Research Data Management at UNM - An Overview and Exchange

Karl Benedict (kbene@unm.edu), Jon Wheeler (jwheel01@unm.edu) · February 20, 2018

UNM University Libraries

Outline

- Overview of Data Management Principles
- Data Management Planning
- Data Management Resources

UNM University Libraries

Data Management Planning - First Principles

Karl Benedict (kbene@unm.edu), Jon Wheeler (jwheel01@unm.edu) · February 20, 2018

UNM University Libraries

Context - Data Management Requirements

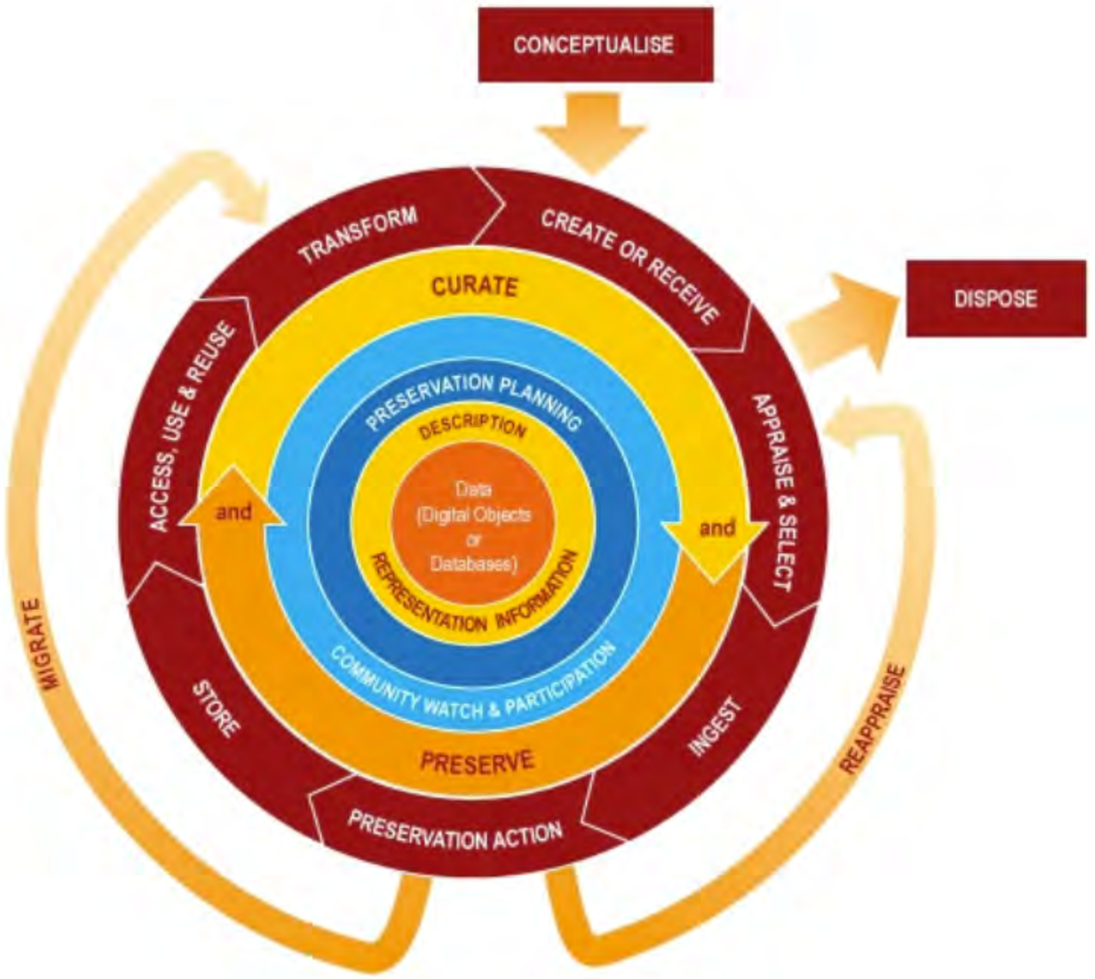
- Data Management Plan requirements from funding agencies
- Data sharing requirements from funding agencies and publishers
- IRB Protocol requirements for explicitly defining how collected data will be handled, de-identified, shared and/or destroyed, expected risks
- Increased collaborative research requiring efficient sharing of data within research teams
- Data intensive research magnifies inefficiencies in data management

Relationship Between the Research and Data Lifecycles



JISC Research Lifecycle

http://www.jisc.ac.uk/whatwedo/campaigns/res3/jischelp.aspx



DCC Data Curation Lifecycle

http://www.dcc.ac.uk/resources/curation-lifecycle-model

UNM University Libraries

Some Definitions

UNM University Libraries

Data

*For the purposes of this document, data are **any and all** complex data entities from observations, experiments, simulations, models, and higher order assemblies, **along with the associated documentation** needed to describe and interpret the data. (emphasis added)*

- National Science Foundation (2007). Cyberinfrastructure Vision for 21st Century Discovery. National Science Foundation, Cyberinfrastructure Council. Washington, DC. http://www.nsf.gov/pubs/2007/nsf0728/nsf0728.pdf. pg. 22

Documentation (AKA Metadata)

Metadata are a subset of data, and are data about data. Metadata summarize data ***content, context, structure, interrelationships, and provenance (information on history and origins)****. They add relevance and purpose to data, and enable the identification of similar data in different data collections. (emphasis added)*

- National Science Foundation (2007). Cyberinfrastructure Vision for 21st Century Discovery. National Science Foundation, Cyberinfrastructure Council. Washington, DC. http://www.nsf.gov/pubs/2007/nsf0728/nsf0728.pdf. pg. 22

Embargo

*A **period during which access to research data is not allowed** to certain types of users. This is either to protect the revenue of the publisher or (more generally) to protect the interests of other parties (for example, partner research organizations). (emphasis added)*

- University of Bristol: Research Data Management Glossary

License

*A licence in this context is a legal instrument for a rights holder to **permit a second party to do things that would otherwise infringe** on the rights held.The first thing to note is that only the rights holder (or someone with a right or licence to act on their behalf) can grant a licence; it is therefore imperative that the intellectual property rights (IPR) pertaining to the data are established before any licensing takes place. (emphasis added)*

- How to License Research Data. Digital Curation Centre. http://www.dcc.ac.uk/resources/how-guides/license-research-data#x1-20002

Some Recommendations

UNM University Libraries



UNM University Libraries

What you need to know …

- Who?
- What?
- Where?
- When?
- Why?
- How?
- Access?



UNM University Libraries

Huh?

Who? Credit (researchers, sponsors), Questions, Responsibility, Role

What? What was measured, Units, Aggregation

Where? Geographic Location (define datum, Coordinate System, Method)

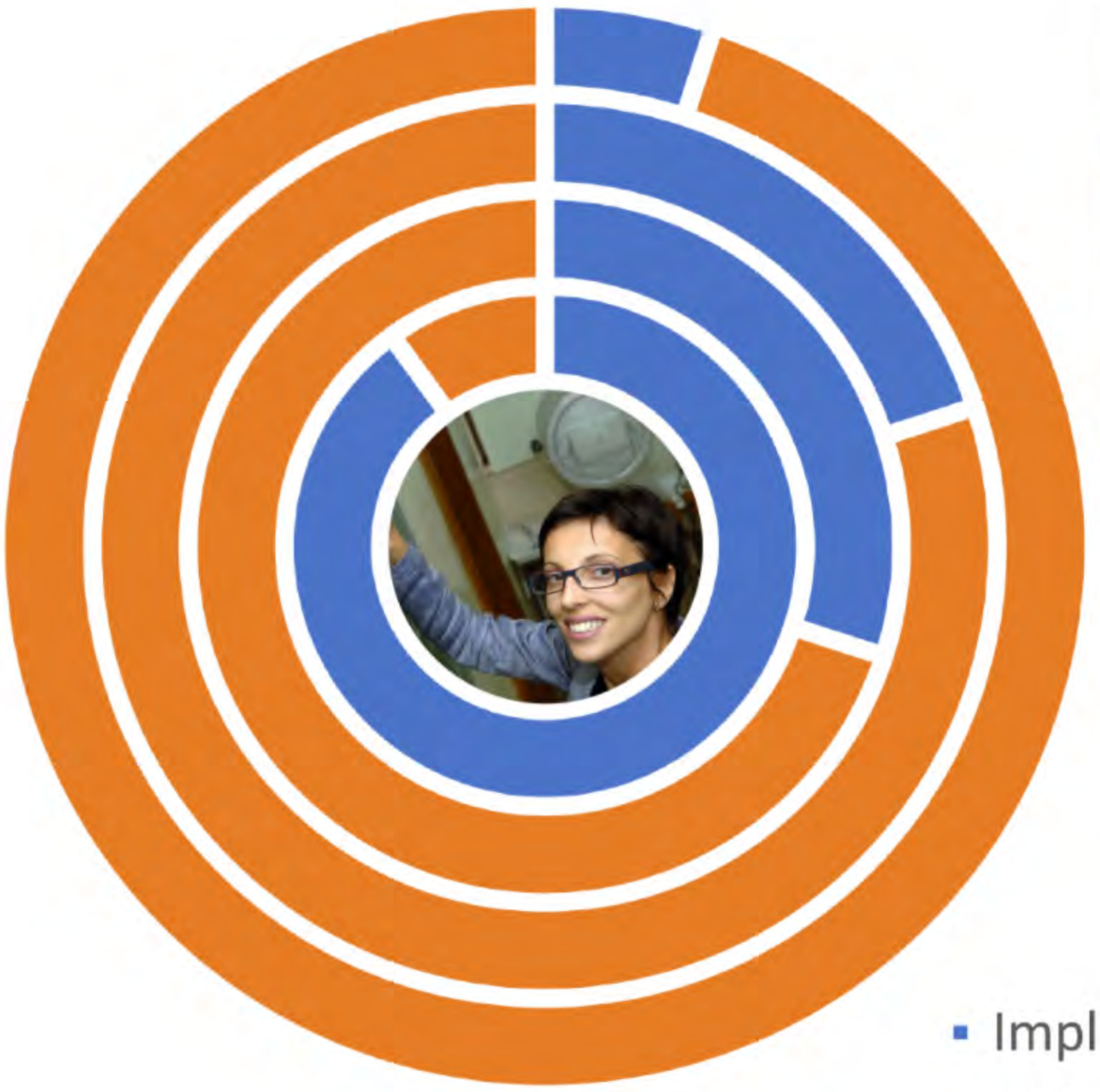
When? Date, Time - structured, Consistent, Time Zone, Standards-based

Why? Purpose for Data collection, Suggested Use, Known Limitations

How? Instruments, Sensors, Algorithms, Models, Software, Methods

Access? Licensing Terms, Embargo, Redistribution, Modification, Privacy Concerns

Who Needs What Information?







- Implicit Knowledge
- Written Documentation

UNM University Libraries

Organization



Define folder and file names and structure - and use it

Use meaningful names that include basic information (e.g. date, measurement, collection, etc.)

Unique

Avoid Spaces

ASCII Characters only

Security of Files & Backups

Backups

UNM University Libraries

Structure / Content

Consistent content

Separate data from analysis

Focus on tabular structure for tabular data

Explicitly encode missing data, and document that encoding

Use meaningful column headings - while keeping short without spaces

Include units

Data dictionary



UNM University Libraries

Formats


Bruegel's Tower of Babel

Plan for data & metadata integration into an archive

Open Standards
>>>
Proprietary ASCII
>
Proprietary Binary - Documented
>
Proprietary Binary

Documentation

Many documentation standards

Machine and human readable

Commonly based on Extensible Markup Language (XML)

Wide variety of strategies/methods/tools for creating documentation

Enables Discovery, Use, and Understanding

Work with experts in documentation for your discipline to identify relevant standards for your data



Rosetta Stone - Hans Hillewaert (http://tinyurl.com/bsdppy8) CC-by-sa 3.0

UNM University Libraries

Data Curation - Selecting and Transition to Preservation System

Identify what data need to be shared & preserved

- Data directly related to publications
- Raw data - especially when they cannot be replicated or replaced
- Derived data that have a high cost of production
- Retention required by protocol or other legal requirements

Documentation

- Enables discovery, understanding and use/reuse
- Supports long-term preservation: format information, dependencies
- Citation information

Who is responsible for providing continuity of access?

Bottom Line

Get help … Consult early and often (rds@unm.edu) - your target archive

Maintain documentation from the project planning stage and throughout your work

Adopt a systematic model for organizing your data: naming, file structure, formats, storage, backups

Adopt consistent and documented data structures

Separate data from analysis products

Always have the entire data lifecycle in mind when you are managing your data

Resources

PLOSONE Data Sharing Requirements:
http://www.plosone.org/static/policies#sharing

Library of Congress: Sustainability of Digital Formats:
http://www.digitalpreservation.gov/formats/index.shtml

Digital Curation Centre: Disciplinary Metadata:
http://www.dcc.ac.uk/resources/metadata-standards

Creative Commons Licensing Information: https://creativecommons.org/

Karl Benedict - kbene@unm.edu / Research Data Services - rds@unm.edu

UNM University Libraries

Acknowledgements

Unless otherwise noted, all images are from shutterstock.com or iStockPhoto

The products, platforms and development work upon which this work is based have been funded by

- NSF EPSCoR Program (Track 1 [Awards: 0447691, 0814449, 1301346] and Track 2 awards [0918635, 1329470])
- New Mexico Resource Geographic Information System
- NASA ACCESS Program
- UNM's College of University Libraries and Learning Sciences

Dilbert cartoon used with permission:
http://universaluclick.com/licensing_permissions/educational_use

UNM University Libraries

Data Management Planning

Karl Benedict (kbene@unm.edu), Jon Wheeler (jwheel01@unm.edu) · February 20, 2018

UNM University Libraries

Overview

- Reduce, reuse, recycle?
- Diverse DMP requirements
- Talking through the sections of a DMP

Reduce, reuse, recycle?

Just as each research project is unique - the plan for managing the data for that project is unique.

A common set of concepts apply, but there is no “one size fits all”.

Policies, requirements and reviewer expectations change with time.

Diverse DMP requirements

University Libraries Research Data Services Research Guide:
http://libguides.unm.edu/data

DMP Tool Online: https://dmptool.org/

UNM | University Libraries

Talking through a DMP

- Roles and Responsibilities
- Expected data
- Period of Data Retention
- Data Format and Dissemination
- Data Storage and Preservation of Access
- Additional Possible Data Management requirements

Data Management Resources

Karl Benedict (kbene@unm.edu), Jon Wheeler (jwheel01@unm.edu) · February 20, 2018

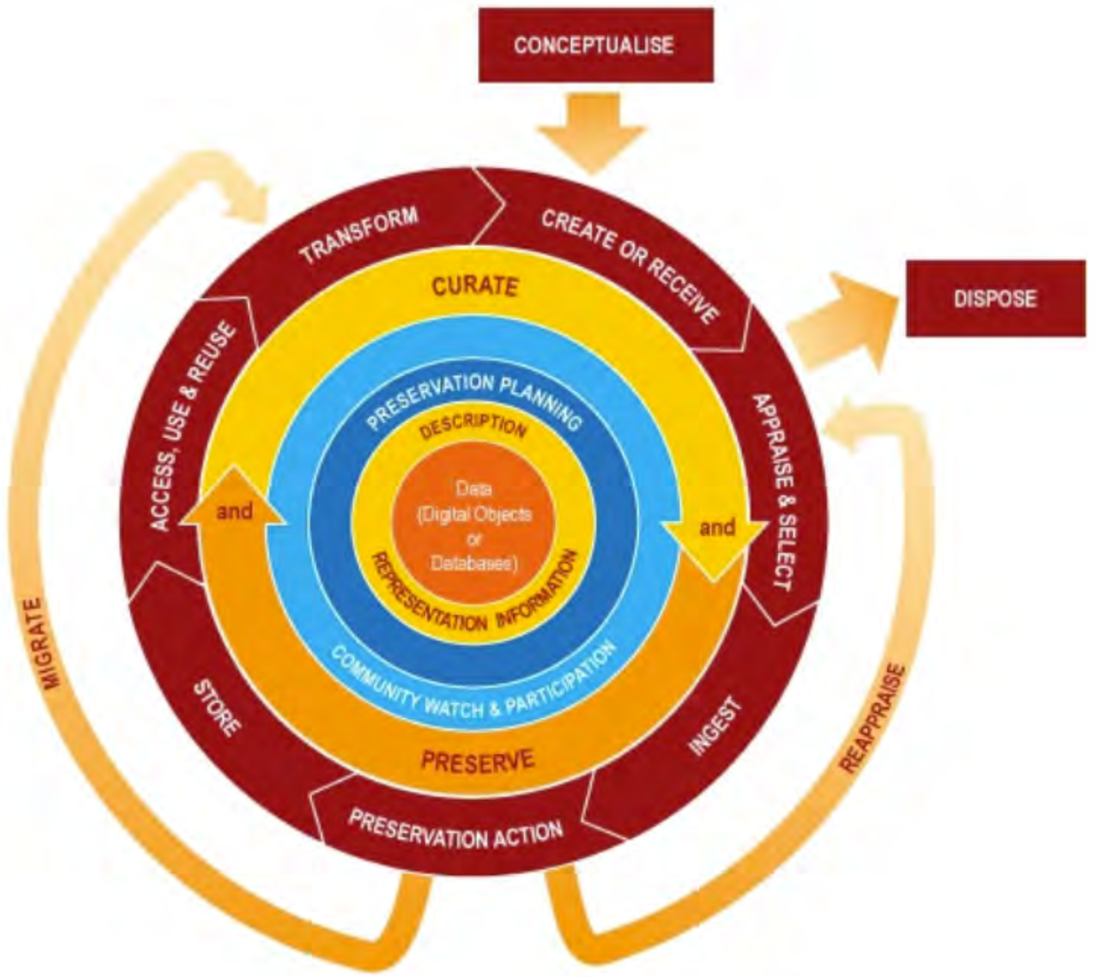
UNM University Libraries

Relationship Between the Research and Data Lifecycles



JISC Research Lifecycle

http://www.jisc.ac.uk/whatwedo/campaigns/res3/jischelp.aspx

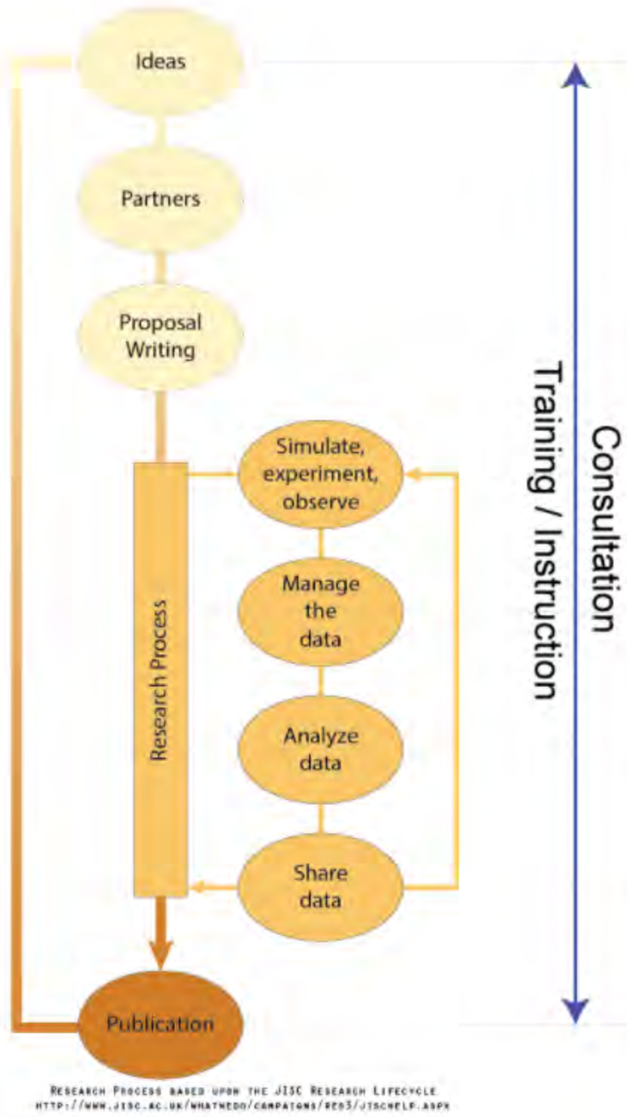


DCC Data Curation Lifecycle

http://www.dcc.ac.uk/resources/curation-lifecycle-model

UNM University Libraries

A Mapping of Resources to Research Lifecycle Phases



Training / Instruction

Consultation

Reference Services		
Reference Services		ORCID, OSF
Reference Services *Data Management Planning*	*FRDO, OIRB, ORC, ISD, ECO*	ORCID, DMPTool
CSEL DEN *GIS Workstations* *LoboGit*	*Center for Advanced Research Computing*	aws
LoboDrive	*Center for Advanced Research Computing* *OneDrive, Sharepoint*	aws, OSF
CSEL DEN *GIS Workstations* *LoboGit*	*Center for Advanced Research Computing*	aws
LoboDrive	*OneDrive, Sharepoint*	aws, OSF
UNM Digital Repository *Library Preservation Platform* *Digital Preservation Network*		ORCID, OSF *Multiple Repositories*

UNM University Libraries