

Connecting Education Students to Research Data Management

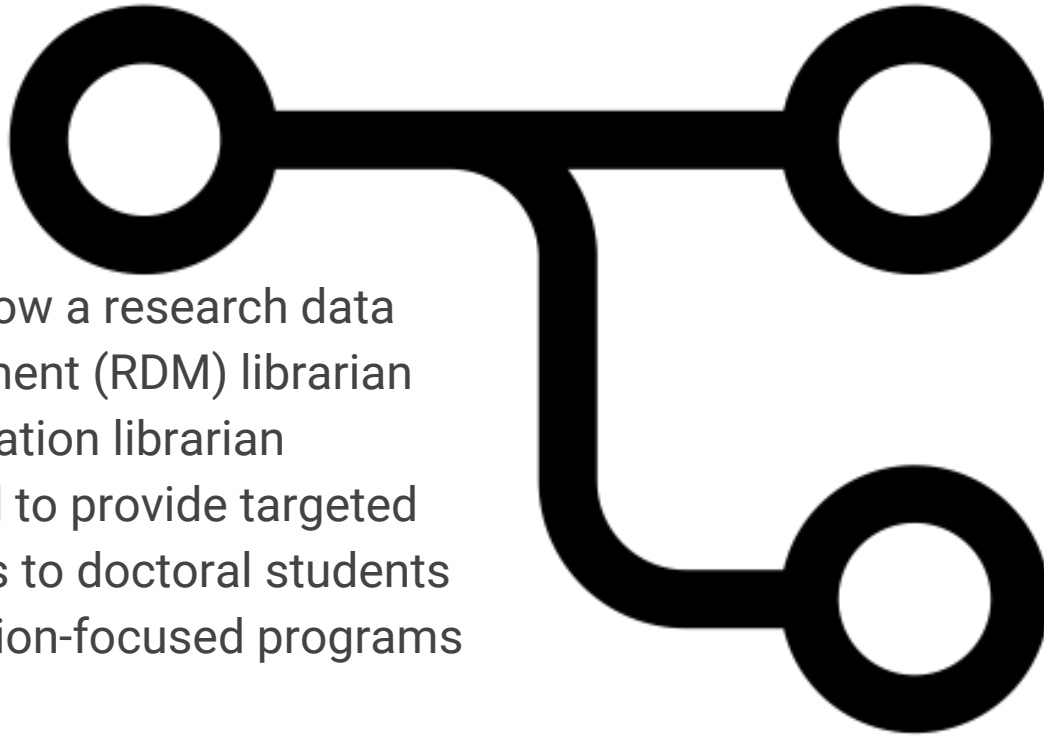
Creating intra-library partnerships and expanding RDM instruction to the social sciences

MI-ALA Conference 2017

Joanna Thielen, jthielen@oakland.edu
Amanda Nichols Hess, nichols@oakland.edu
Oakland University Libraries



Session Goals



Explain how a research data management (RDM) librarian and education librarian partnered to provide targeted resources to doctoral students in education-focused programs

Highlight the value of RDM practices to students and faculty in non-STEM disciplines

Identify ways others can develop partnerships around RDM or provide RDM instruction specific groups



Data is...

materials generated or collected during the course of conducting research.

- National Endowment for the Humanities Office of Digital Humanities

recorded factual material commonly accepted in the scientific community as necessary to document and support research findings

- National Institutes of Health

Data is...

Anything you perform analysis on.

- *Data Management for Researchers*
by Kristin Briney

A few types of data

- research notes or lab notebooks
- survey responses
- software and code
- measurements from laboratory or field equipment
- images (such as photographs, films, or scans)
- audio or video recordings
- physical samples

Research data management (RDM) is...

the compilation of small practices that make your data easier to find, easier to understand, less likely to be lost, and more likely to be usable during a project or ten years later.



RDM includes...

- Data management planning
- Documenting your data
- Creating metadata about your data
- Organizing your data
- Improving analysis procedures
- Securing confidential data properly
- Having adequate storage and backups
- Taking care of your data after a project
- Sharing data effectively
- Finding data for reuse in a new project




Why is RDM important?

- Save time and resources
- Preserve data
- Maintain data integrity
- Meet grant requirements
- Promote new discoveries
- Support open access & open data initiatives

**Most importantly:
researchers don't want to
lose their research data!**

RDM in academic libraries

- Approx. 2005: Academic libraries started developing data-related services
 - Response to forthcoming federal funding agency requirements
 - Data librarians provide assistance with:
 - Writing data management plans
 - Creating documentation & organization systems
 - Ensuring data is stored safely and securely
 - etc.
 - Training and outreach are important components
 - Prior focus on STEM disciplines
- 



Project Context & Development

Oakland University (OU), Rochester, MI

Doctoral research institution

Fall 2016 enrollment:

- 20,012 total students
- 3,444 graduate students

Masters / Ph.D programs in:

- Humanities
- Social sciences
- Education
- STEM
- Health sciences/nursing

OU Libraries

- 13 full-time faculty librarians
- Have both liaison responsibilities and library-specific specializations



Photo courtesy of Oakland University.

Data Management at OU Libraries

- Supports 2nd institutional goal in OU's 2025 Strategic Plan
 - "Be recognized as a strong research and scholarly environment focused on creative endeavors and on the discovery, dissemination, and utilization of knowledge."
- Research Data Support webpages
 - General RDM guidance
- Workshops on various RDM topics

The screenshot displays the 'Research Data Support' webpage. On the left is a navigation menu with sections: 'Research Help', 'Services' (highlighted), 'Services for YOU', 'Accessing Materials', 'Computers & Technology', 'Library Spaces', 'Instruction Support', 'Research Data Support' (highlighted), 'Scholarly Communication', and 'About'. Below the menu is a notice: 'OU Libraries are open 24 hours - OU ID needed midnight-7am. Hours | Location' with social media icons for Facebook, Twitter, Instagram, and RSS.

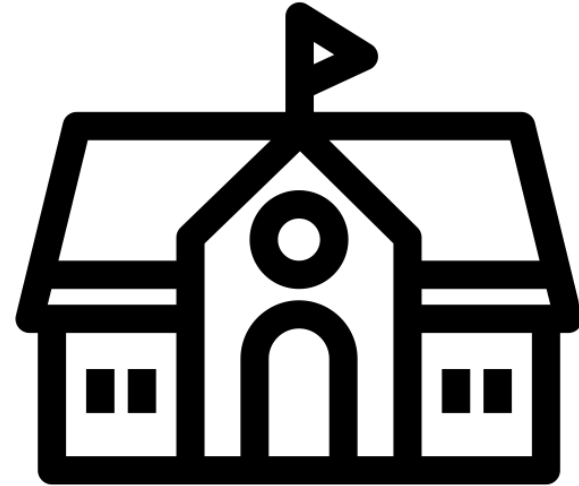
The main content area is titled 'Research Data Support' and includes the following sections:

- Create a file organization system**
 - Data organization provides a structure to your research data. **Using a file naming convention (FNC) is a simple way to organize your files.** It gives each file a unique name that describes both its contents and its relation to other files. The convention can be used for both physical and digital files.
- Examples of file naming conventions**
 - A diagram shows the file name `"yymmdd-FPACEN-FIN-BudgetProposal-v1-00.xls"` broken down into segments: A (Reverse Creation Date), B (Project Name), C (Project Dept/ Section), D (Doc Name), E (Version Number), F (Revision Number), and G (Doc File Extension).
 - A second diagram shows a template: `[Author]_[YYYY]_[JournalName]_[Topic].xxx` with labels: '1st author's last name' pointing to [Author], 'Abbreviated' pointing to [YYYY], 'Abbreviated' pointing to [JournalName], and 'File format' pointing to [Topic].xxx.
- Make a plan for your data**
 - How to write a data management plan (DMP)
 - Funding agency requirements
 - IRB considerations
- Find and re-use existing data**
 - Common data sources
 - Re-using data ethically
 - Citing data sets
- Organize and store your data**
 - Create documentation about your data
 - **Create a file organization system**
 - Confidential data
 - Storing & backing up your data
- Share your data**
 - Why share data?
 - Preserve your data

OU's School of Education & Human Services (SEHS)

Five academic departments:

- Counseling
- Human Development & Child Studies (Early Childhood / Special Education)
- Organizational Leadership (Educational Leadership / Human Resource Development)
- Reading & Language Arts
- Teacher Development & Educational Studies



School by Chanut is Industries from the Noun Project

Total enrollment, Fall 2016: **2,094 students**

Includes undergraduate, graduate, and professional development programs

OU's Graduate Education in SEHS

Fall 2016: 1,031 graduate students in all degrees / certificates; Ph.Ds awarded in:

- Counseling (F16: 20 enrolled)
- Early Childhood (F16: 32 enrolled)
- Educational Leadership (F16: 57 enrolled)
- Reading & Language Arts (F16: 21 enrolled)

Most SEHS graduate students work full-/part-time, have family obligations, and are returning to school after some period of absence.

Average age (including Ed.S students): 42

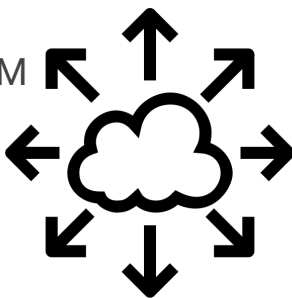


students by Wilson Joseph from the Noun Project

Library Collaboration around RDM

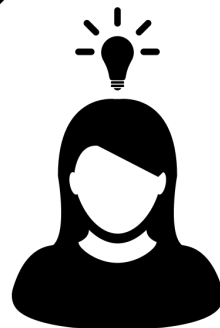
Amanda: Education Librarian **and**
Ph.D candidate in OU's SEHS

- Personal knowledge of Organizational Leadership Ph.D program
- Concrete / recognized need for RDM practices for students
- Existing relationships as starting point to pilot RDM in education graduate programs



Joanna: Research Data librarian

- Subject expertise - essential here
- Aligned RDM knowledge, practices, and experience with SEHS graduate educational experiences



**Together, this knowledge
and experience led to
concrete and useful ideas.**

Adapting RDM Best Practices for SEHS Students

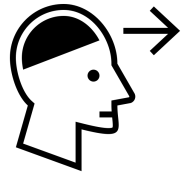
What is important to these students?



Ease of integration



Usefulness



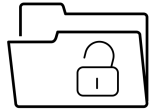
Concrete, real-world connections

Adapting RDM Best Practices for SEHS Students

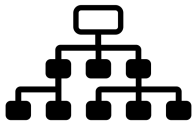
Relevant RDM practices



Security of human subjects data



Storage, documentation

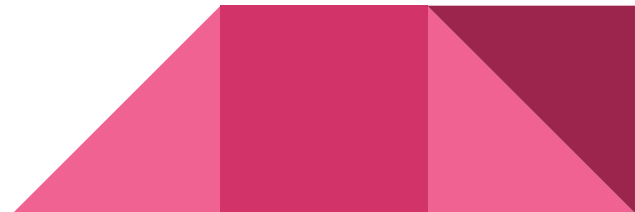


Organization

Irrelevant RDM practices



External grant requirements



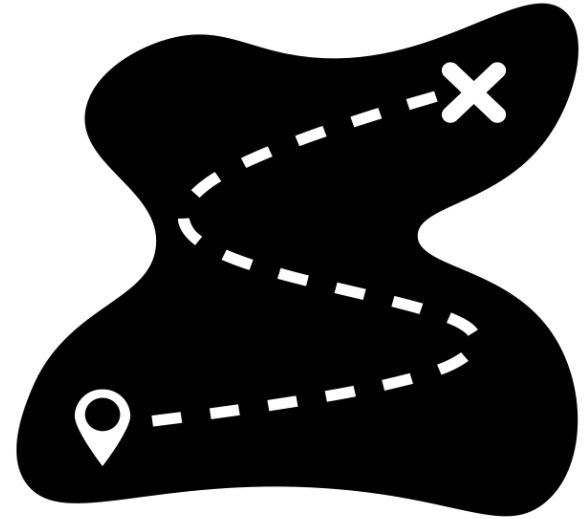


Connecting RDM with doctoral programs in education

Piloting with **One** Program

Beginning with Educational Leadership doctoral students:

- Mapping support into the curriculum to support students at intentional points in their doctoral experience
- Offering standalone workshops to students farther along in the process, but still in need of RDM guidance



First step:

- Identifying students' first experiences with research (qualitative / quantitative)
 - Introductory-level RDM support targeted to course expectations, assignments, needs, etc.

Resulting program: Data Management 101

Second step:

- Building on the introductory RDM support in the last formal class before students work on comprehensive exams
 - Offering more advanced RDM support
 - Providing time to discuss challenges, questions, or successes experienced with RDM so far

Resulting program: Advanced Data Management

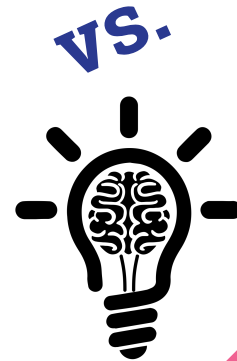


Standalone Workshops Outside of Coursework

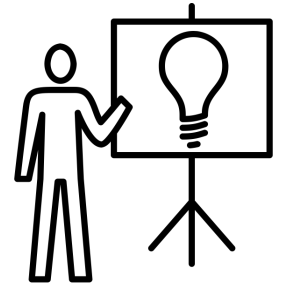
Initiative to help students who were farther along the Ph.D process:

- Monthly doctoral support group for students writing comprehensive exams / dissertation proposals / dissertations
 - Offering training following regular meeting time
 - Repeating this training before SEHS evening courses

Low attendance



vs.



Immediate
relevance to
students' work



Samples of slides from Data Management 101

Common data sources

Previous publications

- Look at article appendices and supplementary files
- Contact the corresponding author

Education sources:

- [Data.gov \(education section\)](#) - “Home of the US government’s open data”
- [National Center for Education Statistics](#)

General sources:

- [ICPSR](#) - Political and social research, 1960s to present
- [Statistical Abstracts of the US](#)* - Social, political and economic data, 1878-1969
- [Roper iPoll](#)* - US public opinion poll data
- [American FactFinder](#) - Demographic data from censuses and other surveys

*Note: Databases provided by OU Libraries. Must be current OU student, faculty or staff to access.

Documentation

- Document the who, what, when, where, why and how of your research
- Types of documentation
 - Research notes
 - Codebooks
 - ReadMe.txt files
- Tools for documentation
 - EverNote
 - Word
 - OneNote

ICPSR 4098

National Household Education Survey, 2003

Part 1: Codebook for Parent and Family Involvement in Education Survey Data

*United States Department of Education
National Center for Education Statistics*



DADEDUC is a measure of the educational attainment of the child's residential father or male guardian. This measure is derived from DADGRADE (PV7) and DADDIPL (PV8).

The values for DADEDUC are:

- 1 = Less than high school diploma
- 2 = High school graduate or equivalent
- 3 = Vocational/technical education after high school or some college
- 4 = College graduate
- 5 = Graduate or professional school
- 1 = No father for the subject child in the household

DAEMPLD indicates the employment status of the child's residential father or male guardian. This measure is derived from DADWORK (PV9), DADLEAVE (PV10), DADHOURS (PV11), DADLOOK (PV13), DADAGN (PV14a), DAEMPL (PV14b), DADREL (PV14c), and DADANSAD (PV14d).

The values for DAEMPLD are:

- 1 = Working 35 hours or more per week
- 2 = Working less than 35 hours per week
- 3 = Looking for work
- 4 = Not in the labor force
- 1 = No father for the subject child in the household

Best practices

1. Practice the 3-2-1 rule



Example storage plan:

I will keep my data on my personal laptop and back it up on my personal external hard drive and OakShare (files.oakland.edu). My laptop and external hard drive will be stored at my house whereas OakShare is stored on servers at OU.

Storage media	Recommended?	Disadvantages
Personal computer	YES	Prone to theft or loss
External hard drive	YES	Subject to degradation; lifetime is ~ 5 years
CD/DVD	YES	Subject to degradation due to mishandling; can be laborious to use
USB flash drive	NO	Easy to lose; very fallible
OakShare (files.oakland.edu)	YES	Initial storage limit is 200 MB; Doesn't sync like some cloud services
Cloud service (Google Drive, Dropbox, Box, etc.)	NO	Don't use for confidential data ; Terms of Service may give the company a license to use your files (including data files)

Main Takeaways

Recap of RDM practices

- Find & re-use data as a good data consumer
- Create documentation about your data
- Create a file naming convention
- Storage - back ups!, use OakShare for confidential data
- Ask for help when needed!

Ways to get started with RDM -- don't be overwhelmed!

1. Storage - start regular back ups!
2. File naming convention
3. Documentation
4. ...

The top right corner of the slide features a decorative graphic consisting of several overlapping triangles in various shades of blue, ranging from light to dark. The main title is centered on the page in a large, white, sans-serif font.

Assessing the Impact of Targeted RDM Instruction

Research questions

Do RDM interventions (workshops, training sessions, individual consultations, etc.) by academic librarians affect the RDM practices of doctoral students in education-focused programs?

If so, how do these interventions affect how nascent researchers collect, organize, store, and manage their research data?

This project has been reviewed by our institution's Office of Research Administration and **classified as not research** under the federal definition of research (i.e. our project didn't need to be reviewed by IRB).

Research = **systematic investigation** with **human subjects** that seeks to create **generalizable knowledge**



Check with your institution to see if this is an option!

BUT we're still following ethical practices for human subjects research (protecting PII, etc.)...

And we can still **present and publish** on this project.



Two Data Sources

Quantitative feedback form, post-interaction

Sent to students after in-class instruction, a workshop, or a 1:1 research consultation

Data Management and SEHS Graduate Students: Feedback Form

Format of your recent research data management support activity:

- Workshop presentation (i.e., a sit-and-get session led by Joanna and/or Amanda)
- Hands-on training session (i.e., time provided to work with tools and resources)
- 1:1 consultation with Joanna or Amanda

Please indicate how much you knew about the following areas BEFORE participating in the workshop, training session, or 1:1 research consultation.

	No knowledge on this topic.	Very little knowledge on this topic.	Some knowledge on this topic.	Average amount of knowledge on this topic.	Significant amount of knowledge on this topic.	Expert amount of knowledge on this topic.
What research data management means.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
How to find and reuse existing data for my own research	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
How to annotate or document my data so I can understand it later.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
How to organize my	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Interviews, baseline and intervention groups

- Baseline interviews about RDM with Ph.D graduates who did **not** receive RDM support / instruction from the library
- Interviews with intervention group after successful Ph.D defense to understand their RDM practices

Data gathered from these sources will help us to understand the effectiveness of our outreach.

Baseline Group: Common themes

n=4; one interview unusable (participant did not meet inclusion criteria)

Commonalities across participants' experiences:

- All recognized the importance of keeping confidential data protected and took procedures to de-identify data
 - IRB requirements as practical guidelines for **data security, data storage**
- All kept multiple copies of data in multiple formats
- None saw future use for their data



Baseline Group: Different themes

Differences across participants' experiences:

- Varied formats of information in:
 - Data collection - paper / electronic instruments
 - Storing / capturing information - locked cabinets, electronic folders
- Only one interviewee had a specific file naming convention



Intervention Group: Initial findings

$n=4$, including one faculty member who participated as part of her course

- Hard to find commonalities or differences in this small sample size
- Different sets of skills, knowledges, and intended research approaches color participants' experiences
- Generally, all attendees found the instructional experiences useful in some way

Going forward, we will continue to ask for feedback from both formal (i.e., class sessions) and informal (i.e., RCs) instructional interactions around RDM.





Conclusions & Key Takeaways

Lessons Learned & Next Steps

- Course-embedded: More meaningful
 - Faculty usually learn new things - useful for students to see them engaged and interested in this topic
 - Faculty can see more places to integrate the library into their courses
- Continue collecting data from baseline group
- Expand RDM instruction to all four education doctoral programs (Fall 2017)
 - Introductory-level RDM support targeted to course expectations, assignments, needs, etc.
- Follow up all interactions with survey
- Prepare for intervention group interviews - Winter 2018 at the earliest



Implementing Targeted RDM at Your Institution

Use institutional expertise!

Where are students collecting or working with data? What RDM practices would be most relevant to them?

- Not just graduate students -- Oakland University Honors College; Business Undergrads
- Not just STEM fields


Remember: data is anything you can perform analysis on!



Implementing Targeted RDM at Your Institution

Use your connections!

With whom can you partner to develop RDM outreach?

- Intra-library partnerships -- RDM librarian, or other librarians who work with students / faculty and data (e.g. Business, STEM)
 - Cross-campus partnerships -- offices of institutional research or research support
 - Department faculty in liaison areas
 - Other ideas?
- 



Questions?

Joanna Thielen, jthielen@oakland.edu

Amanda Nichols Hess, nichols@oakland.edu

Oakland University Libraries