

Data Management 101

February 2017

Joanna Thielen | Research Data Librarian | jthielen@oakland.edu
Amanda Nichols Hess | Education Librarian | nichols@oakland.edu

**Data are anything
you perform
analysis on.**

A few types of data:

- research notes
- survey responses
- images (such as photographs)
- audio or video recordings
- teaching materials - lesson plans, worksheets, etc.

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... includes data management planning, documenting your data, organizing your data, improving analysis procedures, securing sensitive data properly, having adequate storage and backups during a project, taking care of your data after a project, sharing data effectively, and finding data for reuse in a new project.

Why is RDM important?

You don't
want to lose
your data!!!

Saves
valuable time
and
resources

Avoids
duplication of
research (e.g.
re-collecting
data)

Comply with
federal
funding
agency
requirements

RDM is highly individualized

What works well for your classmate
won't necessarily work for you

Data discovery

How do I **find** and
re-use existing data
for my own research?

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](#)

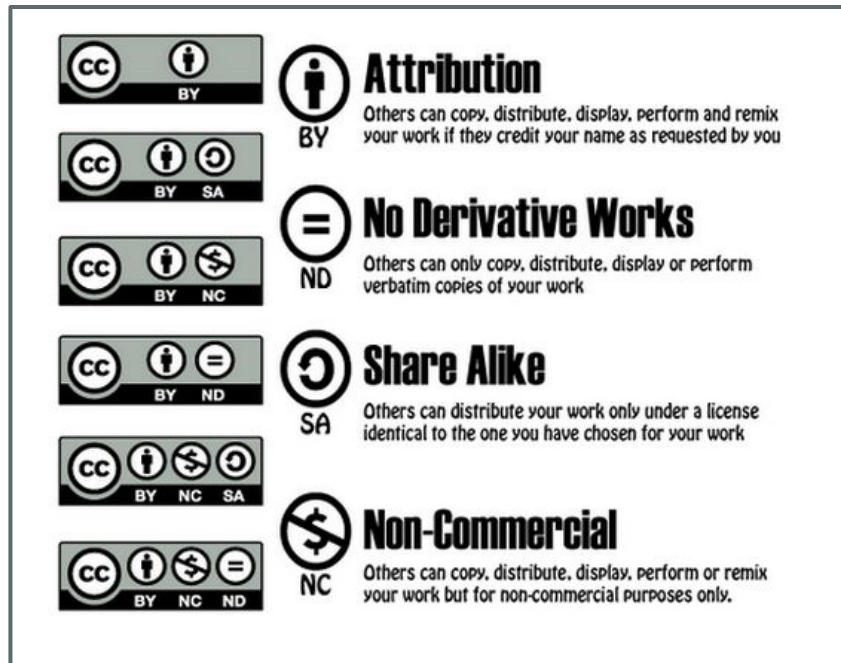
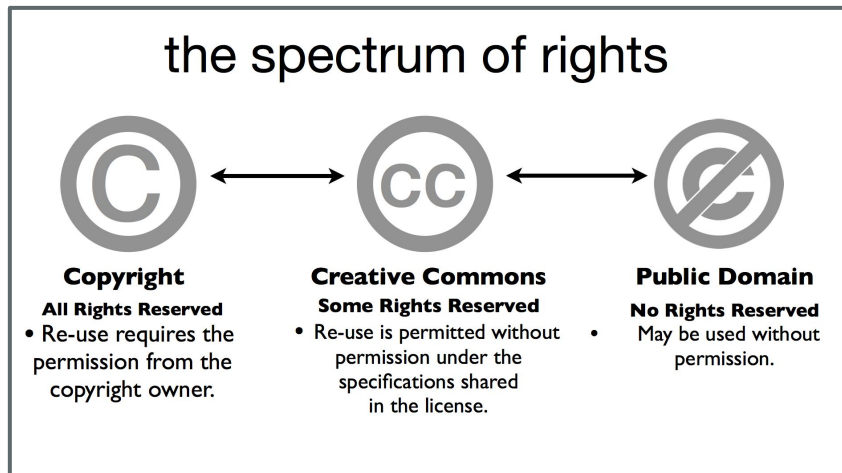
General sources:

- [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.

Being a good data consumer

1. Look at the data's license - how can (and can't) I use this data?



Being a good data consumer

2. Always cite the data you're re-using - like you would a publication

National Household Education Survey, 2003 (ICPSR 4098)

Principal Investigator(s): United States Department of Education, National Center for Education Statistics

Summary:

The National Household Education Survey (NHES) reports on the condition of education in the United States by collecting data at the household level rather than using a traditional, school-based data collection system. The surveys attempt to address many current issues in education, such as preprimary education, school safety and discipline, adult education, and activities related to citizenship. This survey included two topical survey components. The Parent and Family Involvement in Education (PFI) Survey (Part 1) addressed homeschooling... ([more info](#))

Series: [National Household Education Survey \(NHES\) Series](#)

Access Notes

- These data are available to the general public.

Dataset(s)

D50: Study-Level Files

Documentation: [Questionnaire.pdf](#) [User guide.pdf](#)

D51: Parent and Family Involvement in Education (PFI) Survey Data - [Download All Files](#) (32.2 MB)

Documentation: [Codebook.pdf](#)

Citation: U.S. Dept. of Education, National Center for Education Statistics. NATIONAL HOUSEHOLD EDUCATION SURVEY, 2003. ICPSR version. Washington, DC: U.S. Dept. of Education, Institute of Education Sciences [producer], 2004. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2004. <http://doi.org/10.3886/ICPSR04098.v1>

Data Documentation & Organization

How do I document and organize my data so that I can understand it next week or in three years?

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

Child Care & Early Education RESEARCH CONNECTIONS

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

Documentation – Amanda's example

Open Coding Categories:

- Importance of talking to others about practice
 - Sources: survey, conference observation
- Importance of actively engaging in practice
 - Sources: survey, conference observation, job postings
- Importance of reflection and feedback on one's own practices
 - Sources: survey, conference observation
- Importance of engagement -- with learners, with material, and with one's personal practices
 - Sources: survey, job postings
- Importance of technology knowledge
 - Sources: job postings, conference observation
- Importance of assessment activities
 - Sources: job postings, conference observations

A Framework rubric -- Using the ACRL framework

- How this assessment journey started
 - Instituted at one's own institution
- Started out with a pilot program using LEAP rubric
 - Artifacts came in from different sources
 - Opened up new world -- about teaching and learning
- Scoring and norming: unfamiliar process
 - Scoring -- giving an initial grade
 - Norming -- going around with colleagues to make scoring consistent
- What Katie and I did all the time in the classroom
 - Took the language in the Framework and applied it to outcomes
- Rubrics and scoring / norming operation
 - -- the Framework enabled this, too
- From pilot study -- were able to implement
 - Enabled multiple settings, opportunities
 - Moving forward -- involves teaching and learning relevant for their students in the classroom
- Adding more sessions into instruction
 - for more meaningful teaching of the Framework
 - Use of synthesis matrix -- avoid students just cherry picking
 - Also, working with faculty in the classroom
 - faculty had never assessed students in these ways
 - This leads to increased access to students, and access in different ways
- Word of mouth as the best tool in building collaboration

What experiences have most significantly shaped how you think about library / information literacy instruction?

1. My experiences in the classroom teaching. I really had no concept of my students' needs until I met with them and saw their assignments, what they responded to, etc.
2. Early on in my career, I was utterly dismayed at the lack of engagement students had with IL instruction. I made it a focus of my teaching, then, to be as engaging as possible.
3. Honestly, the most significant experience that influenced how I now teach information literacy occurred during graduate school, but not in my coursework. I was a student assistant at an academic library and had to update/modify a previously taught workshop for university faculty on using EndNote. This was my first exposure to teaching and I learned a lot from this first experience about session pacing, incorporating interactive activities, and more.
4. Interactions at the research help desk certainly influence the way that I approach my one-shot (WRT 160) instruction sessions. Students' questions at the desk help me to decide what concepts need more explanation or what skills need more practice.
5. Readings in librarianship journals, especially critical information literacy. I find that IL instruction is all too often based on techniques, not critical thinking.
6. I have found from personal experience that if the students are not able to link the instruction to an immediate task that they need to carry out, they are far less likely to pay attention and/or retain any of the instruction. Also, I have found that even when they are paying attention, they will only retain a small number of learnable "chunks". These ideas seem to be supported by various workshops and presentations that I have attended relating to library instruction. As a result, I hold the belief that instruction ideally needs to be planned in such a way that it supports an actual assignment, that the students should be ready to use the skills being taught to carry out that assignment, and that the instruction time should be spent more deeply with fewer concepts rather than touch lightly on a broad set of concepts.
7. My coworkers, the scholarship of teaching and learning, and the students with whom I work.
8. Reading and reflection on the profession for various research projects has impacted my thinking the most, paired with excellent discussions with colleagues (both casual and professional).

- Amanda Hess
Nov 11, 2015
Hands-on practice -- this is what I was expecting
- Amanda Hess
Nov 11, 2015
Intentional practices from observation and reflection
- Amanda Hess
Nov 11, 2015
Hands-on practice -- learning by DOING
- Amanda Hess
Nov 11, 2015
Different person, but same comments re: the reference desk as an important source of learning... is this in part because of the part-timers being included in this group? Or is this my own bias coming in?
- Amanda Hess
Nov 11, 2015
Hands-on practice -- learning by DOING again
- Amanda Hess
Nov 11, 2015
Again, talking about personal experience here -- what this individual has FOUND
- Amanda Hess
Nov 11, 2015
Getting confirmation for these ideas from outside sources
- Amanda Hess

Important takeaway –

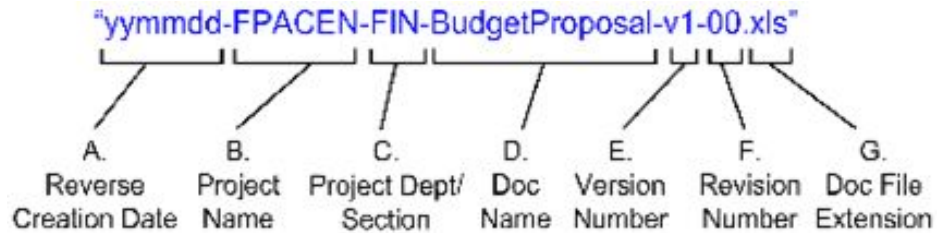
***BE CONSISTENT with your
documentation!***

**Why? Because YOU are the beneficiary of good
documentation!**

Organization - create a file naming convention (FNC)

- **FNC: standardized naming system that gives each file a unique name**
 - Name describes:
 - Contents of file
 - Relation to other files
- **FNC makes your files easy to understand and search**

Example of a FNC

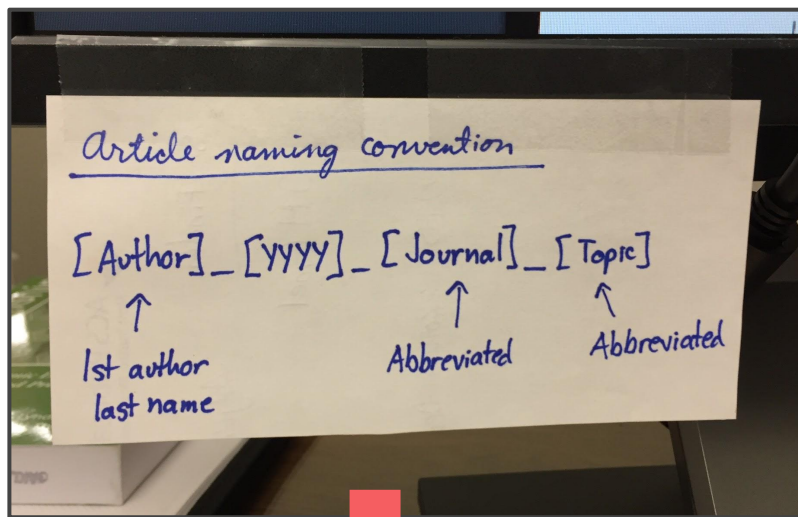


Source:

<http://vannevar.blogspot.com/2011/04/file-naming-convention.html>

Tips for creating a FNC

- **Choose 4-7 elements**
 - Examples: date (YYMMDD), creator, location, project name
- **Avoid spaces** - Some software doesn't recognize file names with spaces
- **Avoid special characters** - Such as ~ ! @ # \$ % ^ & * () { } [] ' " |
- **Strike a balance**
 - Too few elements = ambiguous
 - Too many elements = complicated
- **Be brief** - use meaningful abbreviations
- **Track versions using v01, v02, etc.**
 - DON'T use 'final,' 'initial,' 'revised,' etc.



Kennedy_2017_RevEduRes_ProfDevTeach.pdf

Data storage & security

How do I **store** my data so that...

1. It's secure (because it contains information about human subjects)?
2. It's safe?

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.

Best practices

2. Do regular back ups

- General rule: back ups should occur weekly
- Test back ups to ensure you can access backed up files

3. Don't forget about your physical data

- Examples: observation notes, interview protocols, paper survey responses
- Create digital copies (photographs, scans, etc.)

Most important
thing you will
learn today:

*Do NOT store all
of your research
data on a single
device!*

FOR MY LOST LAPTOP

I am a Rutgers Chemistry 5th year PhD student. On April 19th afternoon, my LENOVO THINKPAD T420S laptop was stolen from room 203 of Wright-Rieman building. If you stole my laptop and now you are reading this letter, I would like to say that you can keep the computer and I would like to pay you money for my data under D drive. The data is my FIVE-YEAR work. I really need the data under the D drive, there is a folder named RESEARCH, under RESEARCH folder, there is a THESIS folder. I only need that folder for my thesis defense, which is coming very soon. I would like to pay you \$1000 and use whatever way you offer to send you the money. The price is negotiable. My laptop password is 850713zd, my email address is battlecrooper@gmail.com and phone number is 732-208-6755. PLEASE contact me and I would appreciate it so so much!!!

Confidential data

According to [OU's Information Security Policy](#), confidential research data includes **information related to human subjects** (such as personally identifiable information) and needs to be stored more securely than non-confidential data



OakShare (files.oakland.edu) is the only approved storage location for confidential data (other than local storage, e.g. your laptop)

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)

How to protect your confidential data

Digital data

- Don't leave your devices unattended
- Encrypt your devices
- Password protect files
- Use a password manager (such as LastPass)

Physical data

- Store in locked building, office, filing cabinet/drawer

Platform specific encryption tools

Windows

- **Whole drive encryption:** BitLocker Control Panel
- **File/Folder encryption:** Right click on object:
Properties>General>Advanced>Compress or Encrypt>Encrypt

Mac OS

- **Whole drive encryption:** FileVault - System Preferences>Security & Privacy>FileVault
- **Folder encryption:** Disk Utility app (Mac OS native) can create a folder that presents as an encrypted hard drive (volume) allowing files to be moved into, out of and edited within the encrypted disk image.

RDM resources

- [Research data in education](#) - guide by University of Michigan Library
- [OU Libraries Research Data Management guide](#)
- Research Data Management Best Practices handouts
- Librarians!
 - Joanna Thielen | Research Data Librarian | jthielen@oakland.edu
 - Amanda Nichols Hess | Education Librarian | nichols@oakland.edu

Research Data Management

What is data? Research data management?

Data is generally defined as recorded factual material commonly accepted in the academic community as necessary to document and support research findings (Modified from the NIH's Data Sharing Policy and Implementation Guidance)

A simpler definition: **Data is anything you perform analysis on.** (Source: *Data Management for Researchers* by Kristin Briney)

Across academic disciplines, what constitutes data can vary greatly. Data can include:

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

(Source: [Research Data Management](#), University of Pittsburgh Libraries)

Research data management (RDM) is broadly defined as the compilation of small practices that make 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 can include data management planning, documenting data, organizing data, improving analysis procedures, securing sensitive data properly, having adequate storage and backups during a project, taking care of data after a project, sharing data effectively, and finding data for reuse in a new project. (Source: *Data Management for Researchers* by Kristin Briney)

[Research Data Management Best Practices](#)

Table of Contents

- [What is data? Research data management?](#)
- [Data Management Plans \(DMPs\)](#)
- [Data Storage & Security](#)
- [Data Organization & Documentation](#)
- [Data Sharing](#)
- [Data Preservation](#)
- [Confused about research data management?](#)
- [Library Contact](#)
- [Why is research data management important?](#)

Confused about research data management?

At the Kresge Library, we're here to help! We provide:

- individualized consultations on developing and implementing research data management practices.
- other methods of instruction as requested (workshops, seminars, etc.)

Please contact Joanna Thielen, Research Data Librarian, to schedule a consultation. (See contact info below.)

Library Contact



Joanna Thielen
Assistant Professor
jthielen@oakland.edu
248.370.2477

Office Hours: W 10-12 and Th 1-3
Click on the button below to set up an appointment or email me if my office

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. ...

Questions?

Joanna Thielen | Research Data Librarian | jthielen@oakland.edu
Amanda Nichols Hess | Education Librarian | nichols@oakland.edu