Data Management 101

Young-Joo Lee
Senior Clinical Librarian

October 27, 2016
Disclosure

Nothing
Learning Objectives

✧ Describe why DMP (Data Management Plan) is needed
✧ Utilize best practices
✧ Locate funder mandates
✧ Identify key elements in DMP
✧ Locate DMP tools/resource (on campus and beyond)
I. INTRODUCTION

Basics of Research Data Management
Laugh or cry?

https://www.youtube.com/watch?v=66oNv_DJuPc
What is research data?

✧ All manner of things that researchers produce in the course of their work.
Why collect research data?

- For the purpose of analysis to produce and validate original research results.
Key research data formats

- Documents (e.g. text, Word, PDF)
- Lab notes, fields notes
- Spreadsheets (e.g. xml, csv)
- Statistical data sets (e.g. SPSS, Stata, SAS)
- Recording (e.g. mp3, wav)
- Photo and other images (e.g. JPG)
Why manage data?

Funder mandate?

NIH Data Sharing Policy

Data sharing is essential for expedited translation of research results into knowledge, products and procedures to improve human health.

The Final NIH Statement on Sharing Research Data was published in the NIH Guide on February 26, 2003. This is an extension of NIH policy on sharing research resources, and reaffirms NIH support for the concept of data sharing.

The new policy becomes effective with the October 1, 2003 receipt date for applications or proposals to NIH.

- Data Sharing Regulations/Policy/Guidance Chart for NIH Awards (08/30/2006) - (MS Word - 58 KB) - This chart is designed as a quick guide only for the purpose of identifying various data sharing regulation/policy/guidance documents applicable to NIH funding.
Why manage data?

Publisher mandate?
Why manage data?

♦ Maintain data integrity
♦ Save time
♦ Enhance data security & minimize risk of data loss
♦ Increase your research impact
II. BEST PRACTICES

1. Data Collection: What & How
“I am going to collect a MRI.”
Just a MRI?

- images
- figures
- Spread sheet
Workflow

✧ How is the data collected?
✧ Who collects data?
Quality Assurance

How to ensure the quality of data?

- Lab notebook
- Excel
- REDCap
II. BEST PRACTICES

2. Organization of Data
Example:

Body mass index

\[
\text{BMI} = \frac{\text{weight}}{\text{(height x height)}} = \frac{\text{kg}}{\text{cm}^2}
\]
File naming convention

✧ File names should stand alone.

Brain

PDF1
PDF2
PDF3

Au_Yr_TK_J
Version Control

- Cancer Paper
- Cancer Paper Revised
- Cancer Paper Revision_YJ
- Cancer Paper Final
- Cancer Paper Final Final
Workflow

Data Dictionary

Meta Data

Authors, date of pub, URL ...

Section 3.2.3 and Grant Proposal Guide, Chapter II C.2.f.i (c)
II. Best Practices

3. Storage & Sharing of Data
Security

✦ Back up
✦ Permission
Preservation & Archiving

✧ Where is data stored?
✧ For how long?
✧ Who can access data?
Is a special hardware or software needed to use the data?

- Data format – will be best if not proprietary
- NSF guideline: “little money...”
Who owns your data?
Repository

🌟 You can deposit the data in multiple places.
III. DMP

Ready for DMP (Data Management Plan)?
A document (≈ 2 pages) that describes what you will do with your data *during* your research and *after* you complete it.
DMP (Cont’d)

Approximately 2 pages

✧ Part one
  A plan of how you’ll manage the data

✧ Part two
  How it will meet the requirement
DMP examples

- NEH
- NSF
- NIH
Final Take Away

✧ Does your data answer your question?
✧ Can it be answered by using existing data?
Further resources

http://hsl-howard.beta.libguides.com/data

What is DMP?
Many federal agencies, including NIH and NSF, now require a DMP (data management plan) for grant applications. The essential elements of a DMP include a description of the data, metadata, a data storage and sharing plan, preservation and associated legal and ethical issues. This guide is intended to help you learn some basics about DMP and so that you can create one of your own.

Data Mgmt 101
The first step in data management is to ask this crucial question: **What data will enable me to answer the question I am asking?**

For there, you create a data management plan (DMP). Start by jotting down answers to the following questions:

- **NEH-ODH:** samples
- **NIH:** sample plans
- **NIH:** genomic data sharing plans
- **NSF templates**
- **ICPSR:** natural sciences examples
  - NEH template
  - NIH template

DMP Templates & Samples

How to Find Data

- CDC FastStats
- Data.gov
- Division of Health Stat, Virginia
- ICPSR
- National Center for Health Statistics
- Statistical Insight
- The World Bank

Howard Resources

- Institutional Review Board
- Office of Research Development
- RCMI Center for Computational Biology and BioInformatics

If you need assistance with statistical analysis, please contact:

John Kwagyan, Ph.D.
Director, Biostatistics Epidemiology &