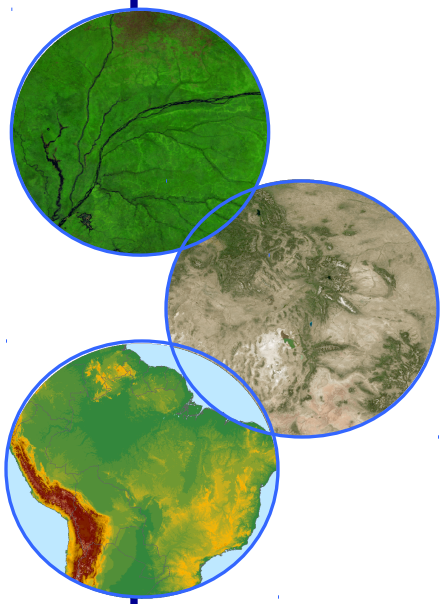
Elements of a Data Management Plan

Bill Michener

University Libraries

University of New Mexico

william.michener@gmail.com



Data Management Practices for Early Career Scientists

NASA Terrestrial Ecology Science Team Meeting

May 2, 2013

Data Management Planning

"Failing to plan is planning to fail"

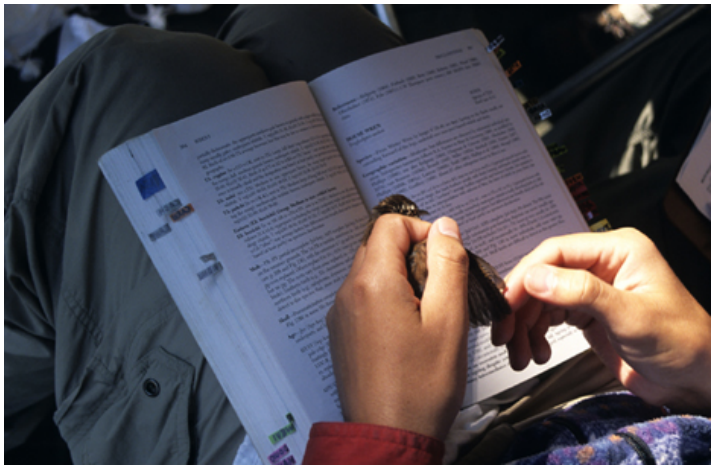
Alan Lakein

"A goal without a plan is just a wish"

Larry Elder

What is a DMP?

A document that describes what you will do with your data **during** and **after** you complete your research



Why Prepare a DMP?

- Saves time
- Increases research efficiency
  - Ensures you and others will be able to understand and use data in future
  - Prevents duplication of effort
- Satisfies funding agency requirements

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

- Visit NASA.gov
- Connect
- Contact Us
- Glossary
- Site Map

NASA Science Live Search GO!

Home | Big Questions | Earth | Heliophysics | Planets | Astrophysics | Missions | Technology | Science News

NASA SCIENCE | EARTH

NAC Science Committee | NASA Science for ... | NASA Celebrates ... | About Us

Home > Earth > Earth Science Data > Data Management Plan Guidance

Earth

- Big Questions
- Focus Areas
- Missions
- Earth Science at NASA
- Earth Science Data
  - Data Management Plan Guidance
  - Open Solicitations in Earth Science Data Systems
  - Data & Information Policy
  - Data Processing Levels
  - Data Maturity Levels
  - Earth Science Disciplines

DATA MANAGEMENT PLAN GUIDANCE

The purpose of a Data Management Plan (DMP) is to address the management of data from Earth science missions, from the time of their data collection/observation, to their entry into permanent archives.

The NASA Procedural Requirements (NPR) 7120.8 document states the requirement for a data management plan. Consistent with this requirement, for each of the Earth science missions, NASA Earth Science Division (ESD) requires the responsible projects to develop a DMP as stated in the Level 1 Requirement below:

> "All terms and conditions of the transfer of data products and associated information to the NASA Earth Science Division (ESD)-assigned data center shall be documented in a Data Management Plan that has been approved by the Earth Science Data and Information System Project."

The purpose of the following document is to provide guidelines for the DMP, to make the assembly of DMPs easier for the projects, and to ensure that DMPs will be in a consistent and useful format. The outline of this document is the recommended outline for all DMPs.

Guidelines for Development of a DATA MANAGEMENT PLAN (DMP), Version 1.0

NASA DMP Guidelines – Vs. 1.0 Jan. 2011

Guidelines for Development of a
DATA MANAGEMENT PLAN (DMP)
Earth Science Division
NASA Science Mission Directorate

NASA Earth Science Division ***Level 1 Requirement****: "All terms and conditions of the transfer of data products and associated information to the NASA Earth Science Division (ESD)-assigned data center shall be documented in a Data Management Plan that has been approved by the Earth Science Data and Information System Project."*
… "It is expected that each project develops the DMP in collaboration with the ESD-assigned data centers, but the primary responsibility rests with the project."

National Aeronautics and Space Administration

NASA Data Management Plan Guidelines – 2011

1. **Introduction**
   2. DMP Development, Maintenance, and Management Responsibility
   3. Change Control
   4. Relevant Documents
1. **Project Overview**
   1. Project Objectives
   2. Science Objectives
   3. Mission or Investigation Summary
   4. Instrument Overview

Source: "Guidelines for Development of a Data Management Plan, Earth Science Division, NASA Science Mission Directorate", January 2011.

3. **Project Data Flow**
   1. Mission Operations
   2. Science Operations
      1. Science Control
      2. Science Planning and Scheduling
      3. Science Data Set Generation
      4. Project Data Storage and Distribution
   3. Post-Mission/Investigation Stewardship and Access
      1. Transition to Science Data Center(s)
      2. Directories and Catalogs
      3. Standards and Policies
      4. Networking Requirements
3. **Products**
   1. Science Data Product Summary
      1. Instrument "A"
   2. Associated Archive Products
3. **Special Considerations**
4. **Acronyms**

daac.ornl.gov/PI/pi_info.shtml

Gmail – Priority ... | Most Visited | Customize Links | Gmail – Inbox (2... | Free Hotmail | Windows Market... | Windows Media | Winc

NASA Earth Data | Data Discovery | Data Centers | Community | Science Disciplines | Search EOSDIS

DAAC Home -> HELP -> FOR DATA PROVIDERS

ORNL DAAC Distributed Active Archive Center for Biogeochemical Dynamics

About Us | About Data | Get Data | Data Tools | Help

home sign-in

search for in Metadata go ?

Data Provider Information

Introduction

At the request of field researchers, investigators, GIS and image specialists, and data managers, we prepared the following data management practices that data collectors and providers should follow to improve the usability of their data sets. We assembled what we feel are the most important practices that researchers should implement to make their data sets ready to share with other researchers and to submit to the ORNL DAAC for archiving and distribution.

This guidance is provided for those who perform environmental measurements, compile data from various sources, prepare GIS coverages, and compile remote sensing images for environmental applications, although many of the practices may be useful for other data collection and archiving activities. These practices could be performed at any time during the preparation of the data set, but we suggest that researchers plan for them before measurements are taken and implement them during measurements.

During Data Preparation

The ORNL DAAC offers the following Best Practices that investigators should perform to improve the usability of their data.

1. Define the Contents of Your Data Files
2. Use Consistent Data Organization
3. Use Consistent File Structure and Stable Formats For Tabular and Image Data
4. Assign Descriptive File Names
5. Perform Basic Quality Assurance
6. Assign Descriptive Data Set Titles
7. Provide Documentation

A version of these Best Practices were published in Cook et al. (2001).

Metadata

In order to archive and distribute various data sets, we need metadata, which is information about the data we distribute. Metadata is used both to describe the data so that others can understand what it represents and to find data of interest. Metadata is used both to describe the data so that others can understand what it represents and to find data of interest. Metadata can be in the form of a document or a specially formatted list of the parameters, keywords, spatial and temporal extent, investigators, and other information about the data set. Please contact the DAAC by e-mail for assistance in preparing

NSF DMP Requirements

From Grant Proposal Guidelines:

Plans for data management and sharing of the products of research. **Proposals must include a supplementary document of no more than two pages labeled "Data Management Plan".** This supplement should describe how the proposal will conform to NSF policy on the dissemination and sharing of research results (in AAG), and may include:

1.the **types of data**, samples, physical collections, software, curriculum materials, and other materials to be produced in the course of the project

2. the **standards to be used for data and metadata** format and content (where existing standards are absent or deemed inadequate, this should be documented along with any proposed solutions or remedies)

3. **policies for access and sharing** including provisions for appropriate protection of privacy, confidentiality, security, intellectual property, or other rights or requirements

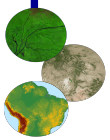
4. policies and **provisions for re-use**, re-distribution, and the production of derivatives

5. **plans for archiving** data, samples, and other research products, and for preservation of access to them

Components of a DMP

1. Information about data & data format
2. Metadata content and format
3. Policies for access, sharing and re-use
4. Long-term storage and data management
5. Budget

1. Information About Data & Data Format

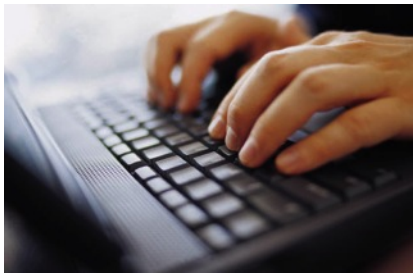


1.1 Description of data to be produced
- Experimental, Observational, Raw or derived,
- Physical collections, Models, Images, etc.

1.2 How data will be acquired
- When? Where? Methods?

1.3 How data will be processed
- Software used, Algorithms, Workflows

1.4 File formats
- csv, tab-delimited, naming conventions



1.5 Quality assurance & quality control

1.6 Existing data
- If existing data are used, what are its origins?
- Will your data be combined with existing data?
- What is the relationship between your data and existing data?

1.7 How data will be managed in short-term
- Version control, Backing up, Security & protection,
- Who will be responsible?

2. Metadata Content & Format

Metadata is the documentation describing all aspects of the data (e.g., who, why, what, when, where)

2.1 What metadata are needed

- Any details that make data understandable and usable

2.2 How metadata will be created and/or captured

- Lab notebooks? GPS units?
- Auto-saved on instrument? Manually entered?

2.3 What format will be used for the metadata

- Standards for community (EML, ISO 19115, etc.)
- Justification for format chosen

3. Policies for Access, Sharing, Reuse

3.1 Obligations for sharing
- Funding agency, institution

3.2 Details of data sharing
- How long?
- When?
- How access can be gained?

3.3 Ethical/privacy issues with data sharing

3.4 Intellectual property & copyright issues
- Institutional policies
- Funding agency policies
- Embargos for political/commercial reasons

3.5 Intended future uses/users for data

3.6 Citation
- How should data be cited when used?
- Persistent citation?

4. Long-term Storage & Data Management

4.1 What data will be preserved

4.2 Where will it be preserved

- Most appropriate archive for data
- Community standards

4.3 Data transformations/formats needed

- Consider archive policies

4.4 Who will be responsible

- Contact person for archive

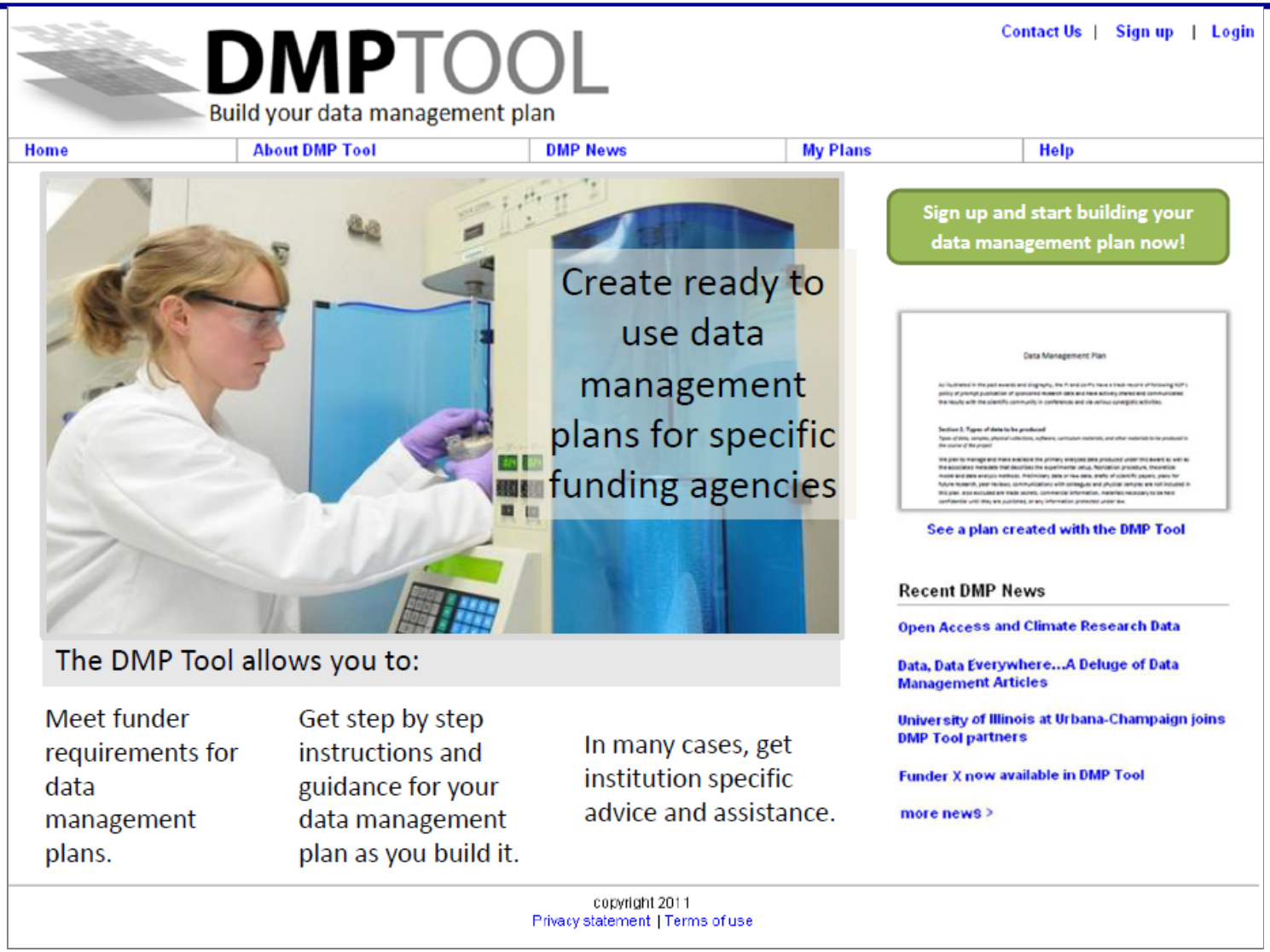
5. Budget

5.1 Anticipated costs

- Time for data preparation & documentation
- Hardware/software for data preparation & documentation
- Personnel
- Archive costs

5.2 How costs will be paid

Data Management Planning Tool



DMPTool
Guidance and Resources for your Data Management Plan

NSF-GEN: Generic: Plan description

Progress

Click on a section below to edit it at any time.

= complete

Plan description

1. Types of data produced
2. Data and metadata standards
3. Policies for access and sharing
4. Policies for re-use, redistribution
5. Plans for archiving & preservation

The NSF-GEN: Generic plan will cover the subject areas listed to the left.

You can save a plan in progress and return later to finish or edit.

Plan Name: (required)
Please give your plan a name to help you identify it in the future

Solicitation Number:

Comment:
Provide any notes you want to appear on your My Plans page. This will not appear in the document

Cancel | Save and next

Resources

General
NSF Data Sharing Policy

NSF Data Management Plan Requirements

DMPTOOL is a service of the University of California Curation Center of the California Digital Library
Copyright © 2010-2012 The Regents of the University of California
Privacy Policy | Terms of Use | Photo Credits

DMP Tool
Build your data management plan

Progress

Sections marked with a check are complete. You can navigate to a section and edit at any time.

NSF-GEN: Generic

Cover page

1. Types of data produced
2. Data and metadata standards
3. Policies for access and sharing
4. Policies for re-use, redistribution
5. **Plans for archiving & preservation**

Plans for archiving data, samples, and other research products, and for preservation of access to them.

Suggested answer text; copy and paste as needed:

As advised by University of Virginia Library staff members, I plan on depositing my research data in the UVA institutional repository – Libra. I will submit the necessary metadata and other resources to make my data accessible for future users. In accordance with the University of Virginia policy RES-002, "Policy: Laboratory Notebook and Recordkeeping," the data will be preserved for a minimum of five years upon completion of the project. However the current preservation plan for Libra will be to preserve the data indefinitely. The Libra backup plan provides for data redundancy including off-site storage.

As advised by University of Virginia Library staff members, I plan on depositing my research data in the UVA institutional repository - Libra. I will submit the necessary metadata and other resources to make my data accessible for future users. In accordance with the University of Virginia policy RES-002, "Policy: Laboratory Notebook and Recordkeeping," the data will be preserved for a minimum of five years upon completion of the project. However the current preservation plan for Libra will be to preserve the data indefinitely. The Libra backup plan provides for data redundancy including off-site storage.

Resources

University of Virginia
UVa Scientific Data Consulting Group

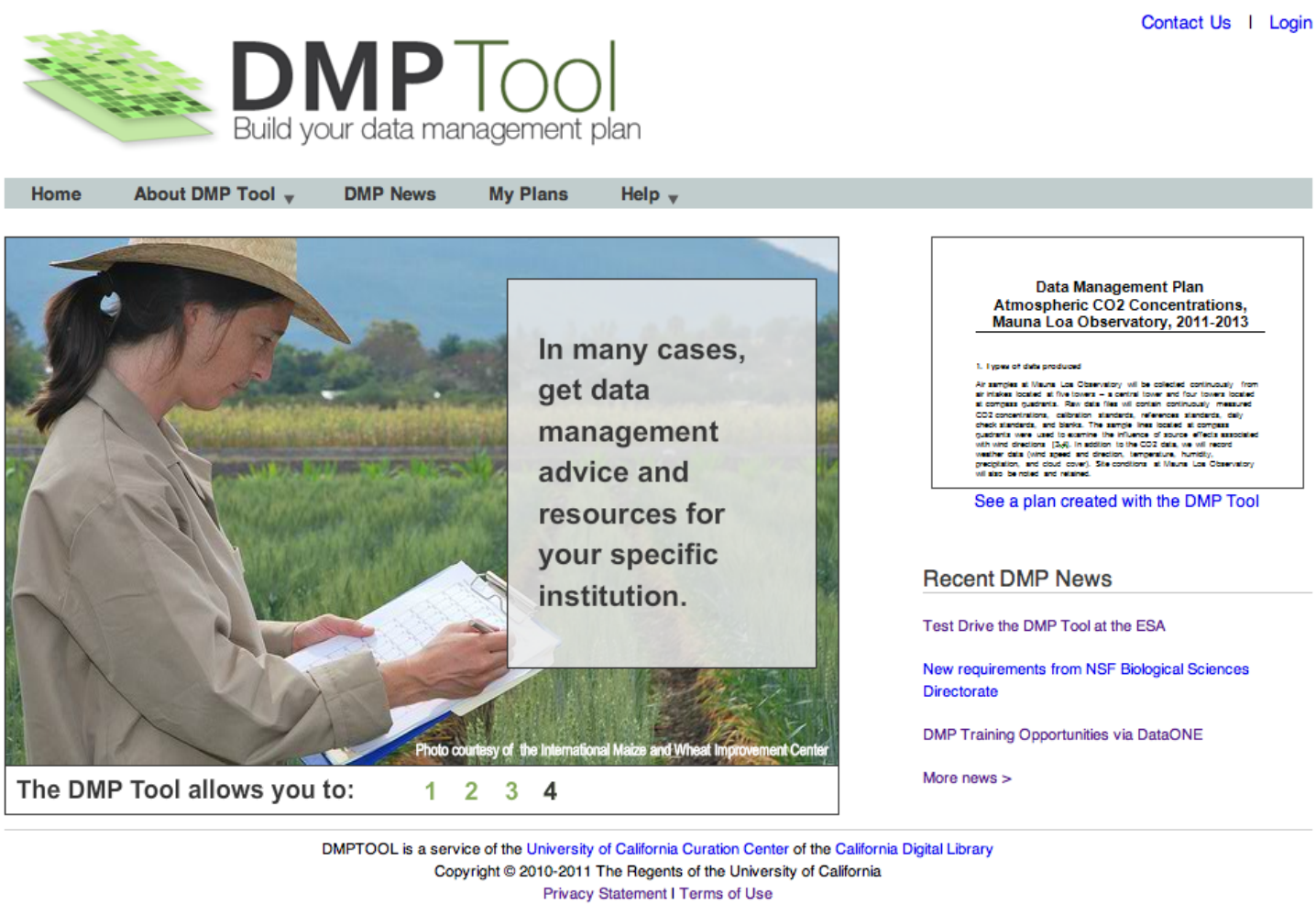
Archiving & Sharing Data Guidance

UVa Policy RES-002: "Laboratory Notebook and Recordkeeping"

General
NSF Data Sharing Policy

NSF Data Management Plan Requirements

http://dmp.cdlib.org/



When do you plan?

- At the beginning … and throughout the research life cycle

From: www.jisc.ac.uk

How do you plan?

- "The better is the enemy of the 'good enough'"
  - Don't hesitate because you don't know all the answers
    - Use metadata from your proposal, publications, etc.
- Review and revise routinely

Who plans?

DMP Resources

- http://dmp.cdlib.org/
- http://dataone.org
- NASA Resources: also see following slides
  - http://science.nasa.gov/earth-science/earth-science-data/data-management-plan-guidance/
  - http://earthdata.nasa.gov/about-eosdis/interfaces
  - http://science.nasa.gov/earth-science/earth-science-data/
  - http://nasascience.nasa.gov/earth-science/earth-science-data/data-information-policy/
  - http://disc.sci.gsfc.nasa.gov/additional/scienceTeam/documents/EvolutionRequirements.doc

Agency Requirements

Agency	Requirement	Details
NSF	Requires a 2 page (max.) data plan with all proposals	Plans for data management and sharing of the products of research
NIH	Requires a paragraph following the Research Plan section of the proposal that describes how data will be shared	Data Sharing Plan
NASA	Requires a data sharing plan as part of the Scientific/Technical/Management section of proposals	Guidebook for Proposers
NIJ	Requires a 1–2 page data archiving strategy	Guidelines and examples
NEH	Digital Humanities Implementation Grants require a 2 page (max.) sustainability and data management plan	Guidelines

NASA Guidebook for Proposers

NATIONAL AERONAUTICS

AND

SPACE ADMINISTRATION

****

GUIDEBOOK FOR PROPOSERS

RESPONDING TO A

NASA RESEARCH ANNOUNCEMENT (NRA) OR COOPERATIVE

AGREEMENT NOTICE (CAN)

Edition: January 2010

NATIONAL AERONAUTICS
AND SPACE ADMINISTRATION

+ NASA Homepage
+ NASA en Español
+ Contact NASA

Home | Data Services | Tools | Documents | Related Sites | About PDS | Sitemap

Documents

- Archive Preparation Guide
- Data Dictionary
- Proposer's Archive Guide
- Standards Reference

Archive Preparation Guide With Examples and Templates

Archive Preparation Guide (June 4, 2008)

1. PDS Home Page (http://pds.nasa.gov/)

2. PDS Standards Reference (JPL Document D-7669, Part 2) - Specific PDS data preparation standards for archive quality data sets. (stdref)

3. PDS Data Dictionary - List of PDS-recognized keywords (and acceptable values where restrictions apply) (psdd.pdf)

4. PDS Nodes and Contact Information - a table of PDS personnel and related contacts. These individuals are your most useful resources - contact them early and often. (Node Contacts)

5. Example Archive Documents:

- MOU – Example Memorandum of Understanding (MRO)
- DMAP – Example Data Management and Archiving Plan (MGS/ODY/MER/MRO/Phoenix, Messenger, Deep Impact, New Horizons, Stardust)
- ICD – Example document describing the detailed interface agreement between the data provider and discipline node
  - CRISM Science Team / Geosciences Node
- SIS – Example document describing volume and data product structure, content, and processing history
  - Draft Huygens - SSP
  - Data Product SIS template
  - Archive Volume SIS template
- MACP – Example Mission Archive Close-out Plan (MGS)

6. Example Catalog Files

- MISSION.CAT (CASSINI-HUYGENS)
- INSTHOST.CAT (CASSINI ORBITER)
- INST.CAT (CASSINI ORBITER)
- DATASET.CAT (CASSINI SATURN CIRS)
- PERSON.CAT
- REF.CAT

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

- Visit NASA.gov
- Connect
- Contact Us
- Glossary
- Site Map

NASA Science Live Search GO!

Home | Big Questions | Earth | Heliophysics | Planets | Astrophysics | Missions | Technology | Science News

NASA SCIENCE | EARTH

NAC Science Committee | NASA Science for ... | NASA Celebrates ... | About Us

Home > Earth > Earth Science Data > Data Management Plan Guidance

Earth

- Big Questions
- Focus Areas
- Missions
- Earth Science at NASA
- Earth Science Data
  - Data Management Plan Guidance
  - Open Solicitations in Earth Science Data Systems
  - Data & Information Policy
  - Data Processing Levels
  - Data Maturity Levels
  - Earth Science Disciplines

DATA MANAGEMENT PLAN GUIDANCE

The purpose of a Data Management Plan (DMP) is to address the management of data from Earth science missions, from the time of their data collection/observation, to their entry into permanent archives.

The NASA Procedural Requirements (NPR) 7120.8 document states the requirement for a data management plan. Consistent with this requirement, for each of the Earth science missions, NASA Earth Science Division (ESD) requires the responsible projects to develop a DMP as stated in the Level 1 Requirement below:

> "All terms and conditions of the transfer of data products and associated information to the NASA Earth Science Division (ESD)-assigned data center shall be documented in a Data Management Plan that has been approved by the Earth Science Data and Information System Project."

The purpose of the following document is to provide guidelines for the DMP, to make the assembly of DMPs easier for the projects, and to ensure that DMPs will be in a consistent and useful format. The outline of this document is the recommended outline for all DMPs.

Guidelines for Development of a DATA MANAGEMENT PLAN (DMP), Version 1.0

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

- Visit NASA.gov
- Connect
- Contact Us
- Glossary
- Site Map

NASA Science Live Search GO!

Home | Big Questions | Earth | Heliophysics | Planets | Astrophysics | Missions | Technology | Science News

NASA SCIENCE | EARTH

NAC Science Committee | NASA Science for ... | NASA Celebrates ... | About Us

Home > Earth > Earth Science Data > Data & Information Policy

Earth

- Big Questions
- Focus Areas
- Missions
- Earth Science at NASA
- Earth Science Data
  - Data Management Plan Guidance
  - Open Solicitations in Earth Science Data Systems
  - Data & Information Policy
    - Data Rights & Related Issues
  - Data Processing Levels
  - Data Maturity Levels
  - Earth Science Disciplines

Data & Information Policy

NASA's Earth Science program was established to use the advanced technology of NASA to understand and protect our home planet by using our view from space to study the Earth system and improve prediction of Earth system change. To meet this challenge, NASA promotes the full and open sharing of all data with the research and applications communities, private industry, academia, and the general public. The greater the availability of the data, the more quickly and effectively the user communities can utilize the information to address basic Earth science questions and provide the basis for developing innovative practical applications to benefit the general public.

A common set of carefully crafted data exchange and access principles was created by the Japanese, European and U.S. International Earth Observing System (IEOS) partners during the 1990s and the early years of the 21st century. From these principles, NASA has adopted the following data policy (in this context the term 'data' includes observation data, metadata, products, information, algorithms, including scientific source code, documentation, models, images, and research results):

- NASA will plan and follow data acquisition policies that ensure the collection of long-term data sets needed to satisfy the research requirements of NASA's Earth science program.
- NASA commits to the full and open sharing of Earth science data obtained from NASA Earth observing satellites, sub-orbital platforms and field campaigns with all users as soon as such data become available.
- There will be no period of exclusive access to NASA Earth science data. Following a post-launch checkout period, all data will be made available to the user community. Any variation in access will result solely from user capability, equipment, and connectivity.
- NASA will make available all NASA-generated standard products along with the source code for

Thanks!

- Amber Budden, Bob Cook, Viv Hutchison, Carly Strasser, and DMP Tool contributors!