policies/2133p3.pdf.

U.S. EPA. 2007. Data Standards Maintenance. June 28, 2007. Available: http://www.epa.gov/irmpoli8/policies/2133p2.pdf.

U.S. EPA. 2007. Data Standards Policy. June 28, 2007. Available: http://www.epa.gov/oamhpod1/adm_placement/ITS_BISS/datastd.pdf.

U.S. EPA. 2007. Management and Interpretation of Data Under a Triad Approach – Technology Bulletin. May 2007. Available: http://www.brownfieldstsc.org/pdfs/Management_and_Interpretation_of_Data.pdf.

U.S. EPA. 2007. National Geospatial Data Policy Procedure for Geospatial Metadata Management. October 25, 2007. Available: http://www.epa.gov/geospatial/docs/2131.pdf.

U.S. EPA. 2007. Requesting Data Standard Waiver Conformance. June 28, 2007. Available: http://www.epa.gov/irmpoli8/policies/2133p4.pdf.

U.S. EPA. 2007. Scientific Data Management Strategy. November 2007 (Final Draft Version 2.0).

U.S. EPA. 2008. National Geospatial Data Policy. CIO Policy Transmittal 05-002. August 24, 2008. Available: http://www.epa.gov/esd/gqc/pdf/epa_natl_geo_data_policy.pdf.

U.S. EPA. 2009. Agency-wide Quality System Documents. December 30, 2009. Available: http://www.epa.gov/quality/qa_docs.html.

U.S. EPA. 2009. E-mail Records Procedures, September 25, 2009. http://www.epa.gov/irmpoli8/expiredpolicies/cio2135p010.pdf.

U.S. EPA. 2009. EPA Enterprise Architecture Target Data Architecture (DRAFT). Received in Email Correspondence with Kevin Kirby 7/14/09.

U.S. EPA. 2009. EPA Records Schedule: Data Standards and Registry Service. July 31, 2009. Available: http://www.epa.gov/records/policy/schedule/sched/096.htm.

U.S. EPA. 2009. Implementing the National Geospatial Data Policy: Lessons Learned. Available: http://intranet.epa.gov/ospintra/Science Council/Related Docs/ORDNGDPPILOTS.pdf.

U.S. EPA. 2009. IT Policy Mega-Matrix, 2009. Available: http://intranet.epa.gov/otop/itpolicy/IT_Policy_Mega-Matrix_Feb2009_external.pdf.

U.S. EPA. 2009. Metadata Standards for the Enterprise Content Management Program. Received in Email Correspondence with Lynne Petterson 6/10/09.

U.S. EPA. 2009. Procedures for Preparing and Publishing Privacy Act Systems of Records Notices. Available: http://intranet.epa.gov/oei/imitpolicy/qic/ciopolicy/2151-p-03.pdf.

U.S. EPA. 2009. Procedures for Preparing Privacy Act Statements. EPA Document: CIO 2151-P-05. Available: http://intranet.epa.gov/oei/imitpolicy/qic/ciopolicy/2151-p-05.pdf.

U.S. EPA. 2009. Procedures for Preparing Privacy Impact Assessments. EPA Document: CIO 2151-P-04. Available: http://intranet.epa.gov/oei/imitpolicy/qic/ciopolicy/2151-p-04.pdf.

U.S. EPA. 2009. QIC Steering Committee – CIO Policy Consolidated Comments Form. Received in Email Correspondence with Lynne Petterson 6/10/09.

U.S. EPA. 2009. Records Management. December 11, 2009. Available: http://www.epa.gov/records/policy/index.htm. U.S. EPA. 2009. STORET Homepage. November 2, 2009. Website available: http://www.epa.gov/storet/.

U.S. EPA. Undated. NHEERL Data Management Policy and Practices: Genomics and Related High Throughput Data. National Health and Environmental Effects Research Laboratory.

U.S. EPA Region 9. 2008. National Tribal WQX/STORET Data Management. November 18, 2008. Available: http://www.epa.gov/region09/water/tribal/storet-training/pdf/WQXTemplate.pdf.

**U.S. Geological Survey (USGS)**

USGS. 2009. Maps, Imagery, and Information. October 15, 2009. Available: http://www.usgs.gov/pubprod/data.html.

## D. Additional Resources

| Reference | Description |
|---|---|
| **Federal Geographic Data Committee** | |
| Federal Geographic Data Committee. 1992. Policy Statements for Federal Geographic Data Sharing (FGDC Steering Committee Endorsement in 1992.) http://www.fgdc.gov/policyandplanning/data%20sharing. | Policy statements for federal geographic data sharing with the objective of facilitating full and open access to federal geographic data by federal users and the general public. |
| Federal Geographic Data Committee. 1998. FGDC Policy on Access to Public Information and the Protection of Personal Information Privacy in Federal Geospatial Databases. April 1998. http://www.fgdc.gov/policyandplanning/privacypolicy.pdf. | This policy articulates the Federal Geographic Data Committee's (FGDC) endorsement of public access to information and appropriate protections for the privacy and confidentiality of personal information in federal geospatial databases. |
| Federal Geographic Data Committee. 2003. Managing Historical Geospatial Data Records. April 2003. http://www.fgdc.gov/library/factsheets/documents/histdata.pdf. | This fact sheet explains the responsibilities of Federal geospatial data producers for properly creating data, documenting data with appropriate metadata, making data available through a clearinghouse, and arranging for the appropriate disposition of the data. |
| Federal Geographic Data Committee. Undated. FGDC Policy Statement Support for International Infrastructure Activities. http://www.fgdc.gov/policyandplanning/International%20Policy.pdf. | Policy statement in support of sharing experiences and resources locally, nationally, and globally. Includes support in the areas of spatial data standards and metadata standards. |
| **International** | |
| Committee on Earth Observation Satellites. 1995. CEOS Data Principles for Operational Environmental Data. CEOS Yearbook. 1995. http://ceos.cnes.fr:8100/cdrom-00/ceos1/policy/policy3.htm. | Data principles relate to the provision of satellite data in support of operational environmental use for public benefit. These data principles were developed at an April 18-19, 1994 meeting that was hosted by NOAA and NASA. |
| Global Biodiversity Information Facility (GBIF). 2005. Global Biodiversity: The GBIF 3rd-Year Review Report from the Review Committee. February 28, 2005. http://www.gbif.org/fileadmin/Temp_for_New_Web_Site/3YR_full.pdf. | This report includes a section on data policy, which presents recommendations on (1) making scientific biodiversity data freely and openly available over the Internet, and (2) appropriate ways for dealing with intellectual property rights, access, and benefit sharing. |
| Global Biodiversity Information Facility (GBIF). Undated. GBIF Data Use Agreement. http://data.gbif.org/tutorial/datauseagreement. | Data use and data sharing agreements for biodiversity data, developed by the Global Biodiversity Information Facility. |
| Global Earth Observation System of Systems (GEOSS). 2008. – Implementation Guidelines for the GEOSS Data Sharing Principles. September 27, 2008. Draft, not for distribution. | Not available. |

| Reference | Description |
|---|---|
| Organisation for Economic Cooperation and Development. 2007. OECD Principles and Guidelines for Access to Research Data from Public Funding. 2007. http://www.oecd.org/dataoecd/9/61/38500813.pdf. | Provides broad policy recommendations to the governmental science policy and funding bodies of member countries on access to research data from public funding. The recommendations promote data access and sharing among researchers, research institutions, and national research agencies. |
| **NASA** | |
| NASA. 2003. NPR 1441.1D. NASA Records Retention Schedules (NRRSs). NASA Procedural Requirements. Effective Date: February 24, 2003; Expiration Date: February 24, 2013. http://nodis3.gsfc.nasa.gov/displayDir.cfm?t = NPR&c = 1441&s = 1D. | NASA records retention schedules. See Chapter NRRS 2, Legal and Technical Records for data-related retention schedules. |
| NASA. 2009. NPD 2200.1B. Management of NASA Scientific and Technical Information (STI). November 19, 2009. http://nodis3.gsfc.nasa.gov/displayDir.cfm?t = NPD&c = 2200&s = 1B. http://nodis3.gsfc.nasa.gov/npg_img/N_PD_2200_001B_/N_PD_2200_001B__main.pdf. | Policy requiring the Chief Information Officer to conduct a Scientific and Technical Information (STI) program. This program includes collection, management, dissemination, safeguarding, and archiving NASA STI for use by NASA and NASA contractors and grantees, and where appropriate, the public. |
| NASA. 2010. Earth Science Data System-RFC for ECHO Metadata Standards. January 2010. | This memo specifies a metadata standard for the Earth Science Data System. It defines the metadata requirements for compiling metadata into the Earth Observing System Clearinghouse (ECHO). |
| NASA. 2010. ECHO Data Partner User's Guide. January 2010. | This guide outlines tasks that will be performed by data partners, requisite skills, data types, metadata models, compiling data, and data management. |
| **National Archives and Records Administration** | |
| National Archives and Records Administration. 2009. Title 36, Code of Federal Regulations, Subchapter B – Records Management. Effective November 2, 2009. http://www.archives.gov/about/regulations/subchapter/b.html. | This subchapter provides NARA regulations affecting Federal agencies and their records management programs. Includes records disposition and transfer. |
| National Archives and Records Administration. Undated. Title 36, Code of Federal Regulations, Subchapter C – Public Availability and Use (Parts 1250-1258). http://www.archives.gov/about/regulations/. | This subchapter highlights NARA FOIA regulations and regulations on the use of archival records and donated historical materials. |

| Reference | Description |
|---|---|
| **National Institutes of Health** | |
| National Institute of Health. Undated. NIH Manual Chapter 1743 – Keeping and Destroying Records. http://www1.od.nih.gov/oma/manualchapters/management/1743/. | NIH requirements on retaining and destroying records. |
| **National Oceanic and Atmospheric Administration** | |
| NOAA. 2006. NOAA Information Quality Guidelines. November 6, 2006. http://www.cio.noaa.gov/Policy_Programs/IQ_Guidelines_110606.html. | Guidelines for ensuring and maximizing the quality, objectivity, utility, and integrity of disseminated information. |
| NOAA. 2008. NOAA Procedure for Scientific Records Appraisal and Archive Approval: A Guide for Data Managers. September August 15, 2008. http://www.joss.ucar.edu/daarwg/feb09/NOAA_Procedure_document_final_12-16-1.pdf. | Defines the procedure for NOAA to identify, appraise, and decide what scientific records are preserved in a NOAA archive. The procedure applies to accepting or rejecting newly acquired scientific records for a NOAA archive and also to retaining or disposing of existing records already held in a NOAA archive. |
| NOAA. 2008. NOAA Procedure for Scientific Records – Appraisal and Archive Approval: A Guide for Data Users and Producers. September 2008. http://www.joss.ucar.edu/daarwg/feb09/NOAA_Records_Brochure_4_pages_Dec_9.pdf. | Brochure describes a four-step process that NOAA data managers use to determine what scientific records are preserved in a NOAA archive. |
| NOAA. 2010. Coral Reef Information System Web Site. December 2009. http://coris.noaa.gov/data/supportingdocs.html#sensitive. | Web site provides a series of documents to assist contributors in providing data and metadata for the Coral Reef Information System (CoRIS). Includes policy for limiting access to sensitive data and technical guidelines for developing metadata. |
| NOAA. Undated. NOAA Records Disposition Handbook. http://www.corporateservices.noaa.gov/~ames/Records_Management/disposition_handbook.html. | This document lists NOAA records disposition schedules. Separate schedules, which include schedules for scientific data, are provided for NOAA offices. |
| **National Park Service** | |
| National Park Service. 2008. Information Management and Archiving Plan Southeast Coast Inventory and Monitoring Network, Natural Resource Report. NPS/SECN/NRR – 2008/062. September 2008. https://science1.nature.nps.gov/naturebib/biodiversity/2008-10-23/SECN_Data_Management_Plan.pdf. | This Information Management and Archiving Plan is part of the National Park Service's effort to "improve park management through greater reliance on scientific knowledge." It covers issues such as data documentation, data dissemination, data storage and archiving, and records management. |

| Reference | Description |
|---|---|
| **National Science Foundation** | |
| National Science Foundation. 2006. National Science, Technology, Engineering, and Mathematics Education Digital Library (NSDL) Solicitation – Metadata Requirements of the NSDL. November 7, 2006. http://www.nsf.gov/pubs/2008/nsf08554/nsf08554.htm#pgm_desc_txt. | This is a request for proposals that contains a description of metadata requirements for NSDL and numerous related sources. |
| **Other** | |
| Baker, Mary and R. Cummings. 2008. Retaining Information for 100 Years. Storage Networking Industry. http://www.snia.org/images/tutorial_docs/DataProt_Mng/Baker-Cummings-Retaining_Information_100_Years%282%29.pdf. | PowerPoint presentation that discusses the issues of long-term digital storage. Describes problems and best practices. |
| Indiana University. 2010. University Information Policy Office. Data Management Policies and Guidelines. http://informationpolicy.iu.edu/data/policies/. | Contains policies, standards, and guidelines for managing institutional data. |
| Long Term Ecological Research (LTER). 2005. Network Data Access Policy, Data Access Requirements, and General Data Use Agreement. April 6, 2005. http://www.lternet.edu/data/netpolicy.html. | This LTER data policy covers the release of LTER data products, user registration for accessing data, and licensing agreements specifying the conditions for data use. |
| Long Term Ecological Research (LTER). 2005. Review Criteria for LTER Information Management Systems. Version 1.0. April 12, 2005. http://harvardforest.fas.harvard.edu/data/doc/LTER_IM_Review_Criteria_V1.0.pdf. | Criteria for reviewing the success of the LTER Information Management System to ensure it supports site and network science by (1) facilitating access to data and metadata by LTER scientists, the scientific community, and the public, and (2) ensuring the integrity, security, and usability of those data and metadata for future generations. |
| The National Academies Press. 1995. Preserving Scientific Data on Our Physical Universe: A New Strategy for Archiving the Nation's Scientific Information Resources. 1995. http://www.nap.edu/catalog.php?record_id = 4871. | This book advises the National Archives and Records Administration and federal R&D agencies on the long-term retention of scientific and technical data, particularly in electronic formats. It provides criteria for retention assessment and states "all observational data that are non-redundant, reliable and usable by most primary users should be permanently maintained." |
| The National Academies Press. 1995. Study on the Long-term Retention of Selected Scientific and Technical Records of the Federal Government: Working Papers. 1995. http://www.nap.edu/catalog.php?record_id = 9478. | This report discusses the long-term retention of scientific data generated or held by the federal government. It addresses issues on what data should be preserved, who should save these data, and the roles and modes of operation that are appropriate for the National Archives and Records Administration (NARA) in the preservation of data. |

| Reference | Description |
|---|---|
| National Ecological Observatory Network (NEON). 2009. NEON Data Product Concept and Production Plan. March 2009. http://www.neoninc.org/sites/default/files/NEON_Data_Product_Concept_and_Production_Plan.Mar2009_0.pdf. | Data policy covering data use, availability, metadata, and other issues. |
| North American Regional Climate Change Assessment Program. 2007. NARCCAP Operational Data Management Plan. Version 1.5. September 28, 2007. http://www.narccap.ucar.edu/about/data-mgmt-plan.html. | Data management plan for NARCCAP data. The plan addresses collecting data along with archiving and publishing processes. |
| Northern Illinois University. Undated. Responsible Conduct of Research (RCR): Data Ownership Web site. http://ori.dhhs.gov/education/products/n_illinois_u/datamanagement/dotopic.html. | Describes the different participants in the data management process. Includes information on the data ownership challenges between academic institutions and industry, academic institutions and researcher staff, collaboration among research colleagues, and collaboration between authors and journals. Includes training in the form of quizzes, games, and case studies. |
| Pryor, Graham and M. Donnelly. 2009. Skilling Up to Do Data: Whose Role, Whose Responsibility, Whose Career? International Journal of Digital Curation, Vol. 4, No. 2. 2009. http://www.ijdc.net/index.php/ijdc/article/viewFile/126/133. | This paper addresses the roles that are necessary to ensure effective data management while highlighting the specific kinds of expertise that are needed. |
| United States Global Change Research Program, Office of Science and Technology Policy. 1991. Policy Statements on Data Management for Global Change Research. July 2, 1991. http://www.gcrio.org/USGCRP/DataPolicy.html. | Data management policy to facilitate full and open access to quality data for global change research. |
| University of Maryland. 2003. Consolidated USMH and UM Policies and Procedures Manual. February 7, 2003. http://www.president.umd.edu/policies/vi2200a.html. | University of Maryland's policy for the management and protection of the University's institutional data. The policy also highlights responsibilities for the protection of those data. |
| University of Pittsburgh. 2009. University of Pittsburgh Guidelines on Research Data Management. November 25, 2009. http://www.provost.pitt.edu/documents/RDM_Guidelines.pdf. | This memo describes the rights and responsibilities related to scientific data generated by university research, including data produced from federally sponsored research. It covers data retention, data ownership, access to data, and data sharing. |

| Reference | Description |
|---|---|
| **Smithsonian Institution** | |
| Smithsonian Institution. 1998. AXAF Project Data Management Plan for AXAF Science Center. Prepared for NASA. June 22, 1998. http://iwg.cfa.harvard.edu/twiki4/pub/IWGDD/IwgddAgencyPolicyDataPlans/Data_Mgmt_Plan_from_SI.pdf. | Data management plan that describes gathering, processing, dissemination, access, and long-term preservation of AXAF data. |
| Smithsonian Institution. 2007. Digitization: The Increase and Diffusion of Knowledge. Digitization Steering Committee report, Smithsonian Institution. Draft. March 2007. http://www.si.edu/ocio/PDFs/Digiti2007.pdf. | Discussion of bringing digitization to the Smithsonian Institution. Includes recommendations on funding digitization, developing a digitization strategic plan, developing and implementing standards, and improving accessibility to digitized information. |
| **U.S. Department of Commerce** | |
| Department of Commerce. 1968. Public Law 90-396. Standard Data Reference Act. July 11, 1968. http://www.nist.gov/cfo/legislation/Standard%20Reference%20Data%20Act.pdf. | This Act states: "… reliable standardized scientific and technical reference data are of vital importance to the progress of the Nation's science and technology. It is therefore the policy of the Congress to make critically evaluated reference data readily available to scientists, engineers, and the general public." |
| **U.S. Department of Defense** | |
| U.S. DOD. 1987. 5230.24. Distribution Statements on Technical Documents. March 18, 1987. http://www.darpa.mil/prc/DARPA%20Directives%5CDoD_Dir_5230.24.pdf. | This DOD directive updates policies and procedures for marking technical documents, including production, engineering, and logistics information, to denote the extent to which they are available for distribution, release, and dissemination without additional approvals or authorizations. |
| U.S. DOD. 1995. 5230.25. Withholding of Unclassified Technical Data from Public Disclosure. August 18, 1995. http://jitc.fhu.disa.mil/jitc_dri/pdfs/d523025p.pdf. | This directive establishes policy, prescribes procedures, and assigns responsibilities for the dissemination and withholding of technical data. |
| U.S. DOD. 1998. Directive Number 3200.12. DOD Scientific and Technical Information Program (STIP). February 11, 1998. http://www.dtic.mil/whs/directives/corres/pdf/320012p.pdf. | This policy/directive establishes the Science and Technical Information Program (STIP) to provide maximum contribution to the advancement of science and technology. The STIP serves to record, disseminate, and preserve, as a critical asset, the investment in, and results of, DOD research programs. |
| U.S. DOD. 2001. Instruction Number 3200.14. Principles and Operational Parameters of the DOD Scientific and Technical Information Program. June 28, 2001. http://www.dtic.mil/whs/directives/corres/pdf/320014p.pdf. | This DOD instruction lays out the principles and operational parameters to govern the STIP. |

| Reference | Description |
|---|---|
| U.S. DOD. 2007. Information Sharing Strategy. White paper distributed as Memorandum for Secretaries of the Military Departments, Chairman of the Joint Chiefs of Staff, Undersecretaries of Defense, et al. May 4, 2007. http://www.defenselink.mil/cio-nii/docs/InfoSharingStrategy.pdf. | This paper describes the Department of Defense Information Sharing Strategy. Information sharing is defined as "making information available to participants (people, processes, or systems)." |
| U.S. DOD. 2010. Authoritative Source of Data for Use in Modeling and Simulation: Review of Policy and Some Thoughts on Establishing Sources. January 2010. | This paper presents preliminary findings of research efforts to refine policies, identify best practices, and develop plans to establish authoritative sources of data for DOD modeling and simulation. |
| U.S. DOD. 2010. Information Assurance for Modeling and Simulation in a Net-Centric Environment: Review of Policy and Some Thoughts on Implementation Options. January 2010. | This paper reviews current policies relating to information assurance and examines current and emerging information assurance solutions. The paper then explores DOD M&S and key M&S-enabled business processes. The paper concludes by considering ways to provide information assurance to meet M&S requirements. |
| U.S. DOD. 2010. Providing Modeling and Simulation Data and Tools as Services in a Net-Centric Environment: Review of Policy and Some Thoughts on Implementation Options. January 2010. | This paper reviews current guidance and policies relating to services in the DOD net-centric environment and examines current and emerging services. The paper also "explores the dimensions of DOD M&S and key M&S-enabled business processes, then derives requirements for distribution, interoperability, and integration of M&S data and tools to support these processes." |
| **U.S. Department of Health and Human Services** | |
| U.S. Department of Health and Human Services. Office of Research Integrity. Undated. Guidelines for Responsible Data Management in Scientific Research. http://ori.dhhs.gov/education/products/clinicaltools/data.pdf. | This training course is intended to educate new investigators about conducting responsible data management in scientific research. It covers the following topics: data ownership, collection, storage, protection, retention, analysis, sharing, and reporting. |
| **U.S. Environmental Protection Agency** | |
| U.S. EPA. 1990. 2180.3 Facility Identification Data Standard [PDF]. April 9,1990. http://iwg.cfa.harvard.edu/twiki4/pub/IWGDD/IwgddAgencyPolicyDataPlans/EPA_Facility_ID_Data_Standard_2180_3.pdf. | Establishes a data standard for unique facility identification codes to be maintained in all EPA data collections that contain information on facilities regulated by EPA under authority of federal environmental legislation. (Note: policy has expired.) |
| U.S. EPA. 1993. 2180 Locational Data: Policy Implementation Guidance [PDF]. April 30, 1993. http://www.epa.gov/irmpoli8/expiredpolicies/2180.pdf. | Policy to ensure the collection and documentation of accurate, consistently formatted, fully documented, latitude/longitude coordinates as part of all spatially relevant data gathering activities. (Note: policy has expired.) |

| Reference | Description |
|---|---|
| U.S. EPA. 2002. Employee Separation Checkout List. November 20, 2002. | Document provides a check list of items that must be completed before the following actions: Retirement, Resignation/Termination, Move within EPA, Transfer to Another Agency. |
| U.S. EPA. 2006. Clearance Routing Slip. July 24, 2006. | Document provides a check list of items that must be completed before the following actions: Retirement, Resignation/Termination, Move within EPA, Transfer to Another Agency. |
| U.S. EPA. 2007. Records Management Manual. February 2007. http://www.epa.gov/records/policy/manual/index.htm. | This manual prescribes the requirements and responsibilities for conducting EPA's records management program to ensure that the Agency is in compliance with federal laws and regulations, EPA policies, and best practices. |
| U.S. EPA. 2008. 240-R-09-001. Information Access Strategy. January 2008. http://www.epa.gov/nationaldialogue/FinalAccessStrategy.pdf. | Office of Environmental Information's strategy to enhance access to high quality environmental information for all EPA stakeholders. |
| U.S. EPA. 2008. ORD Quality Assurance Review Form (QARF). July 15, 2008. | The QARF document is meant to ensure quality assurance and includes instructions for properly completing the form. |
| U.S. EPA. 2008. Quality Assurance Training for New and Short-Term Employees in NHEERL's Research Program. November 12, 2008. | Training presentation on use of EPA Records Schedules 501, 503, and 507. |
| U.S. EPA. 2008. Records Schedule 507: Criteria and Health Assessment Documents and Risk Assessment Guidelines. August 31, 2008. http://www.epa.gov/records/policy/schedule/sched/507.htm. | Records schedules for ORD documentation related to the development of health, risk, and exposure assessments; risk assessment guidelines; and air and water quality criteria documents used in assessing the risk of exposure to hazardous pollutants. |
| U.S. EPA. 2009. Records Management Policy. June 2009. http://www.epa.gov/records/loop/2009-06.htm#d12-JUN-2009. | This policy notes that officials are responsible for "ensuring records and other types of required documentary materials are not unlawfully removed from EPA by current or departing officials, employees, or agents." |
| U.S. EPA. 2009. Records Schedule 501: Applied and Directed Scientific Research. December 31, 2009. http://www.epa.gov/records/policy/schedule/sched/501.htm. | Records schedules for ORD projects supporting rulemaking, enforcement, regulatory, or policy decisions, and research of significant national interest. |
| U.S. EPA. 2009. Records Schedule 503: Scientific Research Project Files Related to Basic, Exploratory Research. December 31, 2009. http://www.epa.gov/records/policy/schedule/sched/503.htm. | Records schedules for ORD scientific research project supporting the demonstration or proof of concepts such as method validation studies, and basic, exploratory, or conceptual research. |
| U.S. EPA. 2009. Standard Operating Procedure for the Development and Review of Policies, Procedures, Standards, and Guidance. May 15, 2009. | Describes the operating procedure for members of OSIM in "developing and gaining approval for new policy, procedures, standards, or guidance." |

| Reference | Description |
|---|---|
| U.S. EPA. 2010. Research Cores Transfer of Records Memorandum. January 15, 2010. | Memorandum authorizes the transfer of records from an individual leaving the Research Cores, NHEERL, ORD, U.S. EPA, RTP, and NC. |
| U.S. EPA. Undated. EPA Records: Tools Web page. http://www.epa.gov/records/tools/index.htm. | Provides information about records management relevant to all EPA staffers including definitions, quick references, technical briefs, detailed guides, and pertinent forms. |
| U.S. EPA. Undated. Form 1340-8. Senior Agency Officials and Political Appointees Separation or Transfer Records Checklist. http://www.docstoc.com/docs/7868522/Senior-Agency-Officials-Separation-or-Transfer-Records-Checklist-%28PDF%29. | Required form for departing EPA senior officials and political appointees to report on the transfer of their records. |
| U.S. EPA. Undated. Identify and Transfer: What to Do with the Records of Departing Employees. http://www.epa.gov/records/tools/identify.htm. | Describes procedures related to the records of employees who are separating or transferring from EPA. |
| U.S. EPA. Undated. Instructions for Completing EPA Form 3110-1, Employee Separation Checklist. | Document provides instructions to individuals completing the EPA Separation Checklist. |
| **U.S. Geological Services** | |
| USGS. 2008. Information Policies and Notices. October 12, 2008. http://www.usgs.gov/laws/policies_notices.html. | Provides information that describes the principal policies and other notices that govern information posted on USGS Web sites, including the Agency's data quality policy. |

U.S. Environmental Protection Agency
Office of Research and Development (ORD)
Office of Science Information Management (OSIM)

April 30, 2010