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))

National Archives and Records Administration

code of federal regulations

What is not Research Data?

Preliminary Analysis

Drafts of papers

Plans for future research

Peer reviews

Communication with colleagues

What is not Research Data?

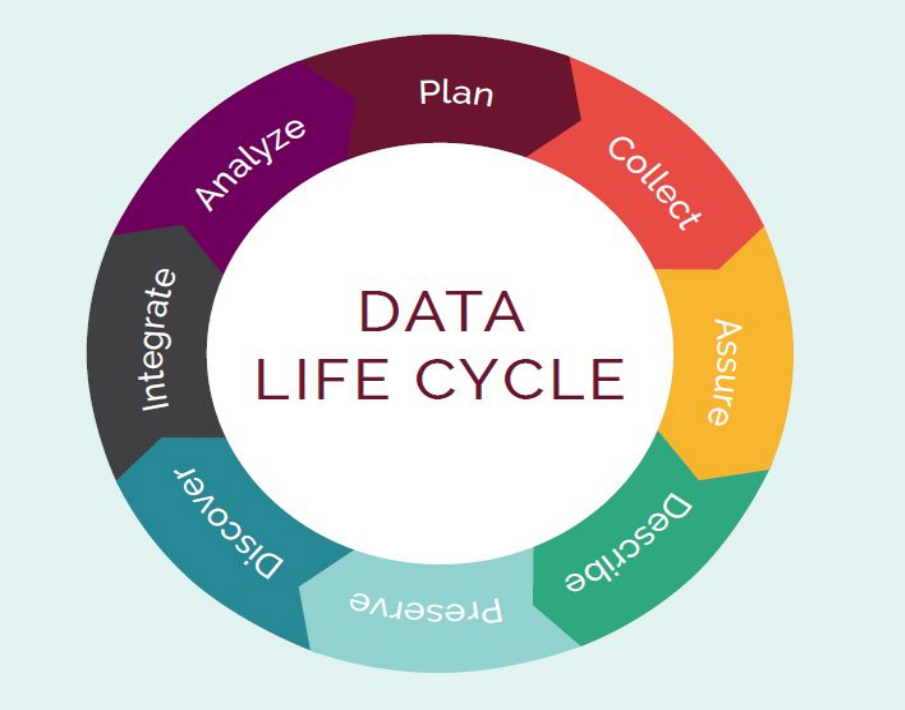
Trade secrets, commercial information, materials necessary to be held confidential by a researcher until they are published, or similar information which is protected under law

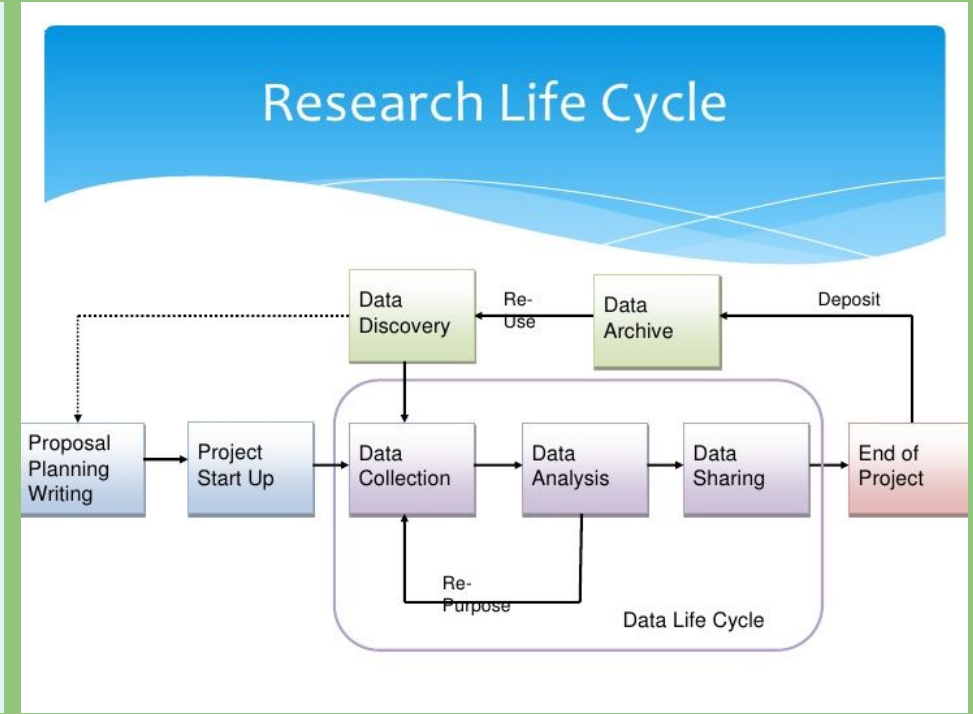
Personnel and medical information and similar information the disclosure of which would constitute a clearly unwarranted invasion of personal privacy, such as information that could be used to identify a particular person in a research study.

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


Plan
Collect
Assure
Describe
Preserve
Discover
Integrate
Analyze
DATA
LIFE CYCLE


Research Life Cycle
Data Discovery
Re-Use
Data Archive
Deposit
Proposal Planning Writing
Project Start Up
Data Collection
Data Analysis
Data Sharing
End of Project
Re-Purpose
Data Life Cycle

Why Manage Research Data?

“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).

BILL & MELINDA GATES *foundation*

ALFRED P. SLOAN FOUNDATION

National Science Foundation

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.

National Science Foundation

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 on-going 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.

National Institutes of Health

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

NIH Data Sharing Policy and Implementation Guidance

National Institutes of Health

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

U.S. Department of Energy

1. Data Types and Sources
2. Content and Format
3. Data Sharing and PReservation
4. Protection
5. Rationale

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

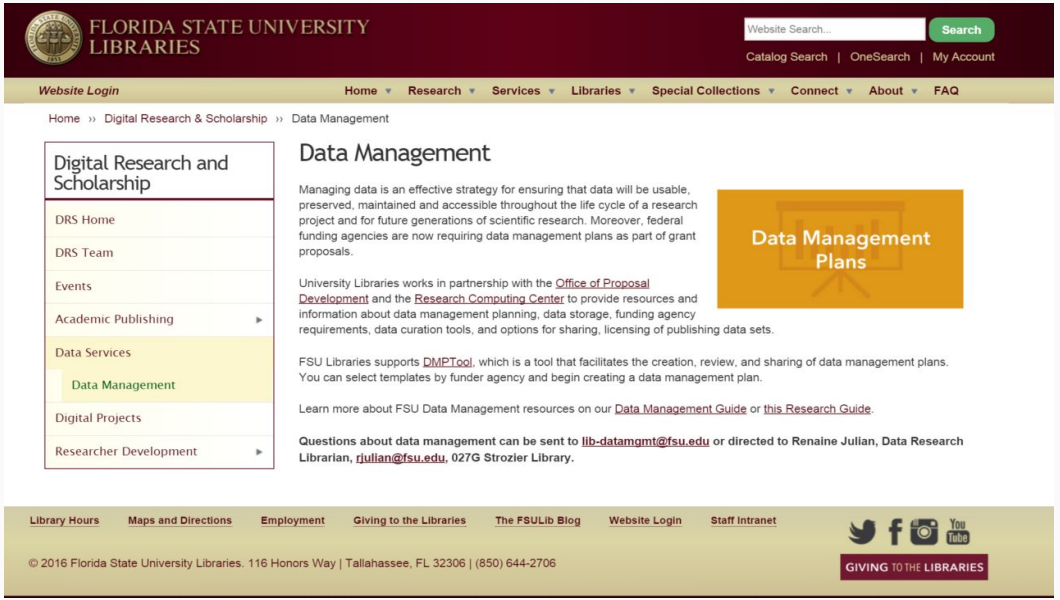
Institutional Repository

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

References

Borgman, C.L. (2007) Scholarship in the Digital Age : Information, infrastructure, and the Internet. Cambridge, MA. : MIT Press

Office of Management and Budget Requirements, Uniform Administrative Requirements, Cost Principles, and Audit Requirements for Federal Awards, 2 CFR 200.315 (2016)

Gold, A. *Cyberinfrastructure, Data, and Libraries, Part 1: A Cyberinfrastructure Primer for Librarians.* D-Lib Magazine, September/October, 2007, Volume 13 Number 9/10 http://www.dlib.org/dlib/september07/gold/09gold-pt1.html

MIT Libraries. (2013) "Data Management and Publishing." MIT http://libraries.mit.edu/guides/subjects/data-management/index.html .

University of Leicester Research Data Definitions: https://www2.le.ac.uk/services/research-data/documents/UoL_ReserchDataDefinitions_20120904.pdf

Office of Research Integrity (2001) "Data Management." United States Department of Health and Human Services. United States Federal Government. http://ori.hhs.gov/education/products/rcradmin/topics/data/open.shtml.

National Science Foundation. *Dissemination and Sharing of Research Results.* http://www.nsf.gov/bfa/dias/policy/dmp.jsp

National Institutes of Health. *NIH Data Sharing Policy and Implementation Guidance. http://grants.nih.gov/grants/policy/data_sharing/data_sharing_guidance.htm*

National Oceanographic and Atmospheric Administration. (2015) *NOAA Plan for Increasing Public Access to Research. http://docs.lib.noaa.gov/noaa_documents/NOAA_Research_Council/NOAA_PARR_Plan_v5.04.pdf*

U.S. Department of Energy. (2015) *DOE Public Access Plan http://www.energy.gov/downloads/doe-public-access-plan*

ICPSR, 2013. "Sustaining Domain Repositories for Digital Data: A White Paper". http://datacommunity.icpsr.umich.edu/sites/default/files/WhitePaper_ICPSR_SDRDD_121113.pdf

Armbruster, Chris and Romary, Laurent, Comparing Repository Types: Challenges and Barriers for Subject-Based Repositories, Research Repositories, National Repository Systems and Institutional Repositories in Serving Scholarly Communication (November 23, 2009). Available at SSRN: http://ssrn.com/abstract=1506905or http://dx.doi.org/10.2139/ssrn.1506905