MEETING THE NSF DATA MANAGEMENT PLAN REQUIREMENT

Co-sponsored by IU Libraries, OVPR, and ORA
Stacy Konkiel, E-Science Librarian
skonkiel@indiana.edu
Learning Objectives

• Reasons behind the mandate
• Understand the basic requirements of the mandate
• Learn about IUB’s cyberinfrastructure and on-campus assistance in creating your data management plan
Overview

• Definitions
• NSF DMP Mandate: The Background
• Getting Prepared
• Five Requirements for your DMP
• Examples
• How IU can help you meet the mandate
• Resources
• Q&A/Feedback
Definitions

- **Cyberinfrastructure**: computing resources and networks, services, and people
- **Data management**: the technical processing and preparation of data for analysis
- **Data curation**: managing and promoting the use of data from its creation, to ensure it is fit for discovery and re-use
- **Data Sharing**: must take into account legal and ethical issues; a spectrum with many options
- **Data Citation**: mechanisms to enable easy reuse and verification, track the impact of data, and recognize and reward researchers
- **DMP** = Data Management Plan

(Coates, 2012)
NSF DATA MANAGEMENT PLAN MANDATE: THE BACKGROUND
Historical Context

Similar policies at Wellcome Trust, Howard Hughes Medical Institute, NOAA, NEH:

http://www.cdlib.org/services/uc3/datamanagement/funding.html
Why do we have a data sharing mandate?
Why we have a data sharing mandate

“Such dissemination of data is necessary for the community to stimulate new advances as quickly as possible and to allow prompt evaluation of the results by the scientific community.” – NSF

- Accelerate scientific discovery
- Reproducible results
- ROI
Why we have a data sharing mandate

- **Organization = Easier Work**
- **Replicated Data = Safe(r) Data**
  - Digital data is more fragile than analog data
- **Open Data = More Citations** (Piwowar et al, 2010)

(Houston, 2011)
Why we have a data sharing mandate

• “Investigators are expected to share with other researchers, at no more than incremental cost and within a reasonable time, the primary data, samples, physical collections and other supporting materials created or gathered in the course of work under NSF grants. Grantees are expected to encourage and facilitate such sharing.”
Why we have a data sharing mandate

• “Investigators are expected to share with other researchers, at no more than incremental cost and within a reasonable time, the primary data, samples, physical collections and other supporting materials created or gathered in the course of work under NSF grants. Grantees are expected to encourage and facilitate such sharing.”
What is “a reasonable amount of time”?

• Engineering Section: “no later than the acceptance for publication of the main findings of the final data”
• Earth Sciences: “No later than two (2) years after the data were collected.”
• Social and Economic Sciences: “within one year after the expiration of an award”
HOW TO PREPARE
DMP Basics

• No more than two pages
• Supplementary document: does not count towards page limit
• Even if no data produced, must submit a DMP
How to Prepare

• Data Inventory
• Audiences
• Obligations
  • Open Data? Intellectual Property? Confidentiality?
• Enduring value?
FIVE REQUIREMENTS FOR YOUR DMP
Five Requirements for your DMP

• Types of data
• Metadata & Data Formats
• Access and Sharing
• Reuse and Distribution Policies
• Preservation
Requirements: Types of Data

“The types of data, samples, physical collections, software, curriculum materials, and other materials to be produced”

• List any and all
  • Observational
  • Experimental
  • Simulation
  • Derived or compiled

• Be specific
Requirements: Metadata

“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...)

- “Data about data”
- Metadata: basic information about data set(s)
- Preservation metadata: assure quality and provenance of data set(s)
- Guiding questions
Requirements: Metadata

- Use existing standards and controlled vocabularies
- Where standards don’t exist, make note!
- Make metadata central to your study design
- Supply minimum information relevant to help others understand and access your data
- Consider supplying preservation metadata
  - Technical specifications
  - MD5 checksums
Requirements: Data Formats

“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…)

• Describe how your data will be recorded and stored
  • XLS or CSV? CML? ArcInfo Coverage?
  • Be specific!

• Common formats above all else
• The more open/interoperable, the better
Requirements: Access and Sharing

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

- With whom/how will you share?
- Will you “open it up” after time? When?
- Encrypt and store your ePHI and other data subject to IRB, HIPAA, FRPAA, etc regulations
Requirements: Reuse & Distribution Policies

“policies and provisions for re-use, re-distribution, and the production of derivatives”

- IU Legal Counsel is final word

- Recommend:
  - Open Data Commons Attribution License
  - Creative Commons Zero License

- Resources
  - Digital Curation Centre's "How to License Research Data"
  - Open Definition's list of recommended data licenses
Requirements: Preservation

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

- Standard at IUB: “At least three years beyond the end of the project”
  - Certain types should be kept longer: patent data, longitudinal data sets
  - Check your directorate requirements!

- Physical samples & Digital data

- Who assumes responsibility?
  - As formats become obsolete, who will migrate data?
    - Think .BMP to .JPG to .PNG
  - As media degrades, who will transfer data?
  - In the future, who will know enough about your research (or even your field) to work with your data?
    - Preservation metadata = Rosetta Stone
FAQs

• If my data is freely available, how will I ensure that I am credited for my work?
• What if my research doesn’t produce data?
• What if it uses existing data?
• Do I have to make my data publicly available?
• How long do I need to keep my data?
• If data or samples are requested before I have completed all analyses on them, must I share them?

→ http://1.usa.gov/MWv5ff <--
EXAMPLES
Example: Atmospheric Sciences

Atmospheric CO$_2$ Concentrations, Mauna Loa Observatory, Hawaii, 2011-2013

Example: Ecology

The influence of plant functional types on ecosystem responses to altered rainfall

Image via http://commons.wikimedia.org/wiki/File:Rainfall_in_Amravati.jpg
HOW CAN IU HELP ME MEET THE MANDATE?
How IU can help you meet the mandate

IU Empowering People Strategic Plan for IT: Action 33:

“IU should provision a data utility service for research data that affords abundant near- and long-term storage, ease of use, and preservation capabilities. This data utility will need to offer a range of services for securing data, providing authorized access within and beyond IU; ensuring metadata description, annotation, and provenance; and providing backup/recovery services.”
How IU can help you meet the mandate

**Staff Expertise**
- Developing your proposal (ORA/PDS)
- Metadata, Checking your DMP (Libraries)
- **Depositing and Preserving** Data (Libraries & UITS)

**Cyberinfrastructure**
- Research File System (UITS)
- Scholarly Data Archive (UITS)
- IUScholarWorks Repository (Libraries)
IUB Data Management Service

• Preparing your data
• Storage (HIPAA-compliant)
• Preservation
• Access
• Data Management Plan consultations
  • Lab-based
  • Funder required
Resources

- Metadata, Data Formats, etc Consultations
- Data Storage and Preservation
  - IUB Data Management Service (UITS & Libraries)
- Privacy, Confidentiality, Reuse & Distribution Policies
  - IU Research Policies
  - Legal Counsel
- Data Management Guide, IU Resources, Boilerplate Text
  - http://libraries.iub.edu/data
- Data Management Plan Consultations
  - Stacy Konkiel, E-Science Librarian
  - iuswdata@indiana.edu
Resources

• Proposal Development Help, Grant Compliance
  • Proposal Development Services
  • Office of Research Administration
• Responsible Conduct of Research (RCR) & DM
  • Poynter Center for Research Ethics
  • RCR classes via ORA
Q&A / Feedback / References

Download this presentation at:

> http://hdl.handle.net/2022/14722 <

