Introduction to Data Management

Mark Stoffan
Head of Digital, Access, and Technology Services

Liz Skene
Digital Initiatives Librarian

Hunter Library
Introduction to Data Management

• Agenda
  – What is a Data Management Plan?
Introduction to Data Management

• Agenda
  – What is a Data Management Plan?
  – Why should I care?
Introduction to Data Management

- Agenda
  - What is a Data Management Plan?
  - Why should I care?
  - Overview of the basics
Introduction to Data Management

• Agenda
  – What is a Data Management Plan?
  – Why should I care?
  – Overview of the basics
  – Introduction to DMPTool
Introduction to Data Management

• **Agenda**
  – What is a Data Management Plan?
  – Why should I care?
  – Overview of the basics
  – Introduction to DMPTool
  – Using repositories
Introduction to Data Management

• What is a Data Management Plan?

A Data Management Plan (DMP) is the part of a grant proposal where researchers describe the types of data they will collect and how they plan to maintain and share that data.
Introduction to Data Management

- Why should I care??
  - Part of the scholarly record
Introduction to Data Management

- Why should I care??
  - Part of the record
  - Increase the visibility and impact of your research
Introduction to Data Management

• Why should I care??
  – Part of the scholarly record
  – Increase the visibility and impact of your research
  – Safeguard your research
Introduction to Data Management

• Why should I care??

  – Part of the scholarly record
  – Increase the visibility and impact of your research
  – Safeguard your research
  – Meet grant requirements for sharing and curation
Introduction to Data Management

Origins

– Federal law: on August 9, 2007, President Bush signed the America COMPETES Act (or the "America Creating Opportunities to Meaningfully Promote Excellence in Technology, Education, and Science Act").
– Requires federal agencies to provide guidelines, policy and procedures for open exchange of data and research between agencies, the public and policymakers.
– Publicly-funded research should benefit the public.
Dissemination and Sharing of Research Results

NSF 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. See Award & Administration Guide (AAG) Chapter 11.0.1.

NSF Data Management Plan Requirements

Proposals submitted or due on or after January 10, 2011, must include a supplementary document of no more than two pages labeled "Data Management Plan". This supplementary document should describe how the proposal will conform to NSF policy on the dissemination and sharing of research results. See Grant Proposal Guide (GPG) Chapter II.C.3.i for full policy implementation.

Requirements by Directorate, Office, Division, Program, or other NSF Unit

Links to data management requirements and plans relevant to specific Directories, Offices, Divisions, Programs, or other NSF units, are provided below. If guidance specific to the program is not provided, then the requirements established in Grant Proposal Guide (GPG) Chapter II.C.3.i apply.

Please note that if a specific program solicitation provides guidance on preparation of data management plans, such guidance must be followed.

- Biological Sciences Directorate (BIO)
  - Directorate-wide Guidance
- Computer & Information Sciences & Engineering (CISE)
  - Directorate-wide Guidance
- Education & Human Resources Directorate (EHR)
  - Directorate-wide Guidance
- Engineering Directorate (ENG)
  - Directorate-wide Guidance
- Geosciences Directorate (GEO)
  - Directorate-wide Guidance
Introduction to Data Management

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.


- NIH Data Sharing Policy and Implementation Guidance (03/05/2003) - Guidance providing the NIH policy statement on data sharing and additional information on the implementation of this policy.

Frequently Asked Questions - Data Sharing (02/16/2004) - Listing of Frequently Asked Questions that will be updated as new questions are received. Please check back periodically for new questions and answers.

- [data sharing Workbook](#) (PDF - 75 KB) or [MS Word - 74 KB] (02/16/2004) - Workbook to show how investigators working in a variety of scientific areas have shared their data.

- [NIH Data Sharing Brochure](#) (PDF - 244 KB) (05/20/2003) - Printable brochure that summarizes main elements of the NIH Data Sharing Policy.

- [Testimonials](#) (MS Word - 22 KB) (03/05/2003) - First-hand accounts from researchers who have shared data.

- [Other Data Sharing Documents and Resources](#) (02/19/2004) - Additional resources relating to data sharing.

Western Carolina University
Introduction to Data Management

- Many other funding agencies, institutions, and publication venues have policies regarding data sharing because openness and transparency are considered by many to be part of the scientific method.
- Data and methods may be requested from authors years after publication.
- With technology-based developments in data curation long-term preservation is becoming more common.
Introduction to Data Management
Introduction to Data Management

• So, what do I do?
Introduction to Data Management

• Think about…
  – What types of data are being produced? Do you have the IT resources you need to collect the data?
Introduction to Data Management

• **Think about…**
  – What types of data are being produced? Do you have the IT resources you need to collect the data?
  – What file formats will be used? Is there special software needed to read the data?
Introduction to Data Management

• Think about…
  – What types of data are being produced? Do you have the IT resources you need to collect the data?
  – What file formats will be used? Is there special software needed to read the data?
  – What directory and file naming conventions will be used?
Introduction to Data Management

• Think about…
  – What types of data are being produced? Do you have the IT resources you need to collect the data?
  – What file formats will be used? Is there special software needed to read the data?
  – What directory and file naming conventions will be used?
  – How much data are being produced, and at what growth rate? Will the data change?
Introduction to Data Management

• Think about…
  – What types of data are being produced? Do you have the IT resources you need to collect the data?
  – What file formats will be used? Is there special software needed to read the data?
  – What directory and file naming conventions will be used?
  – How much data are being produced, and at what growth rate? Will the data change?
  – How long should the data be retained?
Introduction to Data Management

• Think about…
  – What types of data are being produced? Do you have the IT resources you need to collect the data?
  – What file formats will be used? Is there special software needed to read the data?
  – What directory and file naming conventions will be used?
  – How much data are being produced, and at what growth rate? Will the data change?
  – How long should the data be retained?
  – Who controls the data?
Introduction to Data Management

• Think about…
  – What types of data are being produced? Do you have the IT resources you need to collect the data?
  – What file formats will be used? Is there special software needed to read the data?
  – What directory and file naming conventions will be used?
  – How much data are being produced, and at what growth rate? Will the data change?
  – How long should the data be retained?
  – Who controls the data?
  – Who will be responsible for data management?
Introduction to Data Management

• Think about…(continued)
  – How will you store and backup the data?
Introduction to Data Management

• Think about...(continued)
  – How will you store and backup the data?
  – Are there privacy, legal, ethical, or security requirements? Do you need to de-identify your data prior to archiving?
Introduction to Data Management

- Think about…(continued)
  - How will you store and backup the data?
  - Are there privacy, legal, ethical, or security requirements? Do you need to de-identify your data prior to archiving?
  - Can your data be shared? Why or why not?
Introduction to Data Management

• Think about…(continued)
  – How will you store and backup the data?
  – Are there privacy, legal, ethical, or security requirements? Do you need to de-identify your data prior to archiving?
  – Can your data be shared? Why or why not?
  – Are the data properly described and the overall project documented? What metadata standards will be used?
Introduction to Data Management

- Think about…(continued)
  - How will you store and backup the data?
  - Are there privacy, legal, ethical, or security requirements? Do you need to de-identify your data prior to archiving?
  - Can your data be shared? Why or why not?
  - Are the data properly described and the overall project documented? What metadata standards will be used?
  - Will you be publishing in a journal that requires your underlying data?
Introduction to Data Management

• Think about…(continued)
  – How will you store and backup the data?
  – Are there privacy, legal, ethical, or security requirements? Do you need to de-identify your data prior to archiving?
  – Can your data be shared? Why or why not?
  – Are the data properly described and the overall project documented? What metadata standards will be used?
  – Will you be publishing in a journal that requires your underlying data?
  – How should your data be cited?
Introduction to Data Management

• Components of a good plan
  – Project, experiment, and data descriptions
    • What, how much, how long, how often, who is responsible?
Introduction to Data Management

• **Components of a good plan**
  – Project, experiment, and data descriptions
    • What, how much, how long, how often, who is responsible?
  – Documentation, organization, and storage
    • Metadata, file formats, file naming conventions, local storage and backup, tools and software?
Introduction to Data Management

• Components of a good plan
  – Project, experiment, and data descriptions
    • What, how much, how long, how often, who is responsible?
  – Documentation, organization, and storage
    • Metadata, file formats, file naming conventions, local storage and backup, tools and software?
  – Access, sharing, and re-use
    • Who has management/usage rights? Shared? Copyright? Patents? HIPPA/FERPA?
Introduction to Data Management

Data Handling Procedures Related to the “Data Security and Stewardship Policy”, *University Policy 97*

<table>
<thead>
<tr>
<th>Data Sensitivity Levels</th>
<th>Examples</th>
<th>Scope</th>
<th>Protection</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low (Green)</td>
<td>Publicly available data - calendar of events or schedules, sports roster, public articles, phone directories</td>
<td>Intended to be shared with the general public by the official university data owners.</td>
<td>Does not need to be protected. End users may share this information, but may not publish another version of it.</td>
</tr>
<tr>
<td>Guarded (Blue)</td>
<td>Not sensitive – internal policies or procedures, org charts, first name, last name, email address</td>
<td>Access is generally limited to those whose job requires them to have access, but is not as restrictive as the higher sensitivity levels.</td>
<td>Should be protected from unauthorized access.</td>
</tr>
<tr>
<td>Elevated (Yellow)</td>
<td>Personal information, personnel data, certain intellectual property, and education records as defined by FERPA, Banner 92#</td>
<td>Can be accessed by authorized university personnel. This authorization must be granted and documented by the appropriate Data Steward or his/her designee or the appropriate compliance officer. For guidelines and best practices refer to the document Storing PII and FERPA Education Records.</td>
<td>Must be protected from unauthorized access. Education records as defined by FERPA, including FERPA directory information, cannot be shared or published outside of the university by end-users.</td>
</tr>
<tr>
<td>High (Orange)</td>
<td>Collected personal information, such as medical records (HIPAA), financial data, political records, some research data, private contributors, attorney/client data.</td>
<td>Can be accessed only by authorized university personnel. This authorization must be granted and documented by the appropriate Data Steward or his/her designee or the appropriate compliance officer.</td>
<td>Must be protected and possibly encrypted, access must be re-certified, security controls must be audited periodically.</td>
</tr>
<tr>
<td>Severe (Red)</td>
<td>SSN, credit card info, bank account info, driver's license #, passport #, health insurance account #, export controlled data, other data when designated by agreement or the University</td>
<td>Can be accessed only by authorized university personnel. This authorization must be granted and documented by the appropriate Data Steward or his/her designee or the appropriate compliance officer.</td>
<td>Can’t be placed in unmanaged or unsecured storage. Must be protected and possibly encrypted, access must be re-certified, security controls must be audited periodically.</td>
</tr>
</tbody>
</table>
Introduction to Data Management

• **Components of a good plan**
  – Project, experiment, and data descriptions
    • What, how much, how long, how often, who is responsible?
  – Documentation, organization, and storage
    • Metadata, file formats, file naming conventions, local storage and backup, tools and software?
  – Access, sharing, and re-use
    • Who has management/usage rights? Shared? Copyright? Patents? HIPPA/FERPA?

– **Archiving**
Introduction to Data Management

- Librarians at Hunter Library can help support your research data management needs, including archiving your data in the NC-DOCKS Institutional Repository.
Librarians at Hunter Library can help support your research data management needs, including archiving your data in the NC-DOCKS Institutional Repository.

Libraries have been managing information for thousands of years.
Introduction to Data Management

– Librarians at Hunter Library can help support your research data management needs, including archiving your data in the NC-DOCKS Institutional Repository.
– Libraries have been managing information for thousands of years.
– Core services provided:
  • Assistance with preparing your DMP
  • Archiving/sharing of research data through the NC-DOCKS/Odum DataVerse
Introduction to Data Management

- Writing your DMP
  - DMPTool
- Where to submit your work and data
  - NC DOCKS
  - Odum Institute Dataverse
DMPTool

- History
  - University of California Curation Center of the California Digital Library
  - First version release October 2011
  - 6,000+ DMPs from 600+ orgs
DMPTool

• History
  – University of California Curation Center of the California Digital Library
  – First version release October 2011
  – 6,000+ DMPs from 600+ orgs

• Benefits
  – 38 templates
  – Adhere to funder guidelines
  – Streamlines review
  – It’s easy!

DMPTool
Build your Data Management Plan
WCU’s Institutional Repository

- **NC DOCKS**
  - Holds the work of 8 UNC schools
  - Published work and pre-prints
  - Indexed by Google
  - Stable URL
  - Stats
WCU’s Institutional Repository

- **NC DOCKS**
  - Holds the work of 8 UNC schools
  - Published work and pre-prints
  - Indexed by Google
  - Stable URL
  - Stats
Submitting articles to NCDOCKS

**STEP ONE**
Email Ken
Send a copy of your CV to Ken Rogers at krogers@email.wcu.edu.

**STEP TWO**
Submit pre-pubs
Ken will look up the copyright status and add those article with no restrictions to the institutional repository. He'll then send you a list of what journals allow pre-publication versions of the articles. You can send these as email attachments.

**STEP THREE**
Bask in the glory
You're done! Your work is now permanently archived and openly accessible. Plus you can track your statistics! For your next publication, don't forget to [negotiate your copyright](#) and send us your new articles.
WCU’s Institutional Repository

- Odum Institute Dataverse
WCU’s Institutional Repository

- **Odum Institute Dataverse**
  - Gives data producers the assurance that their data and associated materials will be stored in a reliable manner and can be reused
WCU’s Institutional Repository

- **Odum Institute Dataverse**
  - Gives data producers the assurance that their data and associated materials will be stored in a reliable manner and can be reused
  - Provides funding bodies with the confidence that data will remain available for reuse and their investments will not be lost
WCU’s Institutional Repository

- **Odum Institute Dataverse**
  - Gives data producers the assurance that their data and associated materials will be stored in a reliable manner and can be reused
  - Provides funding bodies with the confidence that data will remain available for reuse and their investments will not be lost
  - Enables data consumers to assess repositories where data are held
WCU’s Institutional Repository

- **Odum Institute Dataverse**
  - Gives data producers the assurance that their data and associated materials will be stored in a reliable manner and can be reused.
  - Provides funding bodies with the confidence that data will remain available for reuse and their investments will not be lost.
  - Enables data consumers to assess repositories where data are held.
  - Supports data repositories in the efficient archiving and distribution of data.

Sample
Q&A

Mark Stoffan mstoffan@wcu.edu | Liz Skene emskene@wcu.edu

dmptool.org
researchguides.wcu.edu/datamanagementplans