Crafting the Data Management Plan

Lee Dotson

University of Central Florida, lee.dotson@ucf.edu

Find similar works at: https://stars.library.ucf.edu/ucfscholar

This Other Presentation is brought to you for free and open access by STARS. It has been accepted for inclusion in Faculty Scholarship and Creative Works by an authorized administrator of STARS. For more information, please contact STARS@ucf.edu.

STARS Citation

Dotson, Lee, "Crafting the Data Management Plan" (2016). Faculty Scholarship and Creative Works. 42. https://stars.library.ucf.edu/ucfscholar/42
Crafting the Data Management Plan

NSF CAREER WORKSHOP SERIES 2016
April 21, 2016

Lee Dotson
Digital Initiatives Librarian
John C. Hitt Library
Today’s session:

• What is a DMP?
• Guiding questions to answer for your DMP
• Data preservation and strategies for sharing your data
• Additional Resources
What is a Data Management Plan?

• Brief description of how you will comply with funder’s data sharing policy

• Typically reviewed as part of the grant application

• Requirements vary by funding agency and NSF directorate

• A document that describes what you will do with your data during your research and after you complete your project
Why data management?

• Conform to the NSF guidelines on the dissemination and sharing of research results
• OSTP Memo to Increase Access to Federally Funded Research
• Data sharing allows for reproducibility, transparency, and data re-use in research.
• Sharing is easier if data are managed well from the start of a project.
What does the NSF say?

• “Proposals must include a supplementary document of no more than two pages…”

• “…should describe how the proposal will conform to NSF policy on the dissemination and sharing of research results”

• “Fastlane will not permit submission of a proposal that is missing a Data Management Plan.”

• “…may include only the statement that no detailed plan is needed, as long as the statement is accompanied by a clear justification”

• “…will be reviewed as an integral part of the proposal”

A DMP is a Living Document

• Keep your plan current
• Incorporate changes
• Use as a guide for daily activities
Where to Start?

Small & Simple

- Document what you know **now**
- Share the plan with your team
- Avoid procrastination and immobilization
What to include

- **Types of data**, samples, and other materials to be produced in the course of the project. Retention period?
- **Data and metadata standards** for format and content
- **Access and sharing policies** including provisions for appropriate protection of privacy or other rights
- **Re-use policies** and provisions re-distribution and the production of derivatives
- **Archiving plans** for data, samples, and other research products, and for preservation of access to them.
DMP templates and guidance

• DMPTool [https://dmptool.org/](https://dmptool.org/)
  – Create ready-to-use data management plans for specific funding agencies with step-by-step instructions and guidance for data management plan
Features of the DMPTool
Data Management Planning Tool
Create, review, and share data management plans that meet institutional and funder requirements.

Get Started
Log in through your institution

Select your institution below and you will be directed to your institutional log in page.

Select your institution

Next >>

If you do not see your institution in the list, please select "Not in List" and click Next.
## University of Central Florida

### OVERVIEW

<table>
<thead>
<tr>
<th>My DMPs</th>
<th>My DMPs under Review (where applicable)</th>
</tr>
</thead>
<tbody>
<tr>
<td>13 plans I own</td>
<td>No DMPs are under review.</td>
</tr>
<tr>
<td>0 plans I co-own</td>
<td></td>
</tr>
</tbody>
</table>

Create New DMP
CREATE NEW DMP

Start with a DMP Template
To create a new DMP, select a funder or institutional template.

Copy an Existing DMP

The existing DMPs in this list are either publicly shared by any user, shared within your institution by other DMP creators, or are plans that you have previously created.

Select an existing DMP below and add text to the template.

<table>
<thead>
<tr>
<th>Name</th>
<th>Owner</th>
</tr>
</thead>
<tbody>
<tr>
<td>A unified approach to preserving cultural software objects and their development histories</td>
<td>DMP dmpcurator</td>
</tr>
<tr>
<td>Atmospheric CO2 Concentrations, Mauna Loa Observatory, Hawaii, 2011-2013</td>
<td>DMP dmpcurator</td>
</tr>
<tr>
<td>Daymet Follow-On: Surface Weather Data with Uncertainty Quantification for Terrestrial Ecosystem Process Models</td>
<td>DMP dmpcurator</td>
</tr>
<tr>
<td>Multimedia Text Annotation for Students</td>
<td>DMP dmpcurator</td>
</tr>
<tr>
<td>CAREER: Parietal Cortex and the Transformation of Spatial Cognition into Action</td>
<td>DMP dmpcurator</td>
</tr>
</tbody>
</table>

1 2 3 4 5 ... Next » Last »
### PUBLIC DMPS

Public DMPS are plans created using the DMPTool and shared publicly by their owners. They are not vetted for quality, completeness, or adherence to funder guidelines.

<table>
<thead>
<tr>
<th>Plan Title</th>
<th>Funder Template</th>
<th>Institution</th>
<th>Owner</th>
<th>Download</th>
</tr>
</thead>
<tbody>
<tr>
<td>A unified approach to preserving cultural software objects and their development histories</td>
<td>NEH-ODH: Office of Digital Humanities</td>
<td>University of California, Office of the President</td>
<td>DMP dmcurator</td>
<td><img src="pdf" alt="pdf" /></td>
</tr>
<tr>
<td>Atmospheric CO2 Concentrations, Mauna Loa Observatory, Hawaii, 2011-2013</td>
<td>NSF-AGS: Atmospheric and Geospace Sciences</td>
<td>University of California, Office of the President</td>
<td>DMP dmcurator</td>
<td><img src="pdf" alt="pdf" /></td>
</tr>
<tr>
<td>Daymet Follow-On: Surface Weather Data with Uncertainty Quantification for Terrestrial Ecosystem Process Models</td>
<td>NSF-AGS: Atmospheric and Geospace Sciences</td>
<td>University of California, Office of the President</td>
<td>DMP dmcurator</td>
<td><img src="pdf" alt="pdf" /></td>
</tr>
<tr>
<td>Multimedia Text Annotation for Students</td>
<td>NEH-ODH: Office of Digital Humanities</td>
<td>University of California, Office of the President</td>
<td>DMP dmcurator</td>
<td><img src="pdf" alt="pdf" /></td>
</tr>
<tr>
<td>CAREER: Parietal Cortex and the Transformation of Spatial Cognition into Action</td>
<td>NSF-BIO: Biological Sciences (2013-)</td>
<td>University of California, Office of the President</td>
<td>DMP dmcurator</td>
<td><img src="pdf" alt="pdf" /></td>
</tr>
<tr>
<td>Biosignature Suites: Using Connections between Microbes and Minerals to understand Biogenic Carbonates</td>
<td>NSF-BIO: Biological Sciences (2013-)</td>
<td>University of California, Office of the President</td>
<td>DMP dmcurator</td>
<td><img src="pdf" alt="pdf" /></td>
</tr>
<tr>
<td>Peer Power</td>
<td>NSF-CISE: Computer and Information Science and Engineering</td>
<td>University of California, Office of the President</td>
<td>DMP dmcurator</td>
<td><img src="pdf" alt="pdf" /></td>
</tr>
</tbody>
</table>
Select one of the funder DMP Templates listed to proceed to the next step. The type of template chosen can affect what information you will need to provide in the following steps.

- Alfred P. Sloan Foundation
- BCC-DMO NSF OCE: Biological and Chemical Oceanography
- Department of Energy: Office of Science
- DMP Template from DCC
- GoMRI Research Consortia DMP Template 2015
- Gordon and Betty Moore Foundation
- Institute of Education Sciences (US Dept of Education)
- Institute of Museum and Library Services
- Joint Fire Science Program
- National Institutes of Health
- National Science Foundation
  - NSF-AGS: Atmospheric and Geospace Sciences
  - NSF-AST: Astronomical Sciences
  - NSF-BIO: Biological Sciences (2013-)
  - NSF-CHE: Chemistry Division
  - NSF-CISE: Computer and Information Science and Engineering
  - NSF-DMR: Materials Research
  - NSF-EAR: Earth Sciences
  - NSF-EHR: Education and Human Resources
  - NSF-ENG: Engineering
  - NSF-GEN: Generic
  - NSF-PHY: Physics
  - NSF-SBE: Social, Behavioral, Economic Sciences
- NEH-OCH: Office of Digital Humanities
- U.S. Geological Survey DMP Guidance
DMP Overview

- **DMP Template**: DMP Template from DCC
- **DMP Title**: FCTL Summer 2015
- **Proposal Solicitation Number**: LIB 1234
- **Proposal Submission Deadline**: 05/14/2015
- **Visibility**:
  - public
  - institutional
  - private
- **Add Co-owners**: Add co-owners; search by name or email address

[Save and Next >>]
Plan was successfully created.

DMP DETAILS

DMP Template from DCC

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

☑ = Complete

* = Mandatory

Template Outline

- Data Types, Formats, Standards and Capture Methods
- Ethics and Intellectual Property
- Access, Data Sharing and Reuse
- Short-Term Storage and Data Management
- Deposit and Long-Term Preservation
- Resourcing

University of Central Florida Resources

UCF Libraries
UCF Office of Research and Commercialization

Instructions
Links

Save Response
Save and Next
Example funder prompts: What data outputs will your research generate? Outline volume, type, content, quality and format of the final dataset. Outline the metadata, documentation or other supporting material that should accompany the data for it to be interpreted correctly. What standards and methodologies will be utilised for data collection and management? State the relationship to other data available in public repositories e.g. existing data sources that will be used by the research project. Gaps between available data and that required for the research; the added value that new data would provide in relation to existing data.

**Guidance**

Outline and justify your choices: You should detail what data you will create and explain why you have opted for particular formats, standards and methodologies. Bear in mind that the choices you make may make it easier or harder to share and preserve your data.

Data can mean many different things, but there are typically four main categories that it can be sorted into for management purposes. The category that you choose will then have an effect upon the choices that you make throughout the rest of your data management plan.

**Observational**

- Captured in real-time
- Usually irreplaceable
- Examples: Sensor readings, telemetry, survey results, images
Open Science Framework

A scholarly commons to connect the entire research cycle

Free and open source. Start now.

Full Name
Contact Email
Confirm Email
Password (Must be 6 to 256 characters)

By clicking "Sign up free", you agree to our Terms and that you have read our Privacy Policy, including our information on Cookie Use.

Sign up free
Example funder prompt: Demonstrate that you have sought advice on and addressed all copyright and rights management issues that apply to the resource. Make explicit mention of consent, confidentiality, anonymisation and other ethical considerations, where appropriate. Are any restrictions on data sharing required – for example to safeguard research participants or to gain appropriate intellectual property protection?

Guidance

Present a strong case for any restrictions on sharing: Explain any constraints, such as embargo periods or restricted access, and ensure these are properly justified as there is a common expectation that publicly funded research data will be openly available as soon as possible. These justifications may also be of use in the event of a Freedom of Information request for

Copyright and Intellectual Property Rights

Sharing data that you produced/collection yourself:

- Data is not copyrightable (yet a particular expression of data can be, such as a chart or table in a book)
- Data can be licensed; some data providers apply licenses that limit how the data can be used, such as to protect the privacy of participants in a study or guide
Example funder prompts: What are the further intended and/or foreseeable research uses for the completed dataset(s)? How do you make the resource accessible to the potential audience(s) identified? Where will you store the data? How will other researchers be able to access the data? Will a data sharing agreement be required? What is the timescale for public release of the data? State any expected difficulties in data sharing, along with causes and possible measures to overcome these difficulties. How will data sharing provide opportunities for coordination or collaboration?

Anticipate and plan for data reuse. It can help to envisage which users your data would be of value to, and address their needs when deciding how to make the data available. Data centres may also ask you to meet minimum quality standards to make sure your data can be understood and reused by other researchers.

You can share your data easily by emailing it to requestors, or posting it to a website, Google, Amazon or Microsoft. However, this method of sharing makes it difficult for people to find your data. Depositing your data in an archive will facilitate its discovery and preservation.

- Post online via a project or institutional web site
- Submit data to a journal article
- Deposit in a discipline specific data repository or archive. Search for an appropriate repository:
  - Databib - an online, community-driven annotated bibliography and
    repository.

Save Response  Save and Next
WHERE DO YOU STORE YOUR RESEARCH DATA?

- USB DRIVE
- DROPBOX
- RESEARCH DATA REPOSITORY

re3data.org
Welcome to STARS!

UCF's Showcase of Text, Archives, Research & Scholarship exists to publicize, disseminate, and provide ready access to works by, for, and about the University of Central Florida. Administered by the UCF Libraries, STARS is available to host and promote research, creative activity, and institutional outputs to...

- Ensure persistent access to your work
- Increase discovery of UCF scholarship and creative endeavors
- Foster scholarly collaborations with colleagues
- Document and record UCF's history and progress
- Discover open access materials and projects created by UCF authors
- Allow you to share your work while retaining your copyright. If you own the copyright to your work, the copyright for materials uploaded to STARS remains with you.

The repository is currently accepting submissions by all faculty, staff, students and affiliates of UCF wishing to share their work with a worldwide audience. Contact us at STARS@ufc.edu with any comments, questions, or suggestions. Read more.

Explore Our Collections

- Research unit, center, or department
- SelectedWorks Gallery
- Journals and peer-reviewed series

At a Glance

STARS of the Day

Back on the Agenda: The Establishment of International Criminal Court
Jennifer Ishenwood

Includes FAQ for researchers and data
For researchers:

• Who can access the materials posted in STARS?

• What if I want to restrict access to my files?

• Could I create an online searchable database of my lab data?

• I have A LOT of files. Do I have to upload them one at a time?

• Would I be able to update my data files as needed?
The collection houses all the materials related to an NSF grant including datasets, conference proceedings, articles and dissertations. The introductory text links out to the original grant proposal and includes substantial context about the project.
DMP Template from DCC

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

- Complete
- Mandatory

Template Outline

- Data Types, Formats, Standards and Capture Methods
- Ethics and Intellectual Property
- Access, Data Sharing and Rouse
- Short-Term Storage and Data Management
- Deposit and Long-Term Preservation
- Resourcing

Example funding prompts:
Describe the planned quality assurance and back-up procedures [security/storage] Specify the responsibilities for data management and curation within research teams at all participating institutions.

Guidance

Define data management support. Outline what provision is available to you within your institution and any additional skills or resources that you need to secure. If local support is available, it helps to demonstrate that you have discussed and agreed requirements. If you need to secure external support, justify the selections made and budget requested. Be clear.

Things to consider when deciding on where and how to store your data:

- Use file formats that will be usable in the long-term, not dependent on software version
- CD & DVDs' media life are not reliable in the long-term: copy or migrate data to new media between 2-5 years after creation
- Appropriate environmental conditions will increase the life-span of media
- Consider security concerns listed above
- Make backups of your data

Save Response  Save and Next
DMP Template from DCC

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

✅ = Complete
* = Mandatory

Template Outline

- Data Types, Formats, Standards and Capture Methods
- Ethics and Intellectual Property
- Access, Data Sharing and Reuse
- Short-Term Storage and Data Management
- Deposit and Long-Term Preservation
- Resourcing

Instructions

Example funder prompts: Identify which of the data sets produced are considered to be of long-term value. Outline the plans for preparing and documenting data for preservation and sharing. Explain your archiving/preservation plan to ensure the long-term value of key datasets.

Guidance

Select data of long-term value. Data sharing and preservation may not be applicable in every case. The DCC provides a 'How to …' guide on appraisal, which offers practical strategies to help you select important data. Deciding what has long-term value and preparing those data to expected standards for deposit are time-consuming processes, for which you should allocate...

Keeping reliable backups is an integral part of data management. Regular backups protect against the risk of damage or loss due to hardware failure, software or media faults, viruses or hacking, power failure, or even human errors. It is recommended that you have three backup copies (original, external but local to your institution, and external at a remote location). Check the integrity of your backed-up files using checksum validation. Create a backup schedule, use reliable backup media, AND test out your backup system by testing file restores. The UK Data Archive provides additional guidelines on data storage, back-up, and security.

- Personal Computer
- Personal Computer

Save Response • Save and Next
**DMP Template from DCC**

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

- ✔ = Complete
- * = Mandatory

### Template Outline

- ✔ Data Types, Formats, Standards and Capture Methods
- ✔ Ethics and Intellectual Property
- ✔ Access, Data Sharing and Reuse
- ✔ Short-Term Storage and Data Management
- ✔ Deposit and Long-Term Preservation
- ✔ Resourcing

### Instructions

Example prompts: What resources will you require to deliver your plan? Outline additional hardware, software and technical expertise, support and training that is likely to be required and how it will be acquired.

### Guidance

Outline and justify costs: If you need to purchase storage, outsource services such as back-up and preservation, or plan to pay for data management support, these costs should be outlined and justified in your proposal. Where institutional provision is available, show that the support you require has been discussed and agreed. It also helps to link resources with roles and responsibilities.

This plan depends on purchased storage, outsourced services such as back-up and preservation, and additional costs for data management support. A server will be required. Two partner institutions have agreed to assist with ongoing data management.

---

**Save Response**  **Save and Next**  **Cancel Changes**  **DMP Preview >>**
### Data Types, Formats, Standards and Capture Methods

Data can mean many different things, but there are typically four main categories that it can be sorted into for management purposes. The category that you choose will then have an effect upon the choices that you make throughout the rest of your data management plan.

#### Observational
- Captured in real-time
- Usually irreplaceable
- Examples: Sensor readings, telemetry, survey results, images

#### Experimental
- Data from lab equipment
- Often reproducible, but can be expensive
- Examples: gene sequences, chromatograms, magnetic field readings

#### Simulation
- Data generated from test models
- Models and metadata
- Input more important than output data
- Examples: climate models, economic models

#### Derived or compiled
FCTL Summer 2015

Data Types, Formats, Standards and Capture Methods

Data can mean many different things, but there are typically four main categories that it can be sorted into for management purposes. The category that you choose will then have an effect upon the choices that you make throughout the rest of your data management plan.

Observational
• Captured in real-time
• Usually irreplaceable
• Examples: Sensor readings, telemetry, survey results, images

Experimental
• Data from lab equipment
• Often reproducible, but can be expensive
• Examples: gene sequences, chromatograms, magnetic field readings

Simulation
• Data generated from test models
• Models and metadata
• Input more important than output data
• Examples: climate models, economic models

Derived or compiled
• Reproducible (but very expensive)
• Examples: text and data mining, compiled database, 3D models

Data formats
• Text (ascii, Word, pdf)
• Numerical (ascii, SPSS, STATA, Excel, Access, MySQL)
• Multimedia (jpeg, tiff, dicom, mpeg, quicktime)
• Models (3D, statistical)
• Software (Java, C)
• Discipline specific (FITS in astronomy, CIF in chemistry)
• Instrument specific (Olympus Confocal Microscope Data Format)

Ethics and Intellectual Property

Copyright and Intellectual Property Rights

Sharing data that you produced/collected yourself:
DMP templates and guidance

- Data Management guide
- Directorate guidance
Data management guide

http://guides.ucf.edu/data
NSF Data Management Plans

The US National Science Foundation (NSF) began requiring formal data management plans as part of the peer-reviewed funding proposal process effective January 18, 2011.

NSF Data Management Plan Requirement Overview

- Proposals submitted after January 18, 2011 must include a Data Management Plan (DMP)
- DMP should be a supplementary document of no more than two pages
- DMP must describe how the researcher will adhere to the NSF policy on sharing of results
- For full DMP policy implementation, see Grant Proposal Guide (GPG) Chapter II.C.2]
- Certain directorates and divisions have issued specific guidelines for the DMP and/or for the implementation of the data sharing policy – make sure to look at your main directorate and/or division or program solicitation website to see if any additional policies apply. For detailed information from the NSF, see the following:

Please note that if a specific program solicitation provides guidance on preparation of data management plans, such guidance must be followed:

- Biological Sciences Directorate (BIO)
  - Directorate-wide Guidance
- Education & Human Resources Directorate (EHR)
  - Directorate-wide Guidance
- Engineering Directorate (ENG)
  - Directorate-wide Guidance
- Geosciences Directorate (GEO)
  - Directorate-wide Guidance
- Mathematical and Physical Sciences Directorate (MPS)
  - Division of Astronomical Sciences
  - Division of Chemistry
  - Division of Materials Research
  - Division of Mathematical Sciences
  - Division of Physics
- Computer & Information Sciences & Engineering (CISE)
  - Directorate-wide Guidance
- Social, Behavioral and Economic Sciences Directorate (SBE)
  - Directorate-wide Guidance

http://guides.ucf.edu/data/dmp
DMP resources

• Data Management (or DMP) Research Guide
  http://guides.ucf.edu/data
• DMP Tool  https://dmptool.org/
• Open Science Framework https://osf.io/
• STARS http://stars.library.ucf.edu/
  – Researcher/data FAQ at
    http://stars.library.ucf.edu/faq.html#faq-30
• Registry of Research Data Repositories
  http://www.re3data.org/
• Schedule a Consultation
  Meet with a librarian for an in-depth one on one consultation specific to
  your research topic
Questions?

- **Lee Dotson**, Digital Initiatives Librarian, [Lee.Dotson@ucf.edu](mailto:Lee.Dotson@ucf.edu)
- **Penny Beile**, Associate Director, Information Services and Scholarly Communication, [pbeile@ucf.edu](mailto:pbeile@ucf.edu)
- **Selma Jaskowski**, Assistant Director, Library Systems & Technology, [selmaj@ucf.edu](mailto:selmaj@ucf.edu)
- **Rich Gause**, Government Information Librarian, [rich.gause@ucf.edu](mailto:rich.gause@ucf.edu)
- **Athena Hoeppner**, Electronic Services Librarian, [athena@ucf.edu](mailto:athena@ucf.edu)