NC STATE UNIVERSITY

DATA MANAGEMENT PLANNING: THE BASICS

NCSU Libraries
May 14, 2013

Will Cross
william_cross@ncsu.edu

Dave Zwicky
david_zwicky@ncsu.edu

AGENDA, OBJECTIVES & OUTCOMES

- **Agenda & objectives**
  - Introductions
  - What is a DMP?
  - Elements of a DMP
  - Services and support at NCSU
- **Outcomes**
  - Understand components of a data management plan
  - Learn about DMPTool
  - Learn about available resources
  - Handouts – Checklist, Resources, NSF FAQ, Example DMP, NSF ENG Template

INTRODUCTIONS

- Name
- Department or affiliation
- Your research (in brief)
- One thing you'd like to learn today

Courtesy of Flickr user Scoobyfoo

WHAT IS A DATA MANGEMENT PLAN (DMP)?

WHAT: A DMP is a document that describes how you will organize, store, and share your research data – during and after you complete your research.

WHO REQUIRES A DMP:

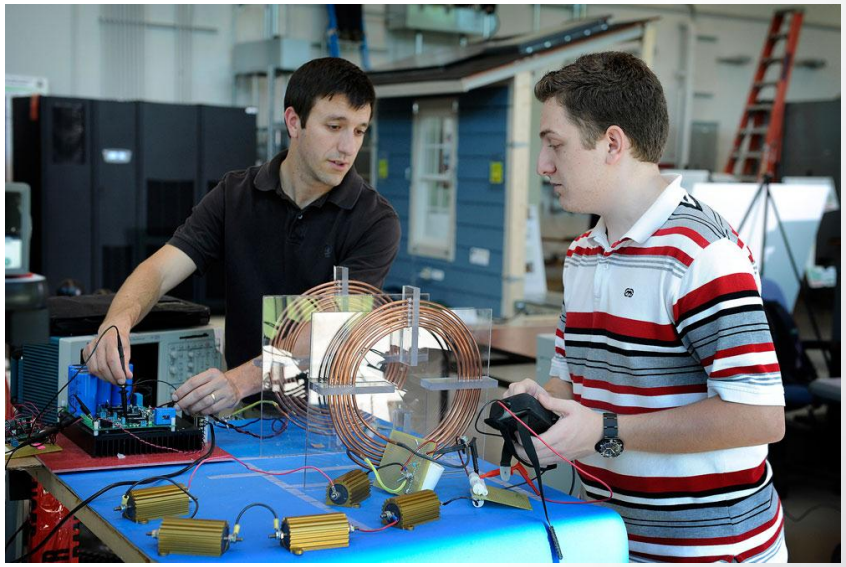
- Funders (e.g., NSF, NEH, NOAA, NIH, GBMF)
- Journals require data sharing (e.g., PLoS, PNAS, Nature, Dryad)
- Institutional Review Boards

WHY:

- Preserve your data for transparency and replication
- Facilitate new discoveries via re-use
- Increase visibility of your work

PRODUCTS OF RESEARCH or DATA TYPES

- What kind of data (observations) are you generating? How much?
- What types of data?
- How are you creating or capturing the data?
- Are you using existing data?



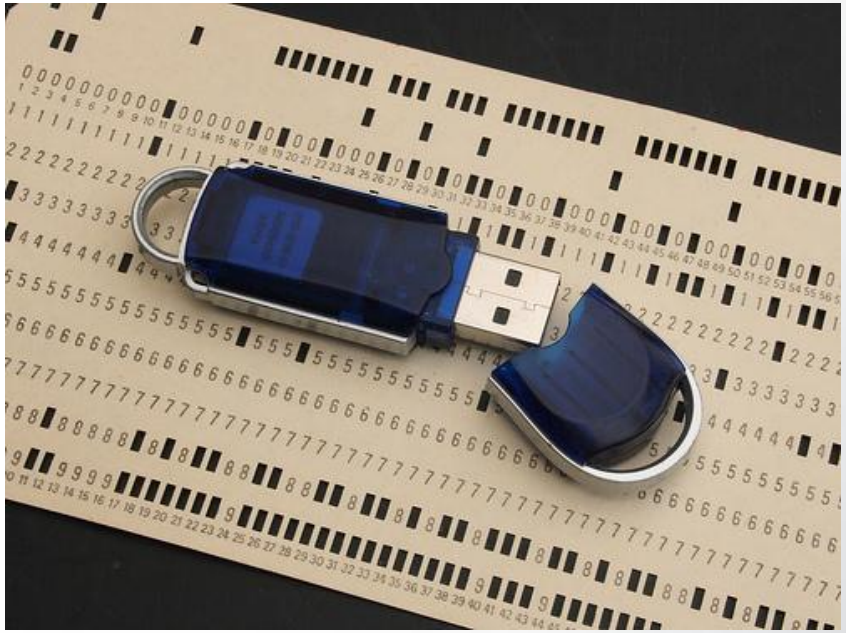
PRODUCTS OF RESEARCH or DATA TYPES

POSITIVE: “This research will generate data resulting from sensor recordings (i.e. earth pressures, accelerations, and wall deformation) during the centrifuge experiments. In addition to the raw, uncorrected sensor data, converted and corrected data (in engineering units) as well as other forms of derived data will be produced. Metadata that describes the experiments with their materials, loads, experimental environment, and parameters will be produced. The experiments will be recorded with still cameras and video cameras. Photos and videos will be part of the data collection.”

NEGATIVE: “The nature of the data or other materials produced under this NSF-sponsored project will include data characteristics such as observational, experimental, reference, derived, simulated and/or other. The data types referenced could include data generated by computer, data collected from sensors or instruments, images, video files, reports and/or other.”

DATA STORAGE & PRESERVATION

- How long should data be kept?
- Where will you deposit your data?
- What procedures are in place for preservation and backup?
- What transformations will be necessary to prepare data for preservation or sharing?
- What documentation is needed to make the data reusable?



Courtesy of Flickr user Ian-S

DATA STORAGE & PRESERVATION

POSITIVE: “For archiving, the data along with any related publications will be deposited in University A's institutional repository with an appropriate licensing statement. DOIs will be attached to all data stored for this project. Since the current preservation plan for the repository is indefinite data storage, preservation of access is assured.”

MIXED: “The infrastructure provided by Agency B is not yet either designed or suitable for distribution of software that evolves and changes over time. Instead, we will release and distribute GPL versions of data analysis and numerical modeling software as tarballs directly from our respective webpages.”

NEGATIVE: “All data will be available at request immediately after the scientific results are published and will be stored at least another five years on magnetic and optical storage devices (hard disks, CDs, DVDs). Optical storage devices will serve the purpose of the Disaster Recover Plan.”

DATA FORMATS & METADATA

- Which file formats will you use and why?
- How will you create or capture the contextual details that describe your data ("metadata")
- What metadata standards will you use?
- What other details will be necessary for others to work with your data?

Courtesy of John Vickery

DATA FORMATS

POSITIVE: “The data will be in standard, non-proprietary file formats to facilitate both data sharing and long-term preservation. The simulation code will be developed in C and provided to the public in source code format for non-commercial use under GNU GPL. The numerical data will be in TXT format so that they are readable by any text editing software. Data and documents in these formats (C, TXT) will likely be standard for a long time due to their wide use. In the situation that these formats become obsolete, we will reformat the documents.”

NEGATIVE: “The data format includes digital data recorded by computers and instruments and metadata recorded in lab notebooks and reports.”

POSITIVE: “Whenever possible, standard formats of data will be used, e.g.: images: TIFF, BMP, JPG.”

NEGATIVE: “Data will be stored in the following formats: Acrobat PDF and Microsoft Excel, PowerPoint, and Word.”

METADATA

POSITIVE: “Metadata will include time, date, and location of measurement, object measured, THz equipment used, and personnel present.”

NEGATIVE: “In addition to data, metadata concerning how the data was generated will be maintained and disseminated on request.”

POSITIVE: “For each code made available, a user's manual will be provided with instructions for compiling the source codes, installing and running the codes, formulating input data streams, and visualizing the output. Documentation will be in TXT format.”

NEGATIVE: “Embedded comments in English will be used to provide sufficient metadata to interpret the meaning of the Microsoft Word files. We have chosen this metadata format because it is flexible and enables later interpretation.”

DATA DISSEMINATION, SHARING & ACCESS

- How/when/how long will data be available?
- Are there any reasons *not* to limit or withhold the data?
- Are there ethical or privacy issues? If so, how will these be resolved?
- What have you done to comply with NCSU's IRB protocols?
- Who will hold the intellectual property rights to the data?

Courtesy of commercialintegrator.com

DATA DISSEMINATION, SHARING & ACCESS

POSITIVE: “These data will provide a detailed experimental look at the mechanical regulation of mesenchymal stem cell osteogensis. The data will further delineate the functional role of the cytoskeletal-focal adhesion-extracellular matrix signaling axis in the mechanoresponsive mesenchymal stem cell osteogenesis, as described in the main body of the proposal. As such, they will be of interest to the tissue engineering and regenerative medicine communities.”

NEGATIVE: “No, or very little, derivative use of the data is expected since the research is aimed at developing machining process models and the data gathered is directed towards this end.”

POSITIVE: “Data will be uploaded into the Project A repository and will conform to Project Data Sharing and Archiving Policies. These policies also define data confidentiality during the research, but require the data be available to the broader community 12 months after experiment completion.”

NEGATIVE: “Data will be shared by mailing a CD to the requestor.”

ROLES & RESPONSIBILITIES

- Who is responsible for implementing the DMP?
- How will responsibility for the data be transferred?
- Who is responsible for long-term decisions about the data?
- What training is necessary to manage the data?

Courtesy of conferencenike.com

ROLES & RESPONSIBILITIES

POSITIVE: “For the proposed collaborative research, Dr. A from University B will take the lead and responsibility for coordinating and assuring data storage and access. However, Dr. C from University D will also be involved in managing, storing, and disseminating the results of the project, particularly in regards to data acquired through testing associated with the E device.”

NEGATIVE: “All the investigators involved in the proposed project have equal rights to access the data generated through this sponsored project. They also have the same obligations to share data with each other and to publish the results in a timely manner.”

POSITIVE: “Should the PI leave University F, the grant would likely be transferred. If not, the co-PI will assume leadership of the project and responsibility for the data.”

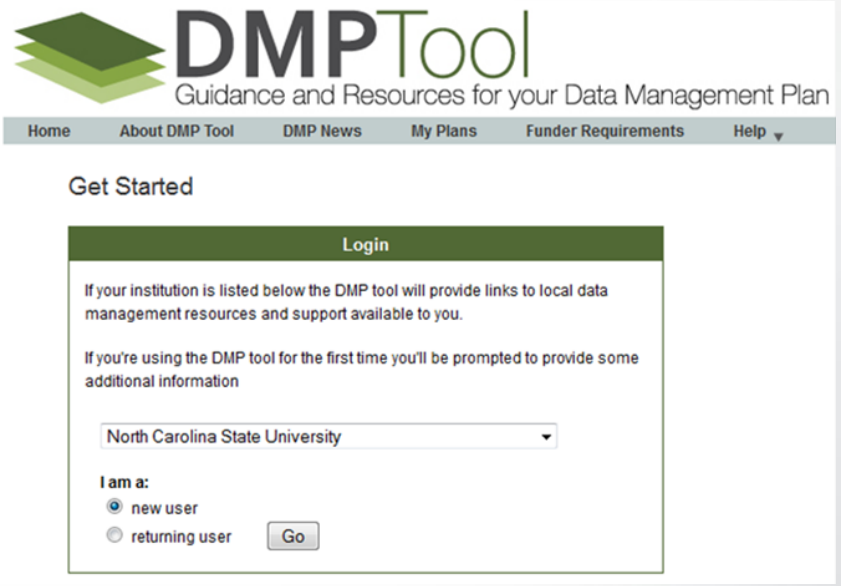
NEGATIVE: “There will be no change to the scope of this data management plan should the PI or co-PI leave University G.”

NCSU DATA MANAGEMENT PLANNING GUIDE

lib.ncsu.edu/guides/datamanagement

DMPTOOL

- Free scratch space for writing a DMP
- Step-by-step guidance
- Log in with NCSU Unity ID
- Template for each funding agency
- Links embedded



NCSU SERVICES & SUPPORT: IP & COPYRIGHT

- **What do you own?**
  - No © for raw data
  - Supplementary materials
- **How will you share?**
  - Preference for open access
- **NCSU's Copyright & Digital Scholarship Center**
  - lib.ncsu.edu/cdsc

NCSU SERVICES & SUPPORT

- **Office of Information Technology**
  - Shared services: oit.ncsu.edu/ac
  - Storage & backup, HPC, software licensing, IT security, web hosting, networking
- **Research Administration**
  - SPARCS: research.ncsu.edu/sparcs
  - Research integrity, compliance, proposal development, IRB, REACH NC, funding opportunities, scholarly data ownership
- **NCSU Libraries**
  - DMP Guide & DMPTool: lib.ncsu.edu/guides/datamanagement
  - Consultation and guidance for storage & file naming, metadata, citing data, sharing data, data repositories
  - Subject librarians ready to help!

oit
Office of Information Technology

SPONSORED PROGRAMS + REGULATORY COMPLIANCE

NCSU LIBRARIES

SUMMARY & QUESTIONS

- NCSU DMP Guide: lib.ncsu.edu/guides/datamanagement
- DMPTool
- Contact us at library_datamanagement@ncsu.edu

Courtesy of Flickr member Gonzalo AR

NCSU LIBRARIES

Positive and negative examples via “NSF ENG Data Management Plan Template.” ©2011, Regents of the University of Michigan and subject to a Creative Commons Attribution 3.0 license.