

Library research data management services: a perspective from the USA

Tobin Magle, PhD
University of Wisconsin-Madison

Slides: <http://bit.ly/2CJhnz2>

Outline

- What is Research Data Management (RDM)?
- Why is RDM important?
- Why do researchers need help with RDM?
- Where does the library fit in?
- What professional development opportunities are there for librarians?

Research data management is...

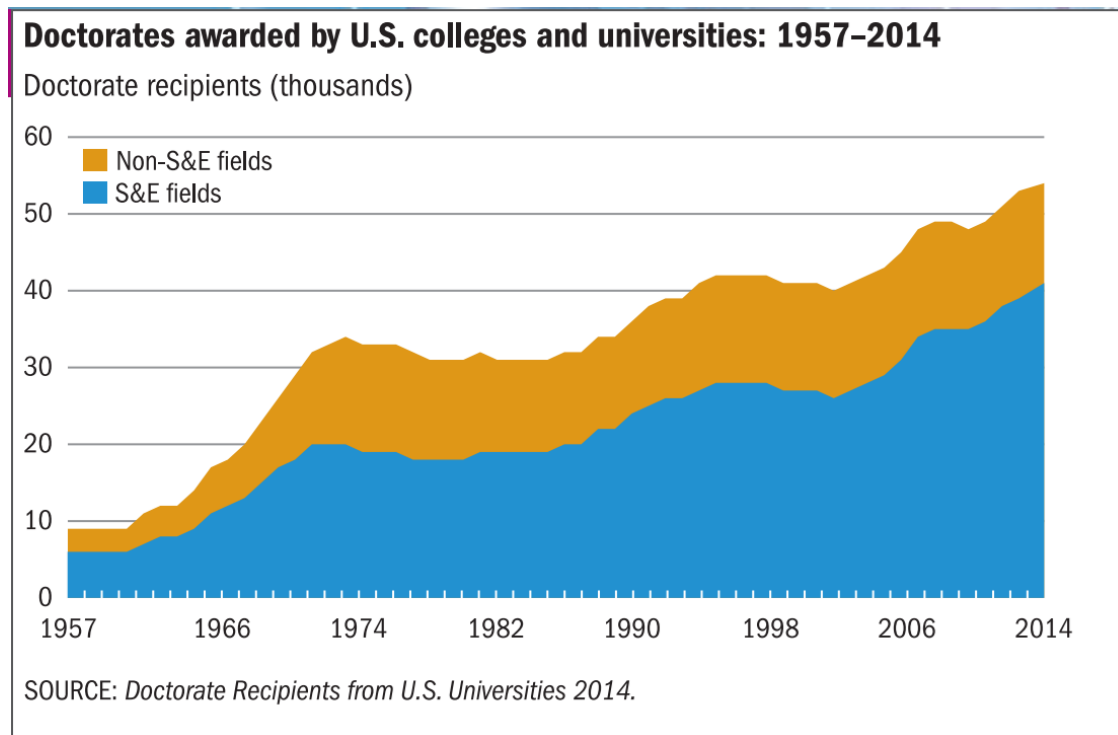
... the policies, practices and procedures needed to manage the storage, access and preservation of data produced from a research project

Why is RDM important?

More digital data

- Data are more fragile
- Facilitates sharing
- Need new skills

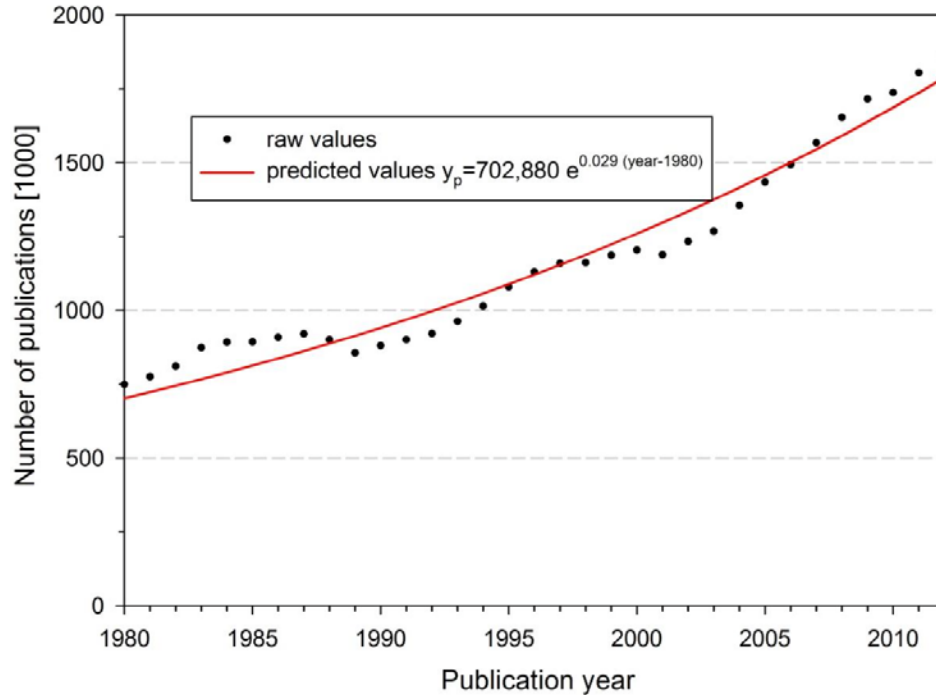
More researchers



<https://www.nsf.gov/statistics/2016/nsf16300/digest/nsf16300.pdf>

Global scientific output doubles every nine years

07 May 2014 | 16:46 BST | Posted by [Richard Van Noorden](#) | Category: [Policy](#), [Publishing](#)



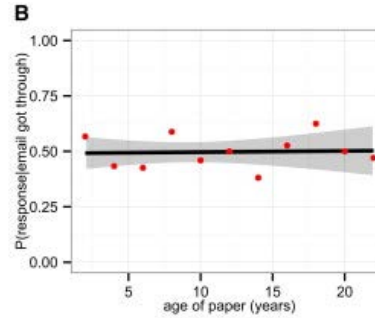
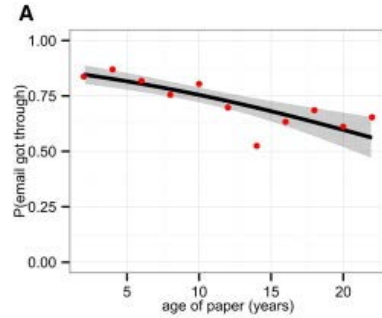
See [arXiv:1402.4578](#) for

[details](#)

Report

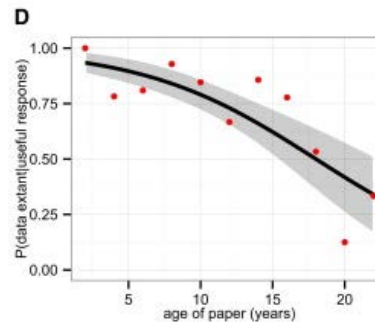
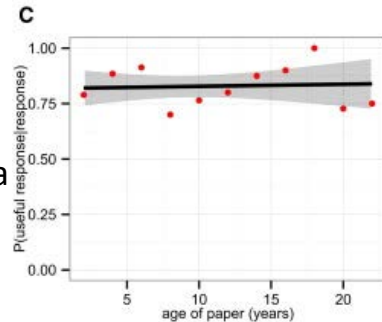
The Availability of Research Data Declines Rapidly with Article Age

Working
Email



Response
(if email
working)

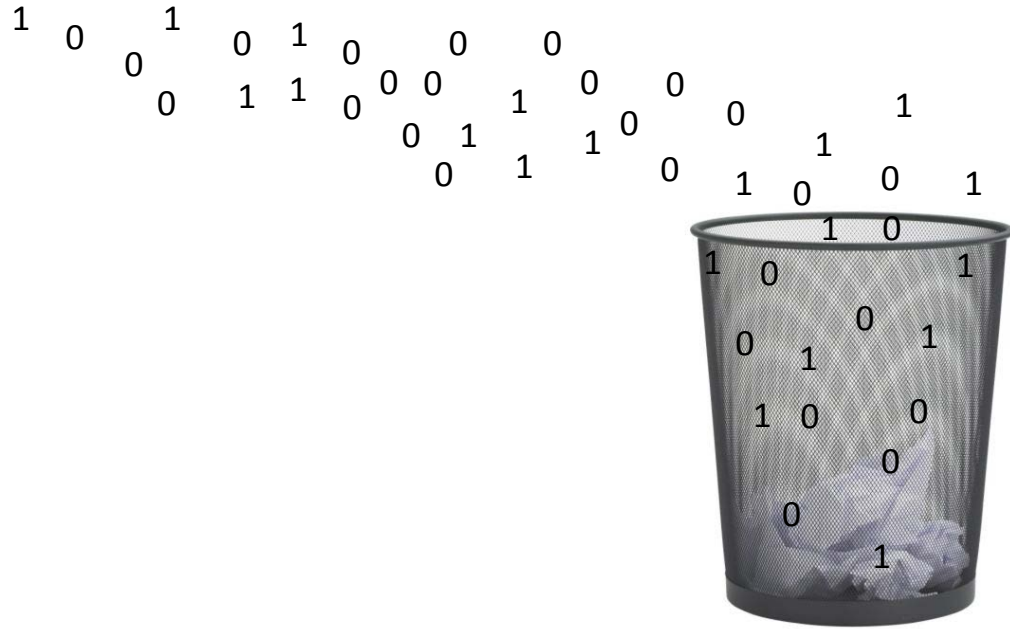
Status of data
(if response)



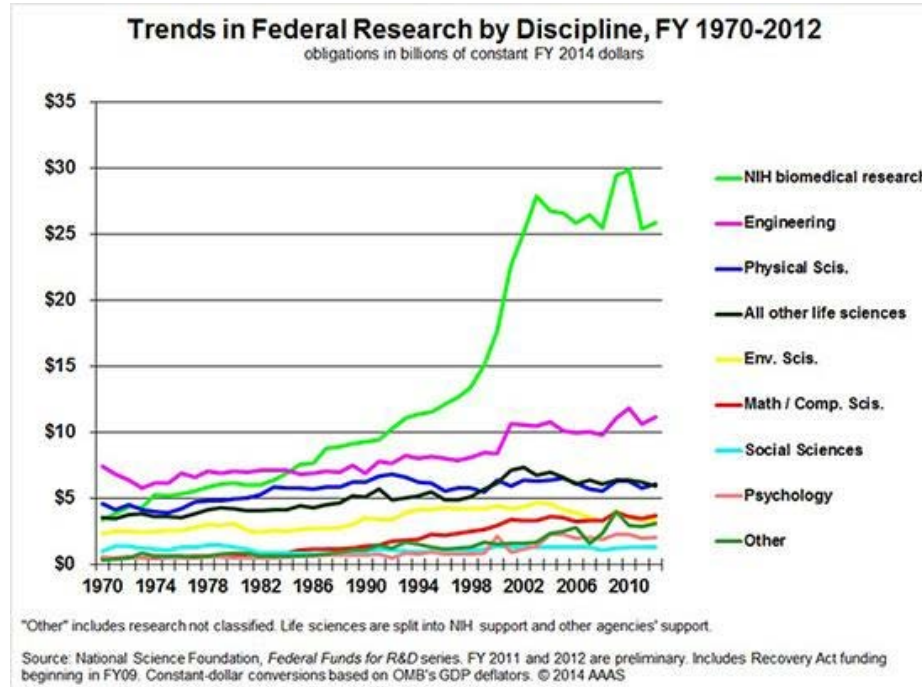
Data are extant
(If status
known)

[doi:10.1016/j.cub.2013.11.014](https://doi.org/10.1016/j.cub.2013.11.014)

We are losing vast amounts of data



Research funding is not increasing in the US



<http://www.bu.edu/research/articles/funding-for-scientific-research/>

US Funders require data sharing

- Reproducibility
 - Data reuse
 - Openness



National Science Foundation

Investigators are expected to **share with other researchers** ... the primary data... Grantees are expected to **encourage** and **facilitate** such sharing.

- NSF Dissemination and Sharing of Research Results

National Institutes of Health



As required by 45 CFR 75.322, recipients that are **institutions of higher education**, hospitals, or non-profit organizations **must release research data** first produced in a project supported in **whole or in part with Federal funds** that are cited publicly.

- [NIH Access to Research Data](#)

National Institutes of Health (Draft)



Submission of a Data Management and Sharing Plan (Plan) outlining how scientific data will **be managed and shared**, taking into account any potential restrictions or limitations ... timely **sharing of scientific data and accompanying metadata** resulting from NIH-funded or conducted research.

[DRAFT NIH Policy for Data Management and Sharing](#)



Department of Energy

Sharing and preserving data are **central to protecting the integrity of science** by facilitating validation of results and to advancing science by broadening the value of research data to disciplines other than the originating one and to society at large. To the **greatest extent**, with the **fewest constraints** possible, and consistent with the requirements and other principles stated in this document, **data sharing should make digital research data available to and useful** for the scientific community, industry, and the public.

[-DOE Policy for Digital Research Data Management](#)

Publishers require data access

- For review process
 - For all readers
 - Restrictions

Nature



Supporting **data must be made available to editors and peer reviewers...** All manuscripts reporting original research published in Nature journals **must include a data availability statement**

- [Nature Availability of Data](#)

Science Journals



After publication, **all data and materials necessary to understand, assess, and extend the conclusions** of the manuscript must be **available to any reader** of a *Science* Journal... **Unreasonable restrictions** on data, code, or material availability may **preclude publication**.

- [Science Journals Editorial Policy](#)

Public Library of Science



PLOS journals require authors to **make all data** underlying the findings described in their manuscript **fully available without restriction** at the time of publication.

- [PLOS Data Availability Statement](#)

Why do researchers need
help with RDM?

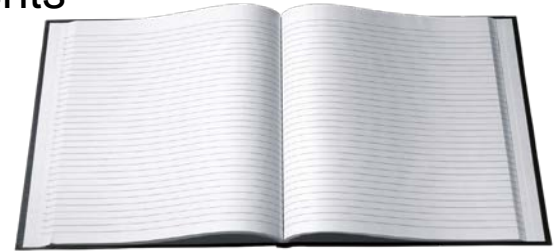
PhD training is free-form



2 years of classwork



~5 years of experiments



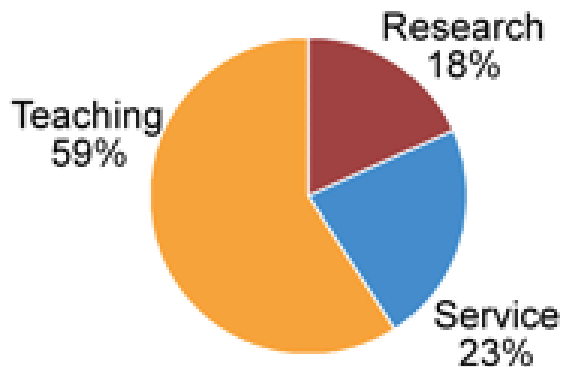
Write a dissertation



Congrats
Dr.!

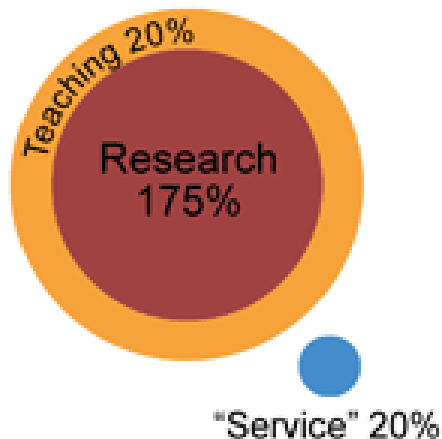
HOW PROFESSORS SPEND THEIR TIME

How they actually spend their time:



Source: Higher Education Research Institute Survey (1999)

How departments expect them to spend their time:



How Professors would *like* to spend their time:

Don't tell me what to do

WWW.PHDCOMICS.COM

How do libraries fit in?

Libraries = books



Libraries = digital information



Libraries = data?



Librarians have relevant RDM skills

- Finding records in databases
- Digital Object Identifiers
- Working with metadata
- Helping with compliance
 - Open access policies
 - Citation management
- Repositories

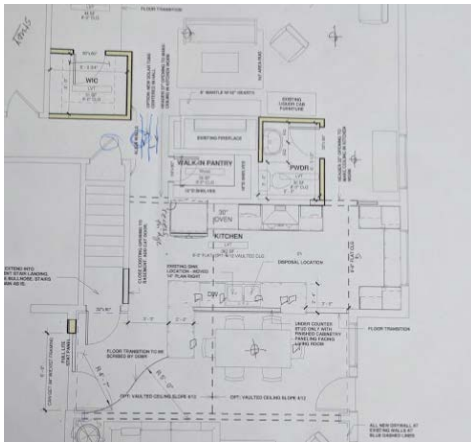


Providing research data
services in libraries

Common services

- Data management planning
- Data Organization
- Metadata standards
- Data Repositories





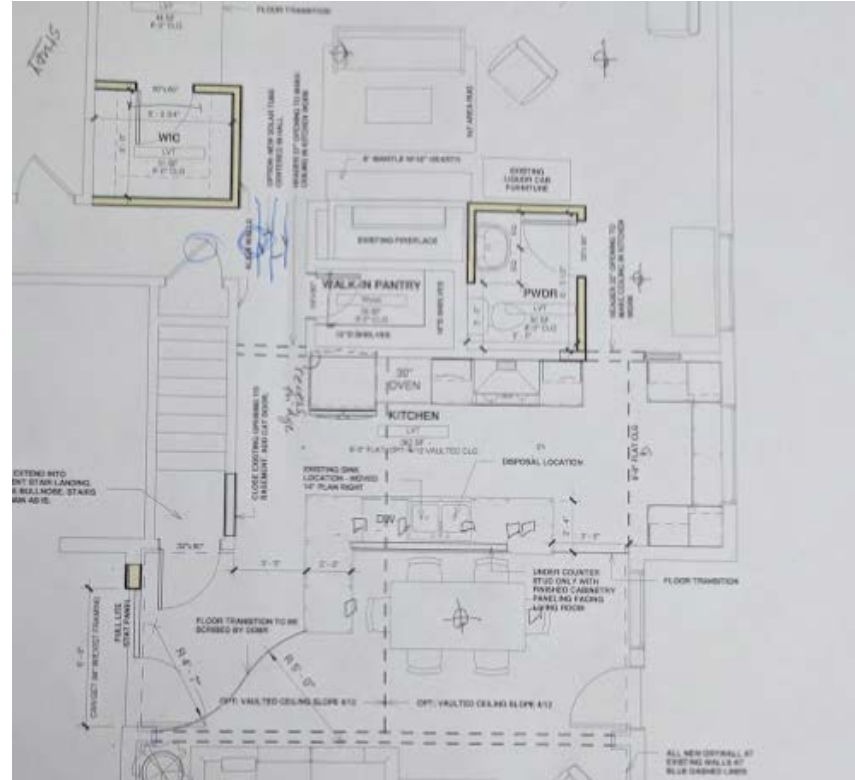
Planning

thinking through and documenting how data and other materials will be organized, saved, prepared, analyzed, and shared over the course of your research project

- What data will you produce?
- Where will you store it?
 - Size
 - Security
 - Backup
- How will you document it?
 - Use standards

Helping with Data Management Planning

- **Novice:** Help researchers find data management best practices
- **Intermediate:** Be familiar with data planning best practices
- **Advanced:** Review data management plans and provide feedback





[Crafternoon](#) by Stephen Kennedy

Data Organization

arranging your data and other research materials so they can be found – by yourself and by others
– as needed

- **Organizing** files into folders
- Using **descriptive file names**
- **Devising** templates that can be used for multiple projects

Helping with Data Organization

- **Novice:** Point researchers to best practices content
- **Intermediate:** Help researchers apply best practices to their research
- **Advanced:** In depth consultations on research data organization and ability to recommend specific tools or strategies based user data





Metadata

describing your data and other research materials so they can be understood – by yourself and by others – as needed

- Including **README** files to describe the data
- Using **commonly used metadata standards**
- Creating a standard where none is available

Helping with Metadata

- **Novice:** Point researchers to information online about metadata standards
- **Intermediate:** Help researchers find appropriate metadata standards
- **Advanced:** Help researchers devise a metadata standard where none are available





[Sharing by ryanrc](#)

Sharing Data

making your data available
so that they can be
accessed and used - by
yourself or by others - in
the future

- **Findable:** How will users know it exists?
 - Put in a repository w/ DOI
- **Accessible:** How will users get the data?
 - Repositories + Data access statement
- **Interoperable:** Can it be combined with similar datasets for meta-analyses?
 - Use common (meta)data standards
- **Reusable:** Can users use it for new purposes?
 - Use open file formats and permissive licenses (CC-0)

Helping with Data Sharing

- **Novice:** Help researchers find data repositories to share their data
- **Intermediate:** Help researchers prepare their data to meet researcher requirement
- **Advanced:** Facilitate research data submission into your institutional repository



What do RDM services look like in practice?

3 libraries, 3 different positions

- Education and Reference
- Digital Collection Services
- Science and Engineering Libraries (Public Services)



Strauss Health Sciences Library
UNIVERSITY OF COLORADO ANSCHUTZ MEDICAL CAMPUS



COLORADO STATE UNIVERSITY | LIBRARIES



LIBRARIES

University of Wisconsin-Madison

FIND

STEENBOCK LIBRARY ▾

Agricultural & Life Sciences, Engineering, Veterinary Medicine

3 libraries, 3 different target audiences

- Basic Science



Strauss Health Sciences Library
UNIVERSITY OF COLORADO ANSCHUTZ MEDICAL CAMPUS

- Everyone



COLORADO STATE UNIVERSITY | LIBRARIES

- STEM researchers



LIBRARIES

University of Wisconsin-Madison

FIND

STEENBOCK LIBRARY ▾

Agricultural & Life Sciences, Engineering, Veterinary Medicine

Services @ Health Sciences

- Informatics workshop series
- Library workshops
- Consultation services



Health Sciences Library

UNIVERSITY OF COLORADO ANSCHUTZ MEDICAL CAMPUS

The Health Sciences Library's Informatics Workshop Series presents:

Biological Interpretation of Gene Expression Data

Michael C. Edwards

University of Colorado Anschutz Medical Campus

Bioinformatics



Data Management



Services @ Colorado State

- Workshops:
 - Data management
 - R programming
- Consultations
- Research data self-submission

CODING & COOKIES



DATA & DONUTS



MOUNTAIN SCHOLAR
DIGITAL COLLECTIONS OF COLORADO

Joined Teams at UW-Madison

- Consulted on RDM topics
- Teach coding workshops, develop new lessons
- Support use of Electronic Lab notebook team



Advice

1. **Get to know your campus**

- Who is your target audience?
- What resources are already available on campus?
- Develop partnerships

2. **Fill in gaps in service**

- What value can you add to your institution?
- Know your limits!

3. **Keep an open mind**

- Unexpected opportunities can be fruitful

Professional development

Conferences; Communities of practice; training programs

Research Data Access & Preservation Association



<https://rdapassociation.org/>

RDAP Mission

RDAP supports an engaged community of **data practitioners** committed to creating, maintaining, advancing, and teaching best practices for research data, access, and preservation.

RDAP Community

The RDAP community brings together a variety of individuals, including **data managers** and **curators**, **librarians**, **archivists**, **researchers**, **educators**, **students**, **technologists**, and **data scientists** from academic institutions, data centers, funding agencies, and industry who represent a wide range of STEM disciplines, social sciences, and humanities.

2020 RDAP Summit

Connecting Through Data

March 11-13, 2020

Santa Fe, New Mexico, USA

<https://rdapassociation.org/summit-info/>

Keep in touch with RDAP

- RDAP listserve:

<http://mail.kunverj.com/mailman/listinfo/rdap>

- Become a member (FREE until July 2020):

<https://rdapassociation.org/membership/>

Research Data Alliance



<https://www.rd-alliance.org/>

RDA Plenary

**15th Research Data Alliance
Plenary Meeting**

Melbourne Convention and Exhibition Centre, MCEC

**Data for Real-World
Impact**

**18-20
MARCH
2020**

**MELBOURNE
AUSTRALIA**

RDA
RESEARCH DATA ALLIANCE

CSIRO

ARDC
Australian Research Data Commons

#rdaplenary
www.rd-alliance.org



**THE
CARPENTRIES**

<https://carpentries.org/>



THE
CARPENTRIES

ABOUT

SPONSORSHIP

CONTACT

CARPENTRYCON 2020

Growing Inclusive, Computational Communities and Leaders

29 June - 1st July, 2020

University of Wisconsin - Madison, USA

<https://2020.carpentrycon.org/>



RESEARCH DATA MANAGEMENT LIBRARIAN ACADEMY

<https://rdmla.github.io/>

Other resources

- Data Curation Profiles Toolkit: <http://datacurationprofiles.org/>
- Best practices
 - University of Wisconsin-Madison: <http://researchdata.wisc.edu/learn-about-data-management/>
 - University of Minnesota: <https://www.lib.umn.edu/datamanagement/best-practices-templates>
- Data One Education modules: <https://www.dataone.org/education-modules>
- DataCure: <https://docs.google.com/document/d/1ojXRTbVy51vkABZDst2r-rEm34VxcxIRSuisOU8pXGo/edit>

Thank you!

Questions?

Slides:

<http://bit.ly/2CJhnz2>

Acknowledgements

- Elsevier
- UW-Madison
 - Science and Engineering Libraries
 - Research Data Services
 - Office of Research Cyberinfrastructure
- RDAP community
- Colleagues
 - Trisha Adamus
 - Cameron Cook
 - Heather Shimon
 - Sarah Stevens