Fall 11-5-2014

Data Management for Grant Seekers

Jennifer L. Thoegersen

University of Nebraska-Lincoln, jthoegersen2@unl.edu

Follow this and additional works at: http://digitalcommons.unl.edu/library_talks


http://digitalcommons.unl.edu/library_talks/110

Part of the Library and Information Science Commons

This Article is brought to you for free and open access by the Libraries at University of Nebraska-Lincoln at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Library Conference Presentations and Speeches by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.
Data Management for Grant Seekers

JENNIFER THOEGERSEN, DATA CURATION LIBRARIAN
JTHOEGERSEN2@UNL.EDU
NOVEMBER 2014
Overview

Introduction
Basic Elements of a DMP
Examine Sample DMPs
UNL Libraries Services
Data Management Resources
Introduction

2003: NIH Data Sharing Policy
2009: Interagency Working Group on Digital Data
2011: NSF requires Data Management Plans
2013: Office of Science & Technology Policy Memo
2014: OSTP Update

“The Administration is committed to ensuring that, to the greatest extent and with the fewest constraints possible and consistent with law and the objectives set out below, the direct results of federally funded scientific research are made available to and useful for the public, industry, and the scientific community. Such results include peer-reviewed publications and digital data.”
Federal Agencies Requiring DMPs

Alfred P. Sloan Foundation
Department of Energy
Gordon and Betty Moore Foundation
Gulf of Mexico Research Initiative
Institute of Education Sciences (US Department of Education)
Institute of Museum and Library Services
Joint Fire Science Program
National Institutes of Health
National Endowment for the Humanities: Office of Digital Humanities
National Science Foundation
U.S. Geological Survey
Federal Agencies Required to Submit Public Access Plans

“The Office of Science and Technology Policy (OSTP) hereby directs each Federal agency with over $100 million in annual conduct of research and development expenditures to develop a plan to support increased public access to the results of research funded by the Federal Government...each agency plan for both scientific publications and digital scientific data must contain...a plan for notifying awardees and other federally funded scientific researchers of their obligations (e.g., through guidance, conditions of awards, and/or regulatory changes);”
Federal Agencies Required to Submit Public Access Plans

HHS Agency for Healthcare Research and Quality
HHS Office of the Assistant Secretary for Preparedness and Response
HHS Centers for Disease Control and Prevention
Department of Homeland Security
Department of Defense
Department of Energy
Department of the Interior
Department of Transportation
Department of Education
Environmental Protection Agency
HHS Food and Drug Administration

National Aeronautics and Space Administration
HHS National Institutes of Health
DOC National Institute of Standards and Technology
DOC National Oceanic and Atmospheric Administration
National Science Foundation
Office of the Director of National Intelligence
Smithsonian Institution
United States Agency for International Development
United States Department of Agriculture
United States Department of Veterans Affairs
Basic Elements of a DMP

Data Description
Content & Format
Protection
Access
Preservation
Responsibility
Data Description

What data gets created by the project and in what form?

What data (raw or processed) are generated?

What data are expected to be managed by the project for sharing and archiving?

How are the data generated and how is it acquisitioned?

When are data generated and how often?

Describe the data that will be generated throughout the duration of the project.
Data Description

**APSF:** “What information products will be created in the course of this project?”

**GBM:** “What data will be collected during this project?”

**IMLS:** “Describe the digital content you will create and the quantities of each type”

**NSF:** “the types of data, samples physical collections, software, curriculum materials, and other materials to be produced in the course of the project”

**NSF-AST:** “Describe the types of data and products that will be generated in the research, such as images of astronomical objects, spectra, data tables, time series, theoretical formalisms, computational strategies, software, and curriculum materials.”

**NSF-AGS:** “Are the data...New observational data, new results from model(s), generated from previous observations or models, other (physical samples, software, curriculum materials, etc.)”
Content & Format

What data formats will be used for data generated?
What tools will be required to read the data?
What data and metadata standards will be used?
How will metadata be generated (automatically or manually, or both)?
Do you have a Data Dictionary and/or controlled vocabulary that should be shared?

Describe plans for documentation, metadata, and file formats, including identification of appropriate standards.
Content & Format

**NIH:** “the format of the final dataset, the documentation to be provided, whether or not any analytic tools also will be provided”

**JFSP:** “Specify the metadata language you plan to use to describe the data. All associated metadata must be documented in a standard metadata language appropriate to the type of data. Spatial data sets must be documented using either the FGDC version 2.0 or the ISO 19115 metadata standard. The Biological Data Profile standard (associated with FGDC) is very useful for created documentation of field- and lab-based work. We recommend use of a metadata documentation tool, e.g., Metavist”

**NSF-BIO:** “the data and metadata formats and standards used”

**NFS-AGS:** “If digital data will be made available what file format(s) will be used (ex: HDFS, NetCDF) File formats?”
Protection

Are there any data with privacy issues?

Does this work involve human subjects? What policies and procedures must be adhered to?

How will you restrict access to sensitive data during and after the project?

Is any of your data copyrightable? If so, who holds that copyright?

Is there any of the data owned by someone else?

What are the conditions of use, sharing and dissemination?

“Statement of plans, where appropriate and necessary, for protection of privacy, confidentiality, security, intellectual property and other rights.”
Protection

**DOE:** “DMPs must protect confidentiality, personal privacy, Personally Identifiable Information, and U.S. national, homeland, and economic security; recognize proprietary interests, business confidential information, and intellectual property rights; avoid significant negative impact on innovation, and U.S. competitiveness; and otherwise be consistent with all applicable laws, regulations, and DOE orders and policies. There is no requirement to share proprietary data.”

**NIH:** “If the research involves human subjects and the data are intended to be shared, the application should discuss how the rights and confidentiality of participants would be protected. In the Human Subjects section of the application, the applicant should discuss the potential risks to research participants posed by data sharing and steps taken to address those risks.”

**NSF:** “policies for access and sharing, including provisions for appropriate protection of privacy, confidentiality, security, intellectual property, or other rights or requirements”
Access

What data will be made available to whom?
How will data be made available?
When will data be made available?
Will there be any restrictions to access? Why?
Access

**NSF-AST:** “Describe your policies regarding the use of data provided via general access or sharing. For example, if you plan to provide data and images on your website, will the website contain disclaimers, or conditions regarding the use of the data in other publications or products? If the data or products (e.g., images) are copyrighted (by a journal, for example), how will this be noted on the website?”

**USGS:** “Dissemination encompasses a range of activities including advertising product availability, serving and maintaining electronic documents, conducting initial mailings (including copies to depository libraries), stocking sales outlets, and managing requests.”

**NEH:** “The DMP should describe data formats, media, and dissemination approaches that will be used to make data and metadata available to others.”
Preservation

What digital and analog data will be stored?

Where will the data be stored and backed up, what policies will be in place?

How will data be stored after the project has been completed?

How long will each type of data be kept?

Where will data be archived?

What policies will be used to manage data after the project has been completed?

Who will manage and administer the archive?
Preservation

“How long beyond the grant term will grant products be maintained and by whom? How will you ensure the long-term durability of grant products after the funding period ends?” (apsf)

“Describe how data will be archived and how preservation of access will be handled. For example, will hardcopy notebooks, instrument outputs, and physical samples be stored in a location where there are safeguards against fire or water damage? Is there a plan to transfer digitized information to new storage media or devices as technological standards or practices change? Will there be an easily accessible index that documents where all archived data are stored and how they can be accessed?” (nsf-che)

“The DMP should describe physical and cyber resources and facilities that will be used for the effective preservation and storage of research data. These can include third party facilities and repositories.” (nsf-ehr)
Responsibilities

Who makes decisions regarding the overall and day-to-day data management?

Who will be responsible for the data in the near-term following project completion?

Who is responsible for preserving the data?

Who will be responsible for the data for long-term archiving?

Describe the roles and responsibilities for management of data
Responsibilities

**APSF:** “Who will be responsible for managing project assets during the grant period?”

**NSF-CISE:** “The plan should outline the rights and obligations of all parties as to their roles and responsibilities in the management and retention of research data”

**NSF-ENG & NSF-SBE:** “It should outline the rights and obligations of all parties as to their roles and responsibilities in the management and retention of research data.”
Examination of DMPs

Take 2+ example data management plans from successful NSF grant proposals

Are the basic elements present?

Is there information included not covered by the basic elements?

Do these examples raise new questions?

How are the examples similar and different?

Data Description
Content & Format
Protection
Access
Preservation
Responsibility
UNL Libraries Services

**Consultations** on Data Management Plans

**Workshops** for department, laboratory, students, and project team

**Data Repository** for inactive data

For more information:

Visit

[libraries.unl.edu/data-management](libraries.unl.edu/data-management)

[dataregistry.unl.edu](dataregistry.unl.edu)

Email datamanagement@unl.edu
Resources

**General Data Management Information:** unl.libguides.com/datamanagement

**Sample NSF DMPs from UC San Diego:** idi.ucsd.edu/data-curation/examples.html

**Data Repository Registries:** Re3data: www.re3data.org

**Metadata Standards:** Digital Curation Centre: www.dcc.ac.uk/resources/metadata-standards

**Privacy:** Office of Research Responsibility: research.unl.edu/researchresponsibility/

**File Format Recommendations:** Library of Congress
www.digitalpreservation.gov/formats/content/content_categories.shtml

**Data Curation Profiles:** http://docs.lib.purdue.edu/dcp/

**DMPTool:** https://dmptool.org/