Innovative Seed Grant
Data Management Plans

February 17, 2015
Presenters

Andrew Johnson  
Research Data Librarian  
University Libraries  
andrew.m.johnson@colorado.edu

Shelley Knuth  
Senior Research Data Specialist  
Research Computing  
shelley.knuth@colorado.edu
Outline

- Introduction and definitions
- Creating a DMP
- Basic elements of a DMP
- Additional DMP resources
- Q&A
What is a DMP?

“A data management plan is a formal document that outlines what you will do with your data during and after you complete your research.”

University of Virginia Research Data Services, http://data.library.virginia.edu/data-management/whymanage/
What is research data?

CU-Boulder Research Data Advisory Committee defines research data as “digital outputs, which include content in structured forms including but not limited to: text files, word processing documents, spreadsheets, websites, calibration information, or simulation outputs; digital information artifacts, such as images, vector-based map products, audio and video products; digital outputs may also include the code used to decode those products, the metadata describing such information artifacts, and required source code that runs computer instructions.”

“Defining Research Data”, https://data.colorado.edu/node/27
Why do we require DMPs?

- Aligns with policies of research funders (https://dmptool.org/guidance)
- Helps to maximize benefits of investment in research
- Improves reproducibility of research
- Enhances reputation of researcher (Piwowar & Vision, 2013)
How do you get started creating a DMP?

Try the DMPTool: [http://dmptool.org](http://dmptool.org)

- Tool for creating custom data management plans to meet specific funder requirements
- No need to register to use:
  - Click “Get Started”
  - Select “University of Colorado Boulder”
  - Log in with IdentiKey
What are the basic elements of a DMP?

Basic data management plans include answers to the following questions:

- What types of data will you produce?
- How will you describe your data and what (if any) standards will you use to do so?
- When and how will you share data?
- What can people do with your data?
- How will you archive and preserve data?
What types of data will you produce?

- List types, formats, amounts of data that your research will generate
- Examples of types: text, images, audio, video, models, simulations, spreadsheets, software, code, etc.
- Examples of formats: TXT, TIFF, MPEG-4, HDF5, CSV, Python, etc.
- Amounts should be estimates of overall total data size (e.g., in GB) and rate of growth
How will you describe your data?

- Documentation (metadata) should include at minimum:
  - Title
  - Creator(s)
  - Date(s)
  - Version
  - Identifiers (for published data)
  - Location (online and/or physical)
  - Archive/Distributor (for archived data)
How will you describe your data? (cont.)

● Documentation (metadata) can also include:
  o Discipline-specific information
  o Instrument settings
  o Image/video/audio specifications
  o Provenance information
  o Spatial information
  o Environmental conditions
  o Anything else needed to understand your data…

● Whenever possible use metadata standards appropriate to your discipline:
  
  http://www.dcc.ac.uk/resources/metadata-standards
When and how will you share your data?

- Describe when you will share data:
  - As it is collected/created?
  - Upon publication of related articles?
  - After a (justifiable) embargo period?
  - Not at all (due to privacy/legal or other restrictions)

- Describe how you will share data:
  - Via a disciplinary data repository/archive: [http://www.re3data.org/](http://www.re3data.org/)
  - Via an institutional repository (e.g., [CU Scholar](https://scholar.colorado.edu), next gen. [PetaLibrary](https://petalibrary.org))
  - Via a general repository (e.g., [figshare](https://figshare.com), [Dash](https://dash.org), [Zenodo](https://zenodo.org))

- Describe how (if possible) restricted data will be shared.
What can people do with your data?

- Describe what (if any) conditions you will place on reuse of your data:
  - Attribution?
  - Non-commercial?
  - Must consult with you?
How will you archive and preserve your data?

● Describe the steps you will take to ensure that your data can be preserved:
  ○ Use or convert data and metadata to preservation-ready formats: [https://dmptool.org/dm_guidance#formats](https://dmptool.org/dm_guidance#formats)

● Describe where and for how long your data will be archived:
  ○ Long-term is better, but some data might not be useful after a certain period
  ○ Data archive can be the same as the mechanism used for sharing data (but not necessarily)
Where can you get additional help?

Research Data Services
http://data.colorado.edu
data-help@colorado.edu

Help with:
• Drafting or reviewing Data Management Plans
• Identifying appropriate places to store, archive, and provide access to your data
• Any other research data questions you have
Thank you! Questions?

- Copyright 2015 by Andrew Johnson and Shelley Knuth
- This work is licensed under a [Creative Commons Attribution 3.0 Unported License](https://creativecommons.org/licenses/by/3.0/).

Twitter: @cu_data
Subscribe to our data events email list: https://lists.colorado.edu/sympa/info/data_community