Effective Data Stewardship: From Planning to Implementation

Presented by Renaine Julian as part of “Managing People and Data”
May 13, 2016
Overview

Defining Research Data

Data Management and the Research (data) Lifecycle

Federal Funder Mandates

Data Repositories

Help with RDM and other library resources
What is Research Data

“Data are outputs of research and inputs to scholarly publications and inputs to subsequent sharing and learning” (Borgman 2007)
What is Research Data

“...The recorded factual material commonly accepted in the scientific community as necessary to validate research findings.” (2 CFR 200.315(3))
What is not Research Data?

Preliminary Analysis
Drafts of papers
Plans for future research
Peer reviews
Communication with colleagues
What is **not** Research Data?

Trade secrets, commercial information, materials necessary to be held confidential by a researcher until they are published, or similar information which is protected under law.

Personnel and medical information and similar information the disclosure of which would constitute a clearly unwarranted invasion of personal privacy, such as information that could be used to identify a particular person in a research study.
What is Research Data Management

“Research Data Management is part of the research process, and aims to make the research process as efficient as possible, and meet expectations and requirements of the university, research funders, and legislation”

Source: University of Leicester, http://www2.le.ac.uk/services/research-data/rdm/what-is-rdm
“And yet, research data is the currency of science, even if publications are still the currency of tenure. To be able to exchange data, communicate it, mine it, reuse it, and review it is essential to scientific productivity, collaboration, and to discovery itself” (Gold 2007)
Why Manage Research Data

Transparency & Integrity
Personal Benefit
Reproducibility
Compliance
Why Manage Research Data

“The Administration is committed to ensuring that...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” (Holdren 2013).
Data Sharing Policy: 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.

Data Management Plan Requirements: As of January 18, 2011, a separate data management plan (DMP) limited to two pages is required for all proposals submitted to NSF. The DMP should include specific details of data standards, accessibility, electronic dissemination, and sustainability.

Directorates from the NSF have their own data management requirements.
1. Describe the types of data, physical samples or collections, software, curriculum materials, and other materials to be produced in the course of the project.

2. Describe the standards to be used for all the data types anticipated, including data or file format and metadata (where existing standards are absent or deemed inadequate, this should be documented along with any proposed solutions or remedies)

3. Describe the roles and responsibilities of all parties with respect to the management of the data

4. Describe the dissemination methods that will be used to make data and metadata available to others during the period of the award, and any modifications or additional technical information regarding data access after the grant ends.

5. Describe the PI’s policies for data sharing, public access and reuse, including re-distribution by others and the production of derivatives. Where appropriate, include provisions for protection of privacy, confidentiality, security, intellectual property rights and other rights.

6. Where relevant, describe plans for archiving data, samples, software, and other research products, and for ongoing access to these products through their lifecycle of usefulness to research and education. Consider which data (or research products) will be deposited for long-term access and where.
NIH Public Access Policy for publications has been a requirement since 2008. Investigators seeking $500,000 or more in direct costs in any year should include a description of how final research data will be shared, or explain why data sharing is not possible.

It is expected that the data sharing discussion will be provided primarily in the form of a brief paragraph immediately following the Research Plan Section of the application form.
Things to include in the Data Sharing Plan

- Description of data collected/created
- Data formats and types
- Plan for disseminating research data
- Timeline for disseminating research data
- Description of data documentation
- Description of Data Sharing Agreement (if applicable)

NIH Data Sharing Policy and Implementation Guidance
National Oceanographic and Atmospheric Administration

1. Type of data and information created
2. Expected Schedule for data sharing
3. Standards for format and content
4. Policies for stewardship and preservation
5. Procedures for providing access
6. Previously published data
Beginning Oct 1, 2015, each DOE sponsoring research office will ensure that the requirements for DMPs are included in all solicitations and invitations for research funding with details about how and when a DMP should be submitted.
DMPTool

Developed by California Digital Library

Data Management Templates by Funding Agency

FSU as a Partner Institution

Community Involvement and Participation
Data Repositories

Preservation
Curation
Interoperability
Support Reproducibility
Citation and Linking
Access Control (restricted-use data)
Community engagement
Types of Data Repository

Research Repository

Domain (subject) Repository

National Repository

Institutional Repository
RDM Services at FSU

Data Management Plan Assistance

Data Management Planning Tools

Consultation on Data Management Implementation
Resources for PostDocs at FSU

Collections: Journals, Databases, Books, Data!

Citation Management

Impact Analysis

Scholarly Publishing


University of Leicester Research Data Definitions: https://www2.le.ac.uk/services/research-data/documents/UoL_ResearchDataDefinitions_20120904.pdf


