2019-02-13

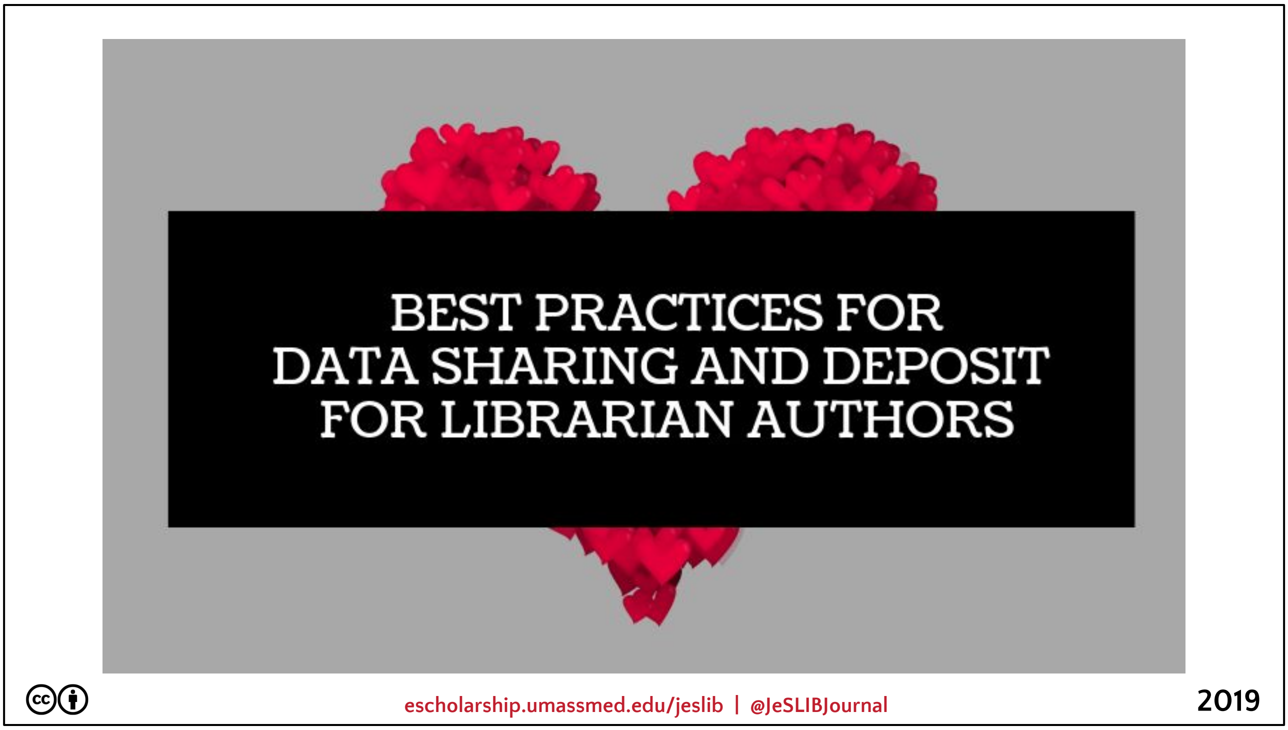
Best Practices for Data Sharing and Deposit for Librarian Authors

Regina Fisher Raboin
University of Massachusetts Medical School

Et al.

Recommended Citation

Raboin RF, Plutchak T, Palmer LA, Goldman J. (2019). Best Practices for Data Sharing and Deposit for Librarian Authors. Journal of eScience Librarianship Presentations. Retrieved from https://escholarship.umassmed.edu/jeslib_presentations/1



BEST PRACTICES FOR DATA SHARING AND DEPOSIT FOR LIBRARIAN AUTHORS

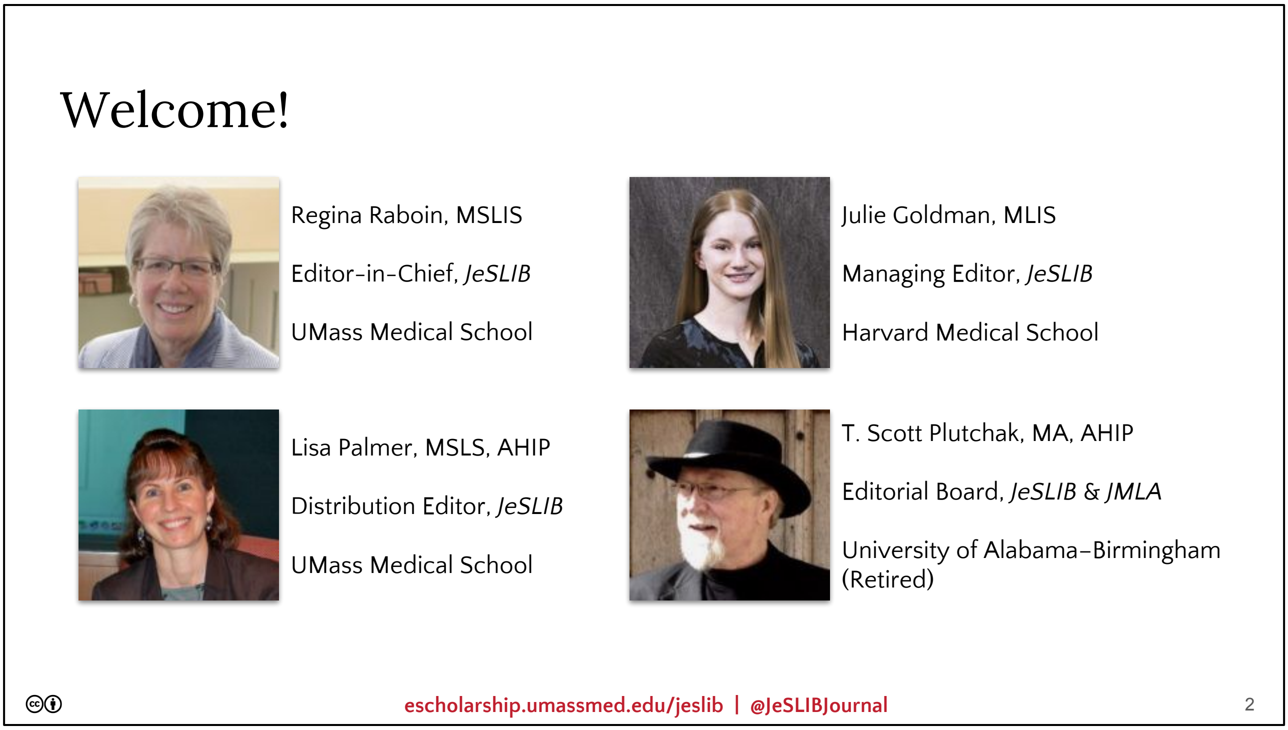
escholarship.umassmed.edu/jeslib | @JeSLIBJournal

2019

Best Practices for Data Sharing and Deposit for Librarian Authors

Thank you to Kate Nyhan and NEASIST, and Cathy Nash and Rodneikka Scott and ASIST for inviting us to present on this important topic.

Now let’s get to the heart of the matter.



Welcome!

Regina Raboin, MSLIS

Editor-in-Chief, *JeSLIB*

UMass Medical School

Julie Goldman, MLIS

Managing Editor, *JeSLIB*

Harvard Medical School

Lisa Palmer, MSLS, AHIP

Distribution Editor, *JeSLIB*

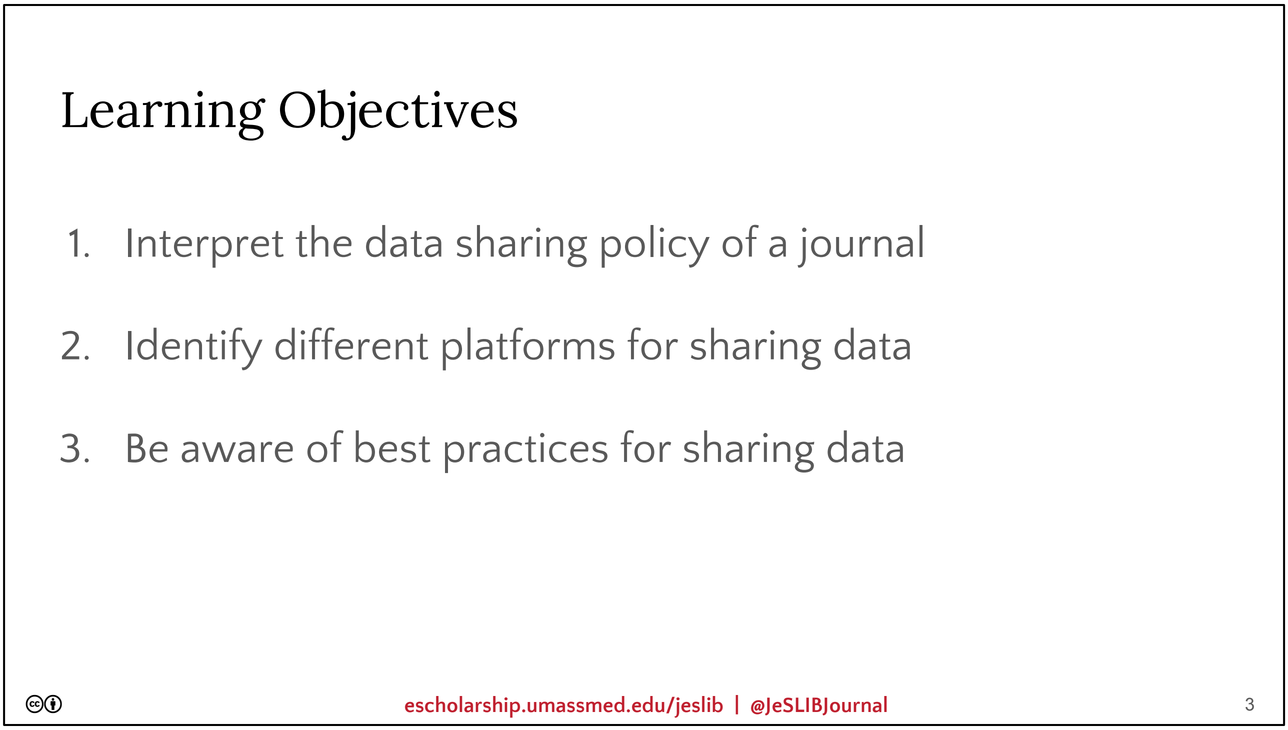
UMass Medical School

T. Scott Plutchak, MA, AHIP

Editorial Board, *JeSLIB* & *JMLA*

University of Alabama–Birmingham (Retired)

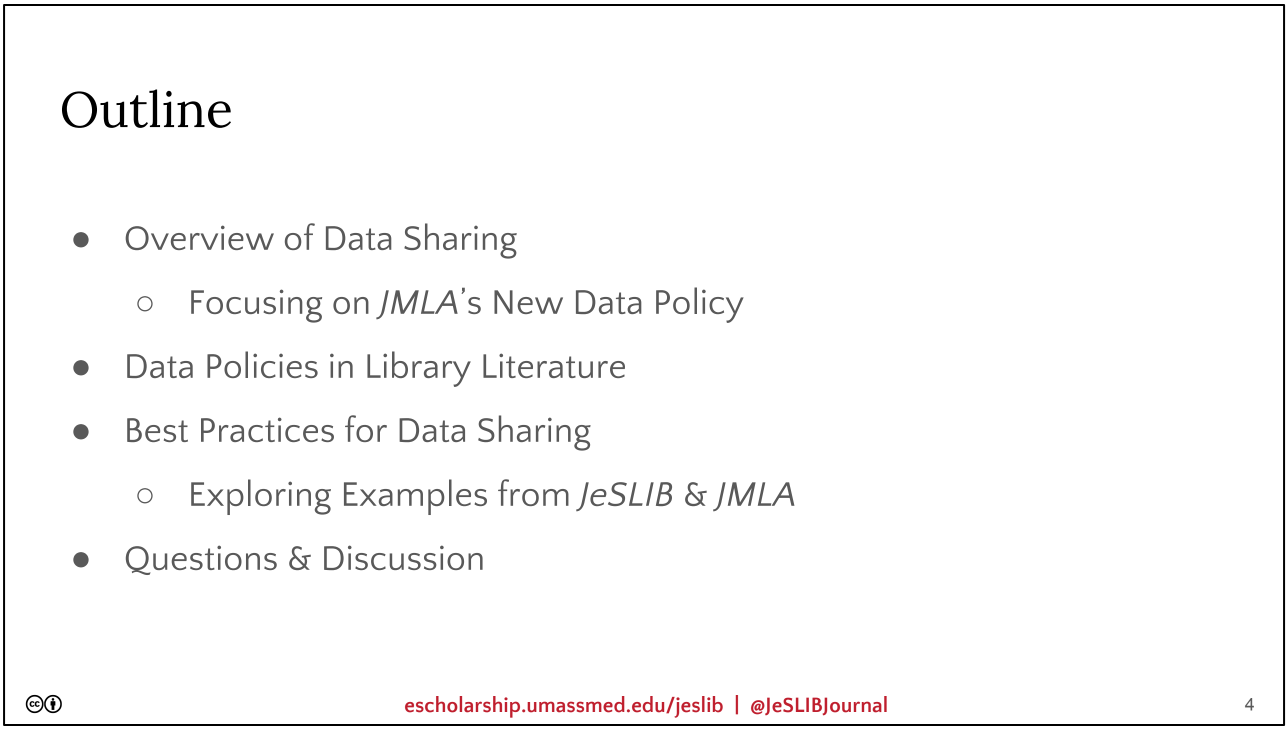
Today’s webinar is designed and presented by the Journal of eScience Librarianship’s Editors Julie Goldman, Managing Editor; Lisa Palmer, Distribution Editor; and me, Regina Raboin, Editor; and we are joined by T. Scott Plutchak, Editorial Board member for JeSLIB (Journal of eScience Librarianship) and JMLA (Journal of the Medical Library Association).



Learning Objectives

1. Interpret the data sharing policy of a journal
2. Identify different platforms for sharing data
3. Be aware of best practices for sharing data

The learning objectives for this webinar are designed for librarians as researchers and authors focused on depositing data to accompany articles and other works submitted for publication.

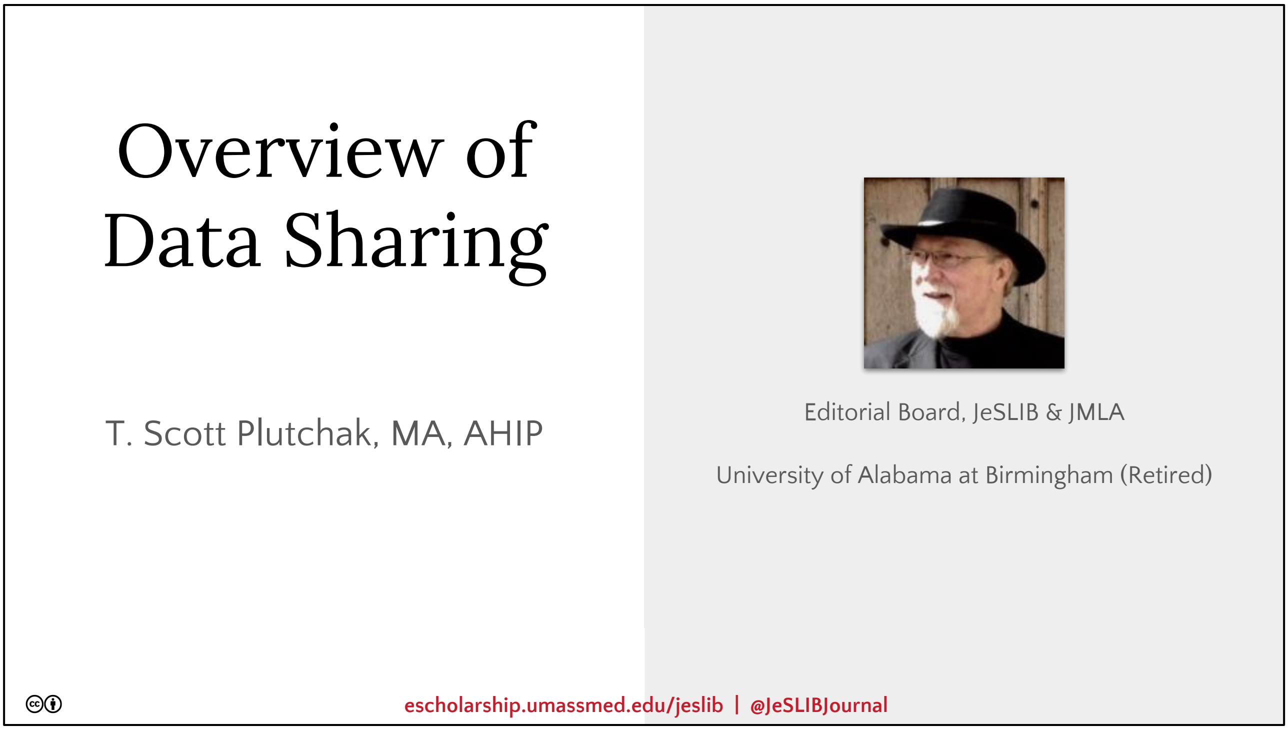


Outline

- Overview of Data Sharing
  - Focusing on *JMLA*'s New Data Policy
- Data Policies in Library Literature
- Best Practices for Data Sharing
  - Exploring Examples from *JeSLIB* & *JMLA*
- Questions & Discussion

The content we will be covering today is:

- A brief outline of why data sharing is important, incentives and impact of funder policies and journal policies
- Everyone knows what data is, but it's important to understand that the definition of data is changing and overtime journals will evolve their policies, different journals define "data" differently; distinction between supplementary materials and data, and authors need to know which journals to target when submitting manuscripts.
- Review the new JMLA open data policy and creation
- A brief comparison of 'data deposit requirement policies' of library science journals to selected science journals
- Best practices and how to approach data sharing
- Explore examples of supplementary materials in JeSLIB and explore specific examples of data deposit in JeSLIB and JMLA
- Questions & Discussion
- And a list of resources will complete the webinar



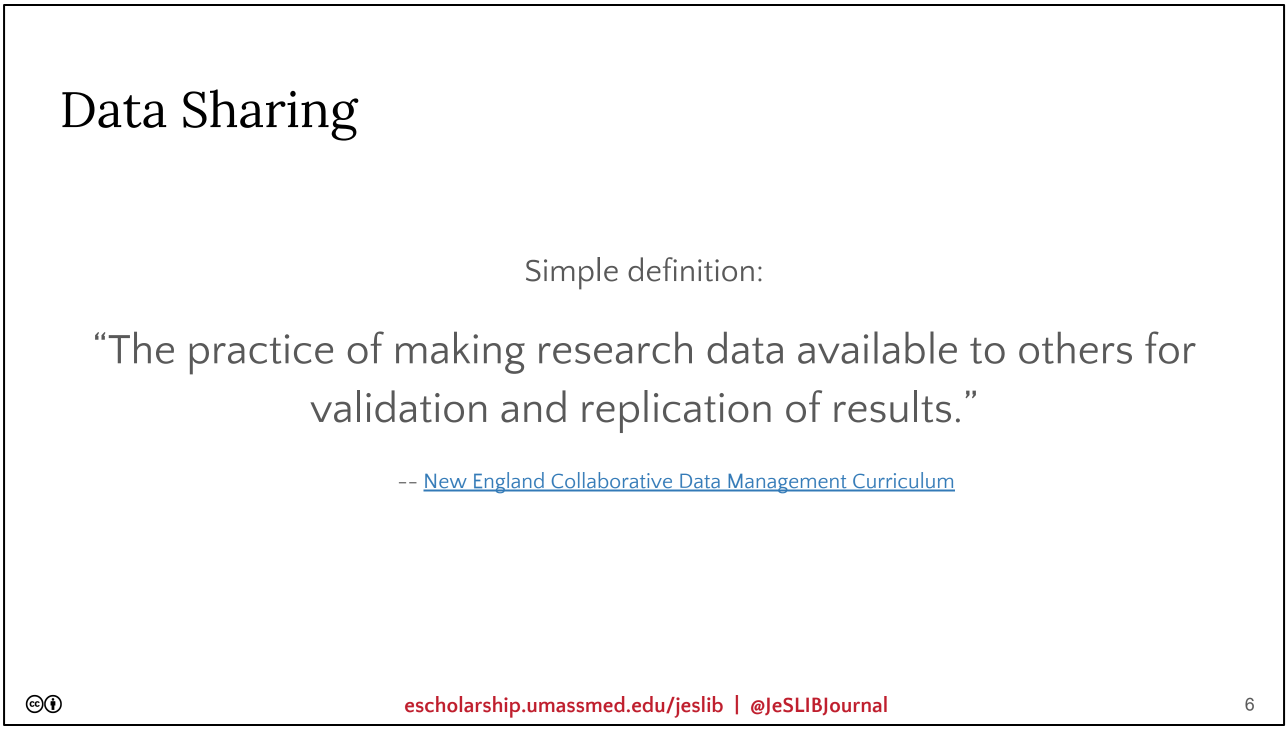
Overview of Data Sharing

T. Scott Plutchak, MA, AHIP

Editorial Board, JeSLIB & JMLA

University of Alabama at Birmingham (Retired)

Overview of Data Sharing: T. Scott Plutchak



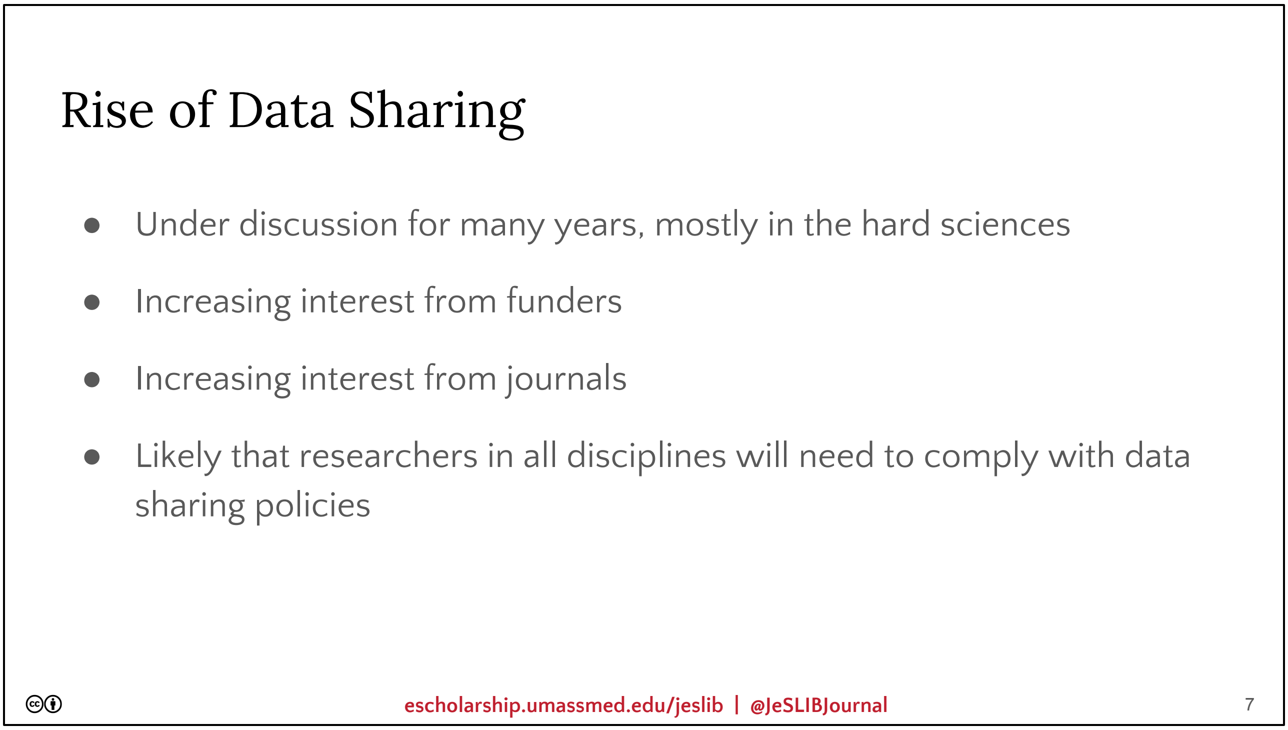
Data Sharing

Simple definition:

“The practice of making research data available to others for validation and replication of results.”

-- New England Collaborative Data Management Curriculum

There are both simple and complex definitions to the term “data sharing,” but the bottom line is that the practice of data sharing brings a higher level of integrity to research (via the ability to validate and replicate results), speeds the research process by saving others the time to generate data that’s already available, and allowing funders the ability to access the data behind the research they support.

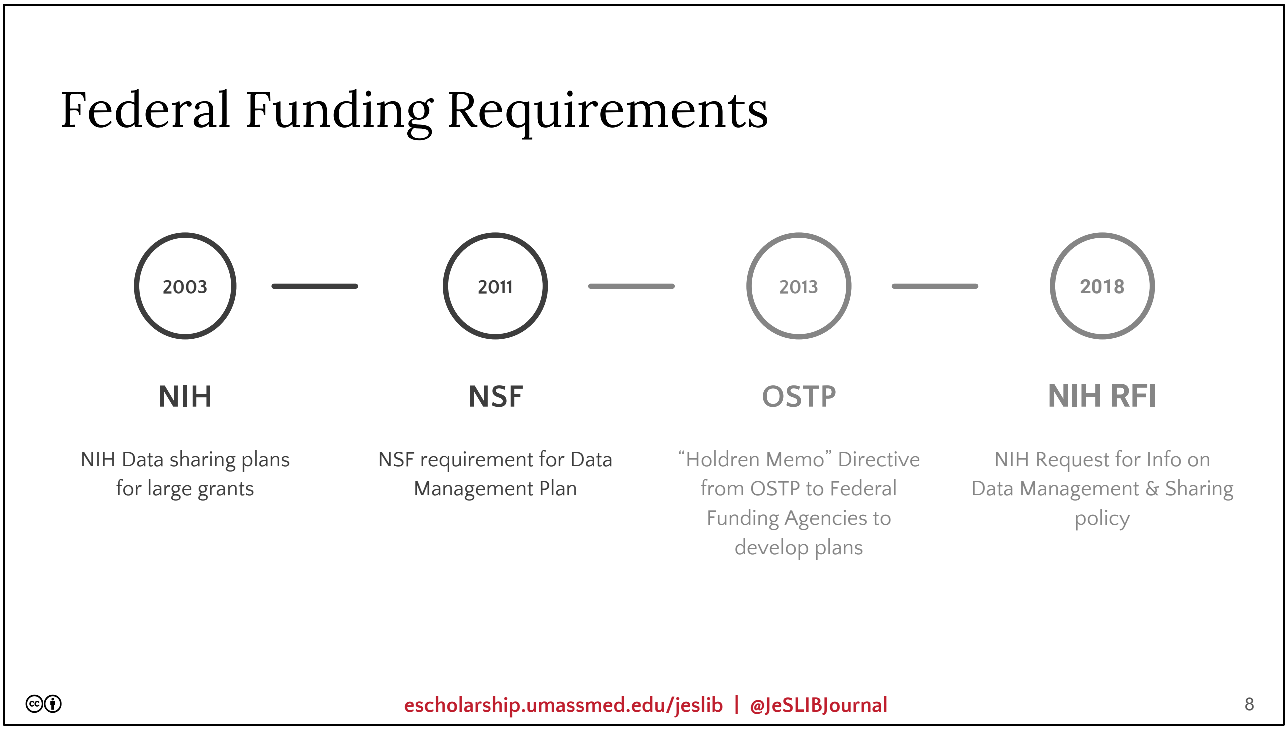


Rise of Data Sharing

- Under discussion for many years, mostly in the hard sciences
- Increasing interest from funders
- Increasing interest from journals
- Likely that researchers in all disciplines will need to comply with data sharing policies

The past few years have seen an increasing wave of interest in open data and in policies and practices for sharing research data. Although this interest is primarily in the hard sciences, increasing interest is being seen in the social sciences and humanities as well.

Related to "Increasing interest from journals," the recent Scholarly Kitchen article about Springer Nature and data sharing highlights recent developments: https://scholarlykitchen.sspnet.org/2019/01/30/guest-post-encouraging-data-sharing-a-small-investment-for-large-potential-gain/



Federal Funding Requirements

2003 — 2011 — 2013 — 2018

NIH

NIH Data sharing plans for large grants

NSF

NSF requirement for Data Management Plan

OSTP

“Holdren Memo” Directive from OSTP to Federal Funding Agencies to develop plans

NIH RFI

NIH Request for Info on Data Management & Sharing policy

- NSF has had a modest data sharing policy for years, although some Directorates adopted early, more prescriptive policies and practices. As the building of subject and agency-specific repositories increased, data sharing policies have evolved. NIH had established a minimal data sharing policy for very large grants and some of the individual programs had specific policies

- Following the release of the “Holdren memo” (White House Office of Science and Technology Policy) in February of 2013, all US Federal Funding agencies have developed plans for requiring their grantees to develop plans for sharing their research data.

- These have been developing slowly – NIH just recently held a request for comment period for their draft data sharing policy – but the trends are clear

- Similar policies have been implemented or are being developed at other funding agencies and foundations as well.

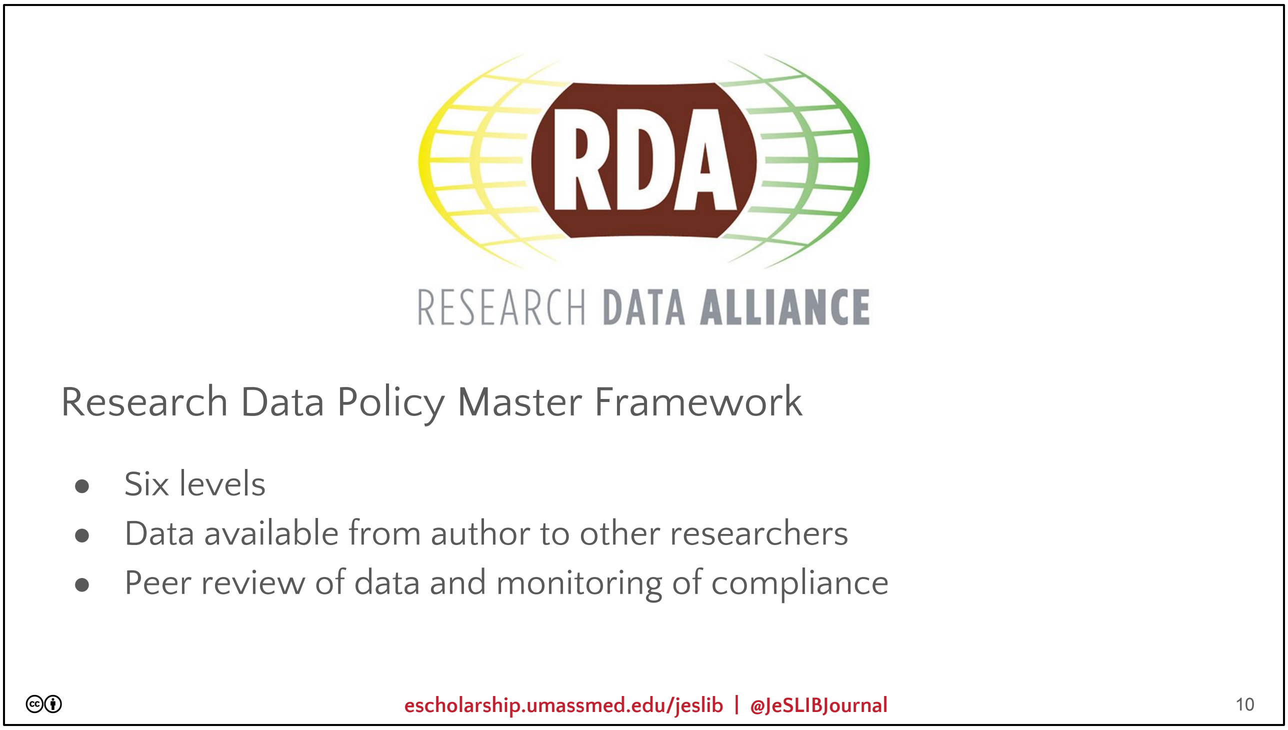
Journal Policies

escholarship.umassmed.edu/jeslib | @JeSLIBJournal

9

In response to these trends and policies, many journals have started to develop policies requiring authors to submit plans for sharing the data underpinning the conclusions of their published articles. Unfortunately, that development has been haphazard, resulting in quite different policies and varying processes for implementation and oversight.

PLOS One: https://journals.plos.org/plosone/
Scientific Data: https://www.nature.com/sdata/
GigaScience: https://academic.oup.com/gigascience
Science: https://www.sciencemag.org/
Springer Nature: https://www.springernature.com/gp

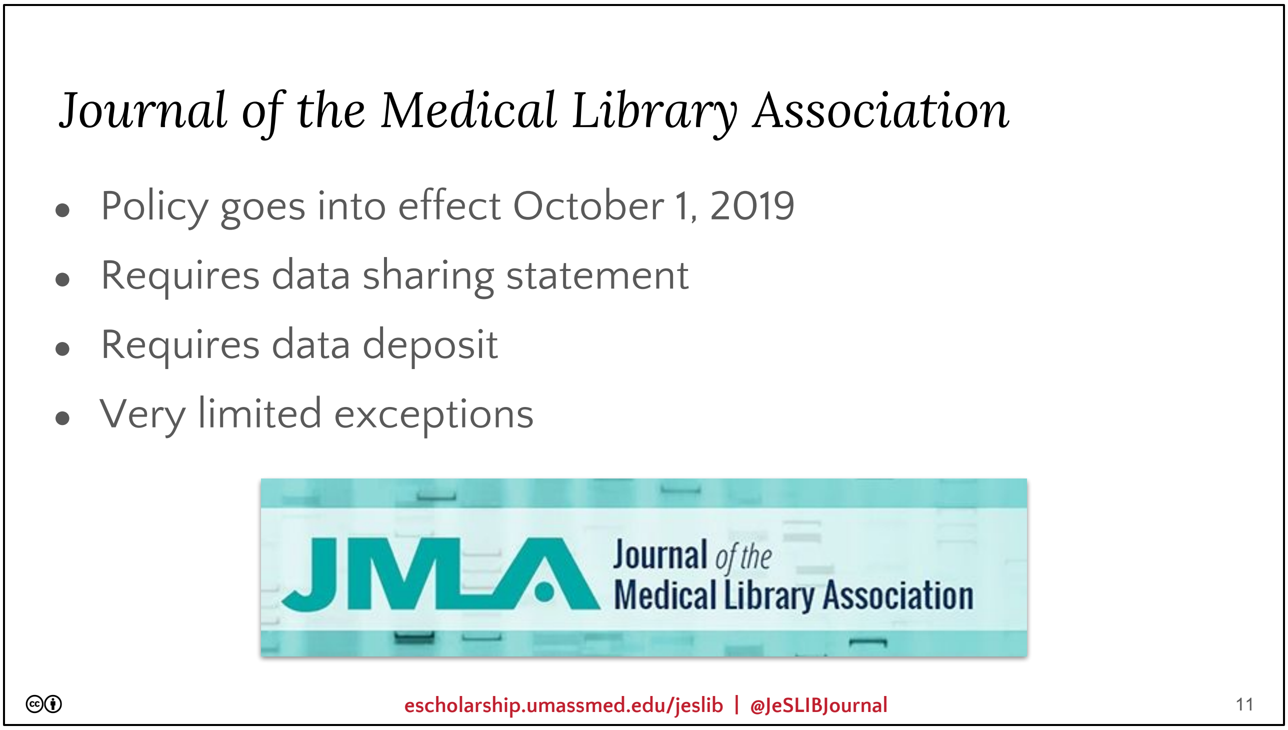

RDA
RESEARCH DATA ALLIANCE

Research Data Policy Master Framework

- Six levels
- Data available from author to other researchers
- Peer review of data and monitoring of compliance

In an attempt to provide some guidance for journals seeking to develop such policies, the Research Data Alliance formed an Interest Group tasked with developing a standardized framework for journals to use in developing such policies. (https://www.rd-alliance.org/groups/data-policy-standardisation-and-implementation-ig)

The RDA guidance provides for several levels of policy, depending on how far a particular journal is ready to go, ranging from a minimal suggestion for data sharing to more robust requirements. The guidance lays out the elements that should be included in a data sharing policy at any level.



Journal of the Medical Library Association

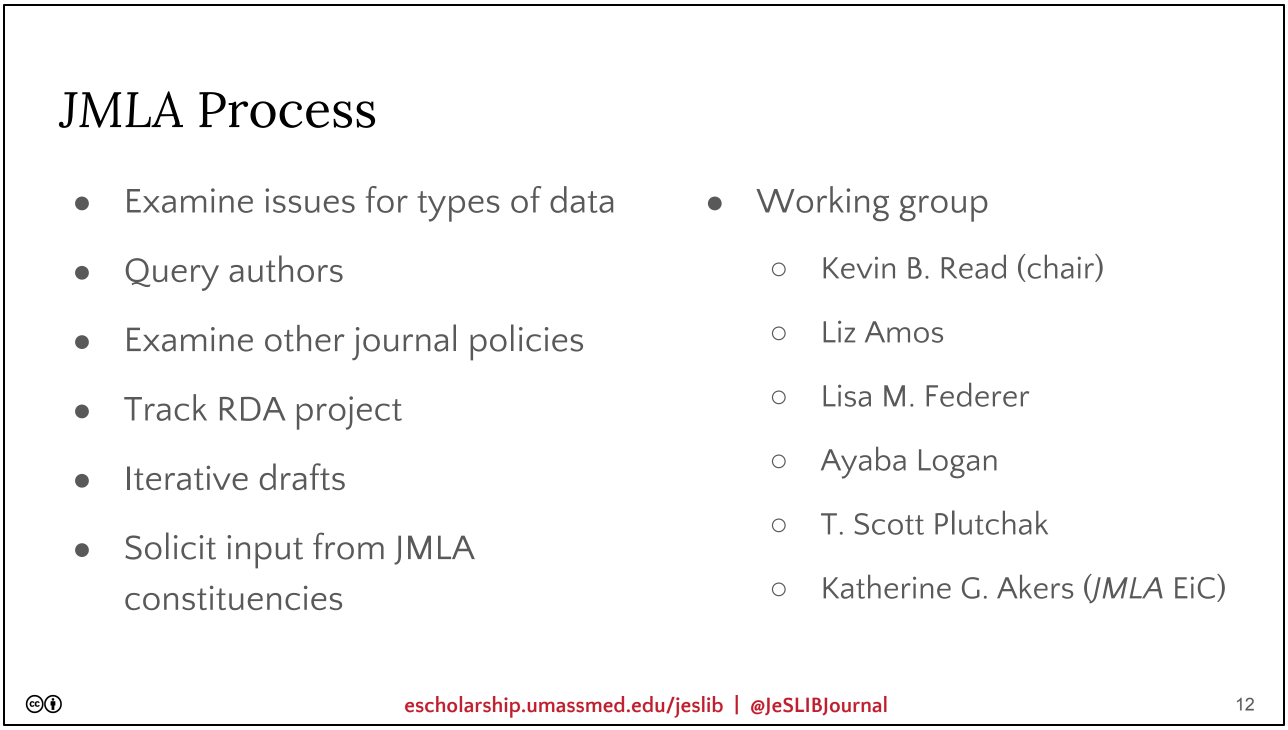
- Policy goes into effect October 1, 2019
- Requires data sharing statement
- Requires data deposit
- Very limited exceptions

JMLA
Journal of the
Medical Library Association

While few library and information science journals have developed such policies to date, it is reasonable to expect that they will, and that librarian authors will need to be prepared to comply.

In June of 2017, the Editor of the JMLA formed a working group to explore developing a policy requiring authors to share their data. The draft policy was presented to the MLA Board of Directors in May 2018 and will go into effect for manuscripts submitted on or after October 1, 2019.

http://jmla.mlanet.org/ojs/jmla/about/editorialPolicies#custom-0

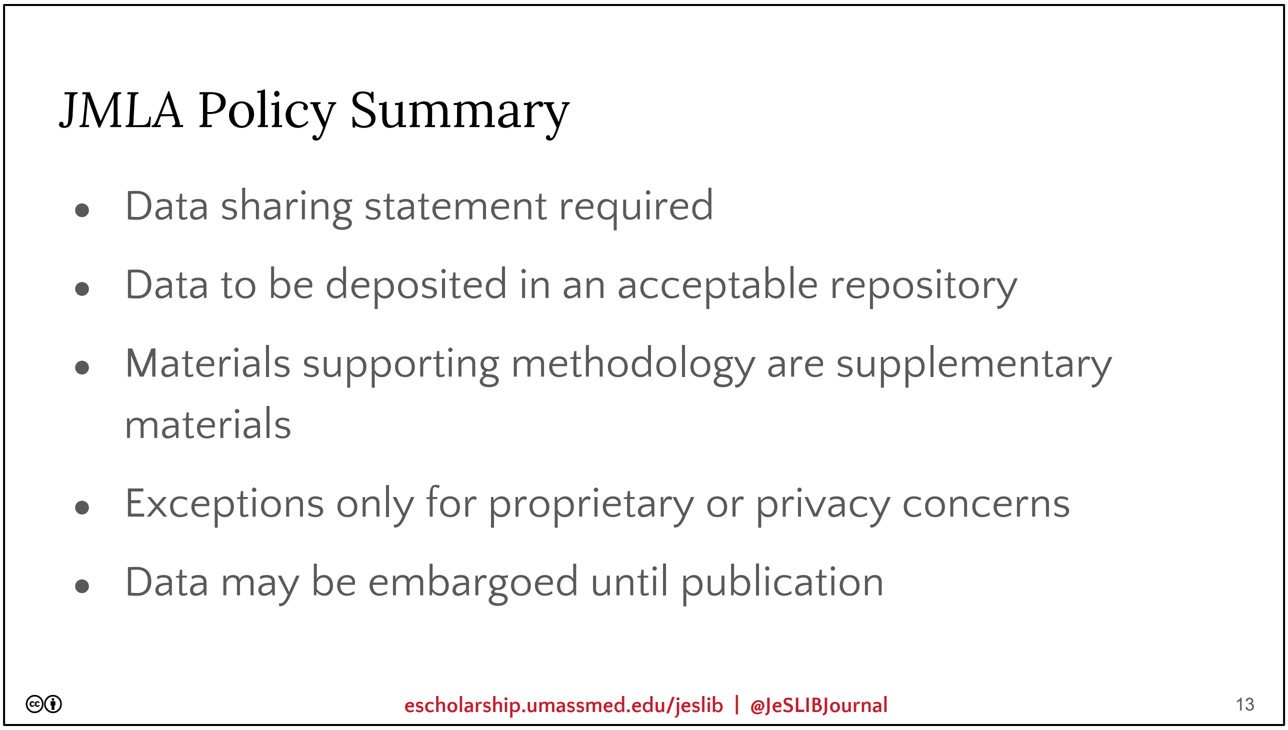


JMLA Process

- Examine issues for types of data
- Query authors
- Examine other journal policies
- Track RDA project
- Iterative drafts
- Solicit input from JMLA constituencies

- Working group
  - Kevin B. Read (chair)
  - Liz Amos
  - Lisa M. Federer
  - Ayaba Logan
  - T. Scott Plutchak
  - Katherine G. Akers (*JMLA* EiC)

JMLA process – analyzed sampling of recent articles to examine types of data; surveyed authors to determine their willingness/ability to share data; worked to align JMLA policy with RDA guidance

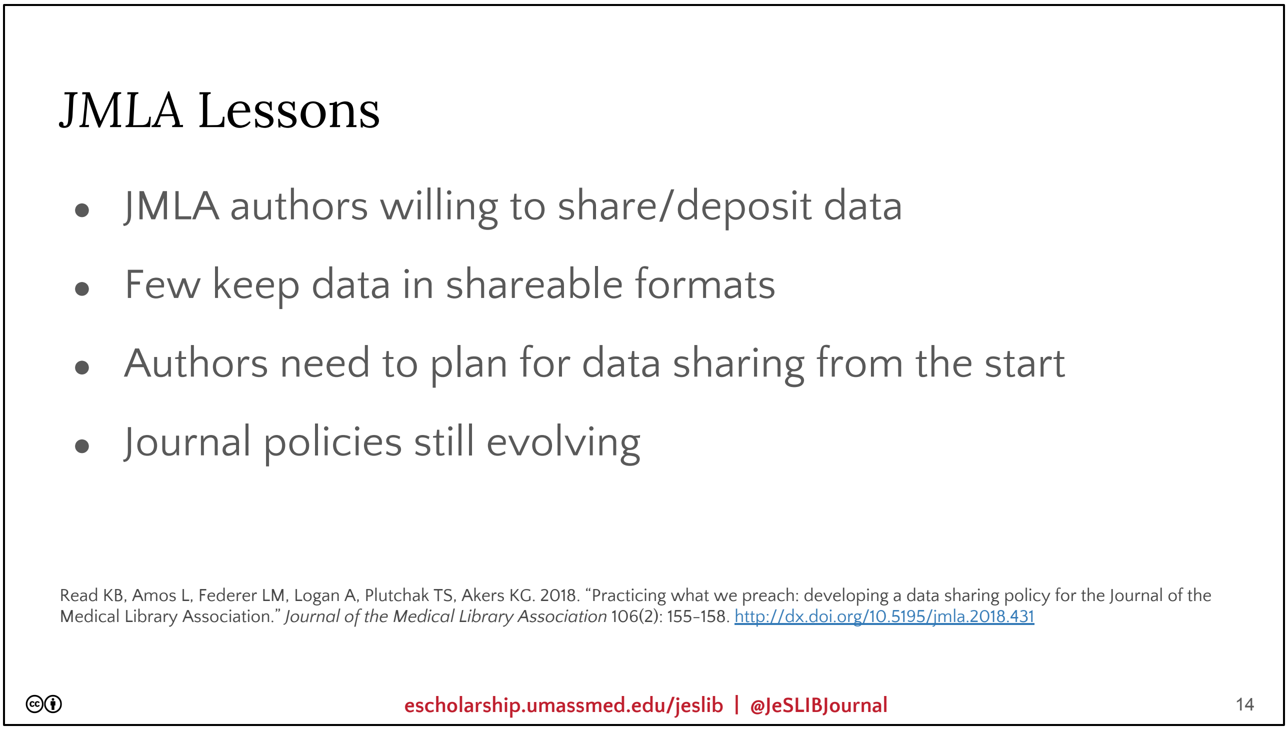


JMLA Policy Summary

- Data sharing statement required
- Data to be deposited in an acceptable repository
- Materials supporting methodology are supplementary materials
- Exceptions only for proprietary or privacy concerns
- Data may be embargoed until publication

JMLA Policy Summary

- Data sharing statement required
- Data to be deposited in an acceptable repository
- Materials supporting methodology are supplementary materials
- Exceptions only for proprietary or privacy concerns
- Data may be embargoed until publication



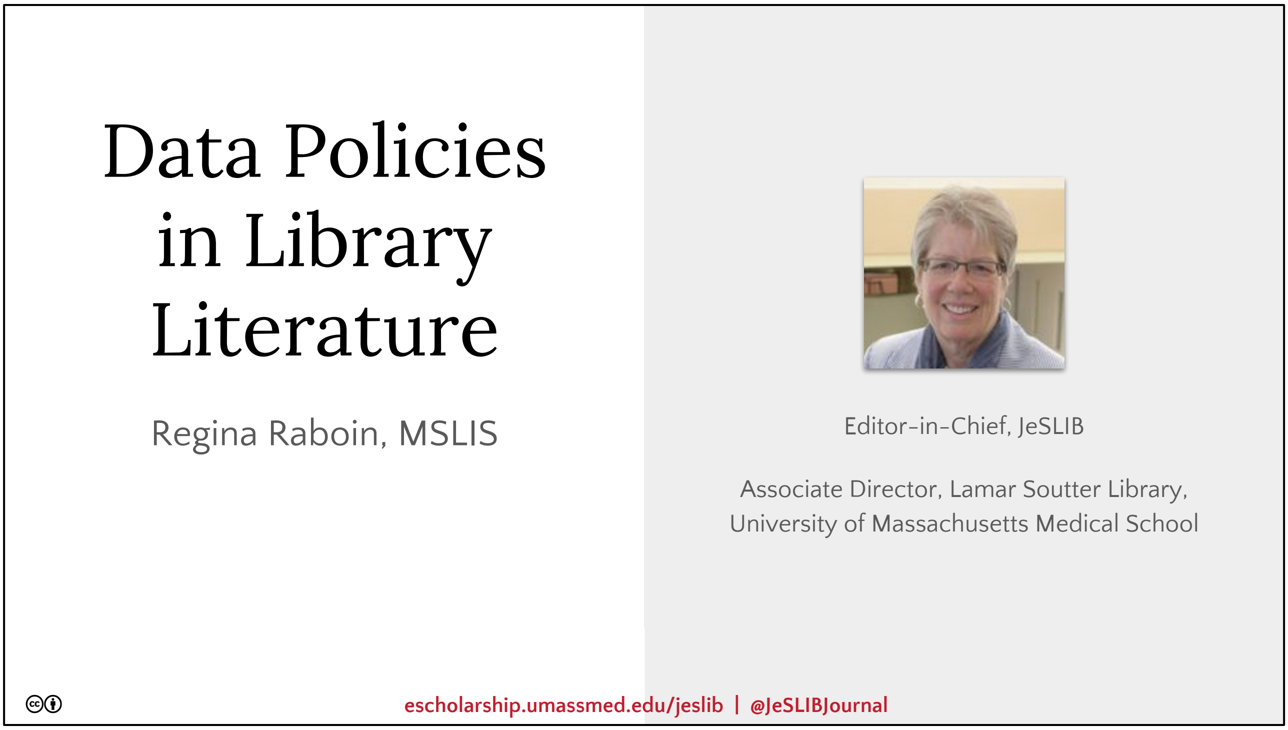
JMLA Lessons

- JMLA authors willing to share/deposit data
- Few keep data in shareable formats
- Authors need to plan for data sharing from the start
- Journal policies still evolving

Read KB, Amos L, Federer LM, Logan A, Plutchak TS, Akers KG. 2018. “Practicing what we preach: developing a data sharing policy for the Journal of the Medical Library Association.” *Journal of the Medical Library Association* 106(2): 155-158. http://dx.doi.org/10.5195/jmla.2018.431

JMLA lessons – most authors willing to share data, but few have data in shareable formats; journal policies are still evolving; authors need to be thinking about data sharing from the beginning of their projects.

Read KB, Amos L, Federer LM, Logan A, Plutchak TS, Akers KG. 2018. “Practicing what we preach: developing a data sharing policy for the Journal of the Medical Library Association.” *Journal of the Medical Library Association* 106(2): 155-158. http://dx.doi.org/10.5195/jmla.2018.431



Data Policies in Library Literature

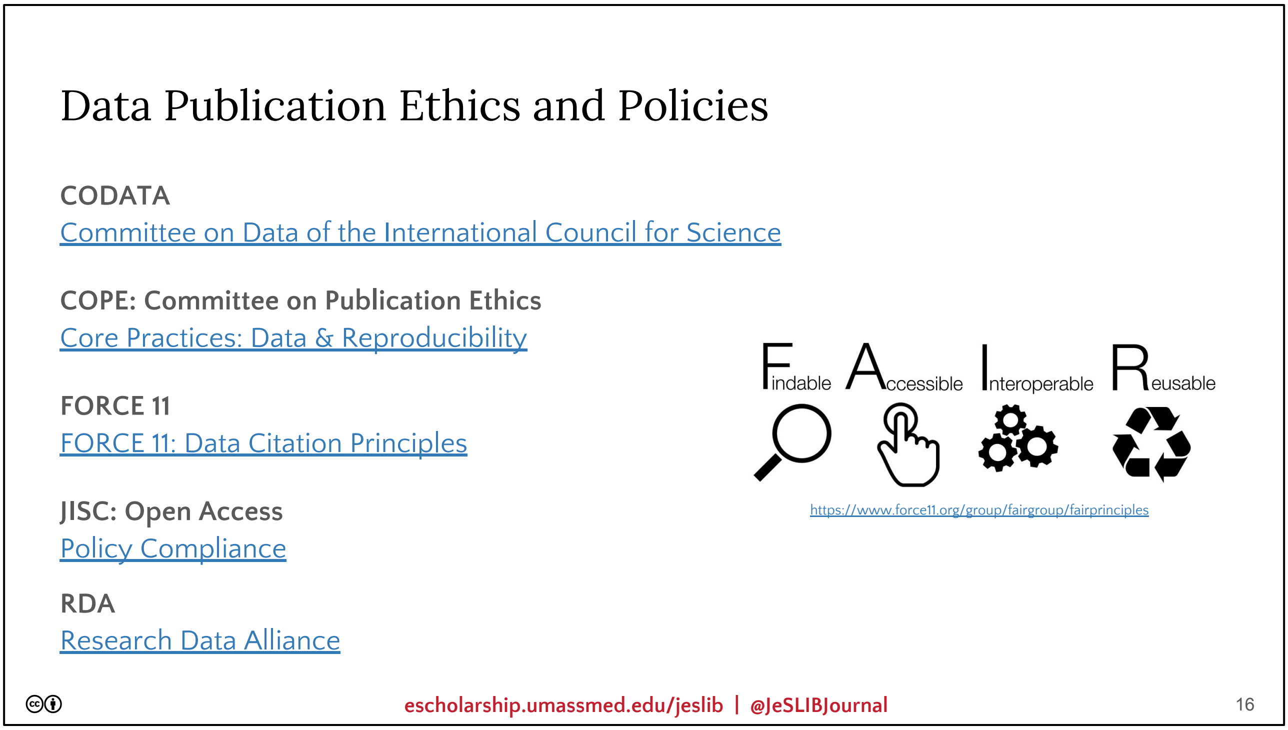
Regina Raboin, MSLIS

Editor-in-Chief, JeSLIB

Associate Director, Lamar Soutter Library,
University of Massachusetts Medical School

escholarship.umassmed.edu/jeslib | @JeSLIBJournal

Data Policies in Library Literature: Regina Raboin



Data Publication Ethics and Policies

CODATA
Committee on Data of the International Council for Science

COPE: Committee on Publication Ethics
Core Practices: Data & Reproducibility

FORCE 11
FORCE 11: Data Citation Principles

JISC: Open Access
Policy Compliance

RDA
Research Data Alliance

Findable Accessible Interoperable Reusable

https://www.force11.org/group/fairgroup/fairprinciples

Before discussing best practices, it's important to understand [that] policies and procedures surrounding data are not created in a vacuum. The organizations listed above, along with others not listed here, are dedicated to providing sound ethical frameworks surrounding the use, distribution, and care of data. These organizations provide safe and informed venues for debate and discussion, fostering shared values and collaboration in order to build important and feasible data policy frameworks.

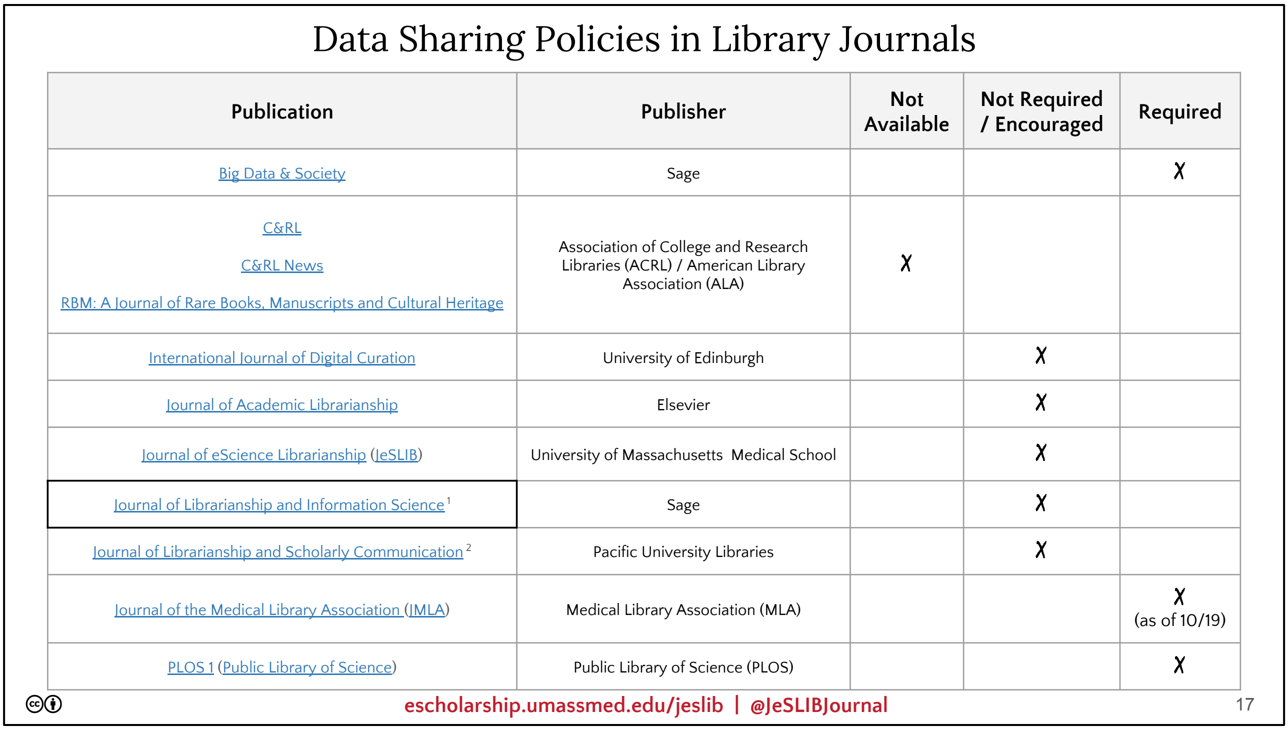
COPE recommends journals should include policies on data availability and encourage the use of reporting guidelines and registration of clinical trials and other study designs according to standard practice in their discipline https://publicationethics.org/core-practices

FORCE11 recommends data should be considered legitimate, citable products of research. Data citation, like the citation of other evidence and sources, is good research practice and is part of the scholarly ecosystem supporting data reuse. In other words, data citations should facilitate access to the data themselves and to such associated metadata, documentation, code, and other materials, as are necessary for both humans and machines to make informed use of the referenced data https://www.force11.org/datacitationprinciples | Wilkinson, M. D. et al. 2016. The FAIR Guiding Principles for scientific data management and stewardship. Sci. Data 3:160018 https://doi.org/10.1038/sdata.2016.18

CODATA has been taking a leadership role in developing and advancing CODATA's Open Research Data agenda at national and international levels. The Science

International Accord on Open Data in a Big Data World, which CODATA also took the lead on drafting, recognizes both the benefits that accrue to science through open data regimes and the responsibility of scientists, and others, to make data available through appropriate means and minimal delay.
https://zenodo.org/record/27872#.XEY1dCBOnD4

RDA: as previously mentioned by Scott, RDA's Data Policy and Standardisation and Implementation Interest Group developed a draft standardized framework intended to assist journal editors in navigating the creation or enhancement of a data availability policy.
https://docs.google.com/document/d/1Z4VWeQ8hMBWOpdky-Ec7GercM-4_cSDKa8gEu0eXz04/edit?usp=sharing



Data Sharing Policies in Library Journals

Publication	Publisher	Not Available	Not Required / Encouraged	Required
Big Data & Society	Sage			✗
C&RL C&RL News RBM: A Journal of Rare Books, Manuscripts and Cultural Heritage	Association of College and Research Libraries (ACRL) / American Library Association (ALA)	✗		
International Journal of Digital Curation	University of Edinburgh		✗	
Journal of Academic Librarianship	Elsevier		✗	
Journal of eScience Librarianship (JeSLIB)	University of Massachusetts Medical School		✗	
Journal of Librarianship and Information Science[1]	Sage		✗	
Journal of Librarianship and Scholarly Communication[2]	Pacific University Libraries		✗	
Journal of the Medical Library Association (JMLA)	Medical Library Association (MLA)			✗ (as of 10/19)
PLOS 1 (Public Library of Science)	Public Library of Science (PLOS)			✗

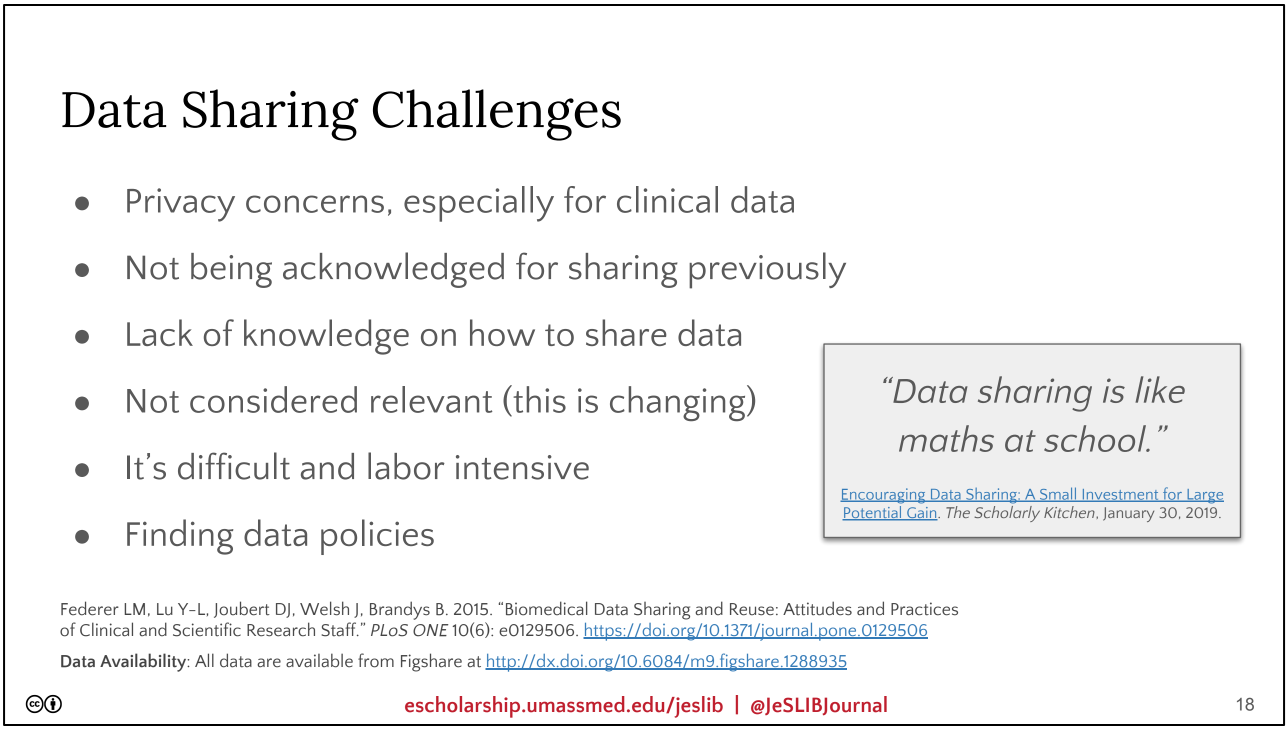
escholarship.umassmed.edu/jeslib | @JeSLIBJournal

17

This slide illustrates library and information science journals varied approaches to data sharing

- The majority of the listed journals have data sharing policies, but don't require or encourage a data sharing statement.
- [1.] Interestingly, the Journal of Librarianship and Information Science offers optional open access publishing via the SAGE Choice program, but the journal Big Data & Society, also published by Sage requires a data sharing policy.
- [2.] Journal of Librarianship and Scholarly Communications offers authors their JLSC Dataverse instance to deposit and share data.

It's important to understand that as authors you need to look for a policy at the journal level - you can't assume that the policy will be same for all of a publisher's journals. As a researcher and author considering how you will share data, among other things, should be considered well before writing and submitting a manuscript.



Data Sharing Challenges

- Privacy concerns, especially for clinical data
- Not being acknowledged for sharing previously
- Lack of knowledge on how to share data
- Not considered relevant (this is changing)
- It's difficult and labor intensive
- Finding data policies

> *"Data sharing is like maths at school."*
>
> Encouraging Data Sharing: A Small Investment for Large Potential Gain. *The Scholarly Kitchen*, January 30, 2019.

Federer LM, Lu Y-L, Joubert DJ, Welsh J, Brandys B. 2015. "Biomedical Data Sharing and Reuse: Attitudes and Practices of Clinical and Scientific Research Staff." *PLoS ONE* 10(6): e0129506. https://doi.org/10.1371/journal.pone.0129506

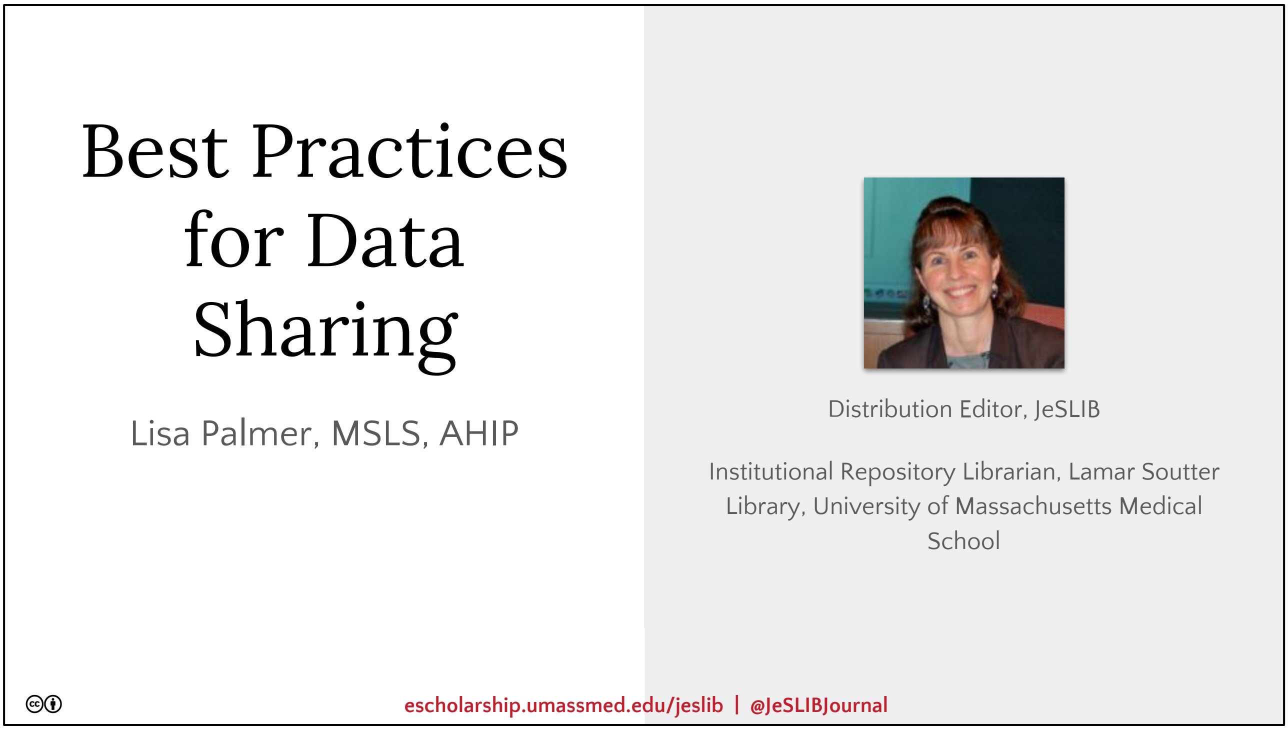
Data Availability: All data are available from Figshare at http://dx.doi.org/10.6084/m9.figshare.1288935

We want to acknowledge that data sharing compliance is not easy. Clinical and basic science researchers perceive many barriers, so it's understandable that librarians will too. The 2015 study by Lisa Federer et al. identifies some of these obstacles.

If your research is funded, the first, and we would say the most important, priority is to be aware of your funder's policies regarding data sharing -- some data sharing policies by funders has been compiled by SPARC at https://sparcopen.org/our-work/research-data-sharing-policy-initiative/funder-policies but if you are unable to locate a policy, you should contact your funder/program officer directly.

Always consult journal data policies before you submit, just as you would consult the author guidelines/instructions. Many library journals do not yet have a policy or guidelines, and the policies vary by journal - so you really need to go to the journal. Information is generally found under "editorial policies" or "instructions for authors"

And please note that if you have human subjects/patient data, you must also be aware of Institutional Review Board (IRB) policies and requirements and consult with this board before gathering or publishing data.



Best Practices for Data Sharing

Lisa Palmer, MSLS, AHIP

Distribution Editor, JeSLIB

Institutional Repository Librarian, Lamar Soutter Library, University of Massachusetts Medical School

escholarship.umassmed.edu/jeslib | @JeSLIBJournal

Best Practices for Data Sharing: Lisa Palmer

Data Sharing Considerations

~~Why to share~~

What to share

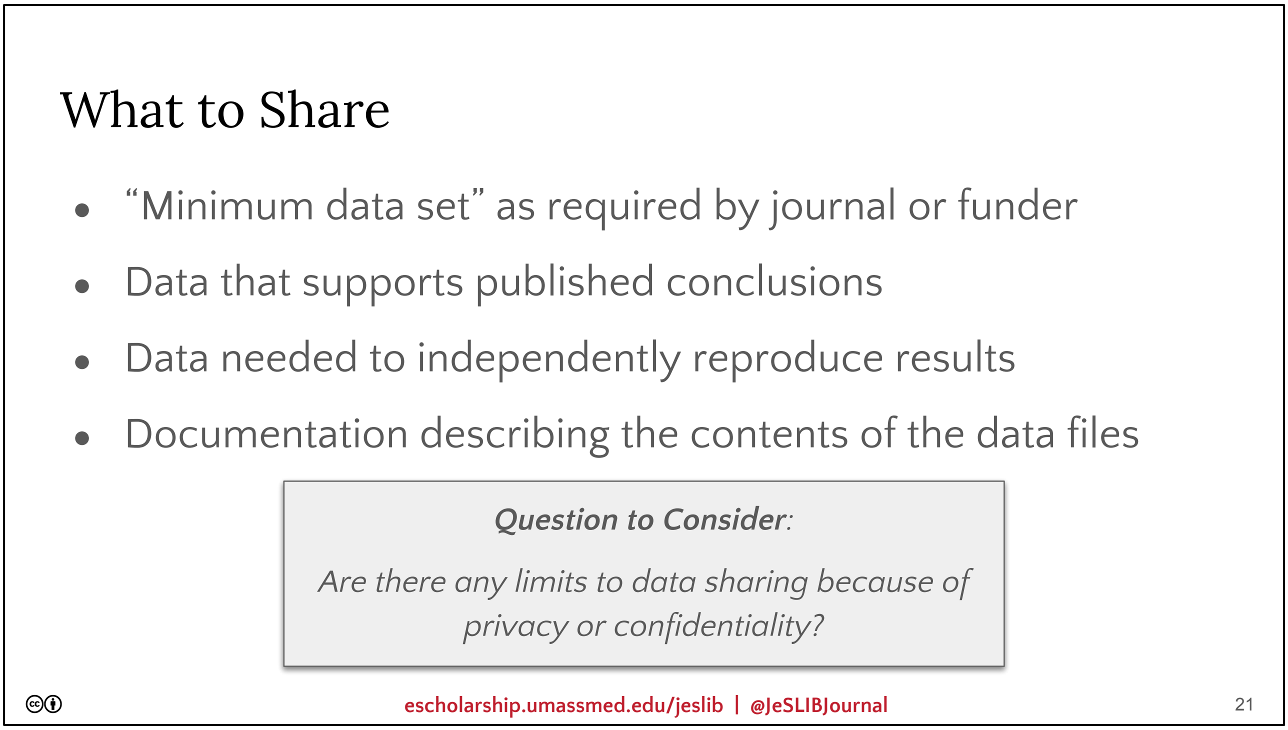
When to share

Where to share

How to share

Thank you, Regina. Data sharing policies place expectations on authors for what to share, where to share, when to share, and how to share. For librarian researchers and all researchers, planning ahead from the start of your project helps in meeting these data sharing challenges.

This presentation doesn't focus on "why to share" research data, as we are hopeful that the benefits of data sharing are known to our audience today. We'll be concentrating instead on some of the practical applications of the what, when, where, and how.



What to Share

- “Minimum data set” as required by journal or funder
- Data that supports published conclusions
- Data needed to independently reproduce results
- Documentation describing the contents of the data files

***Question to Consider*:**

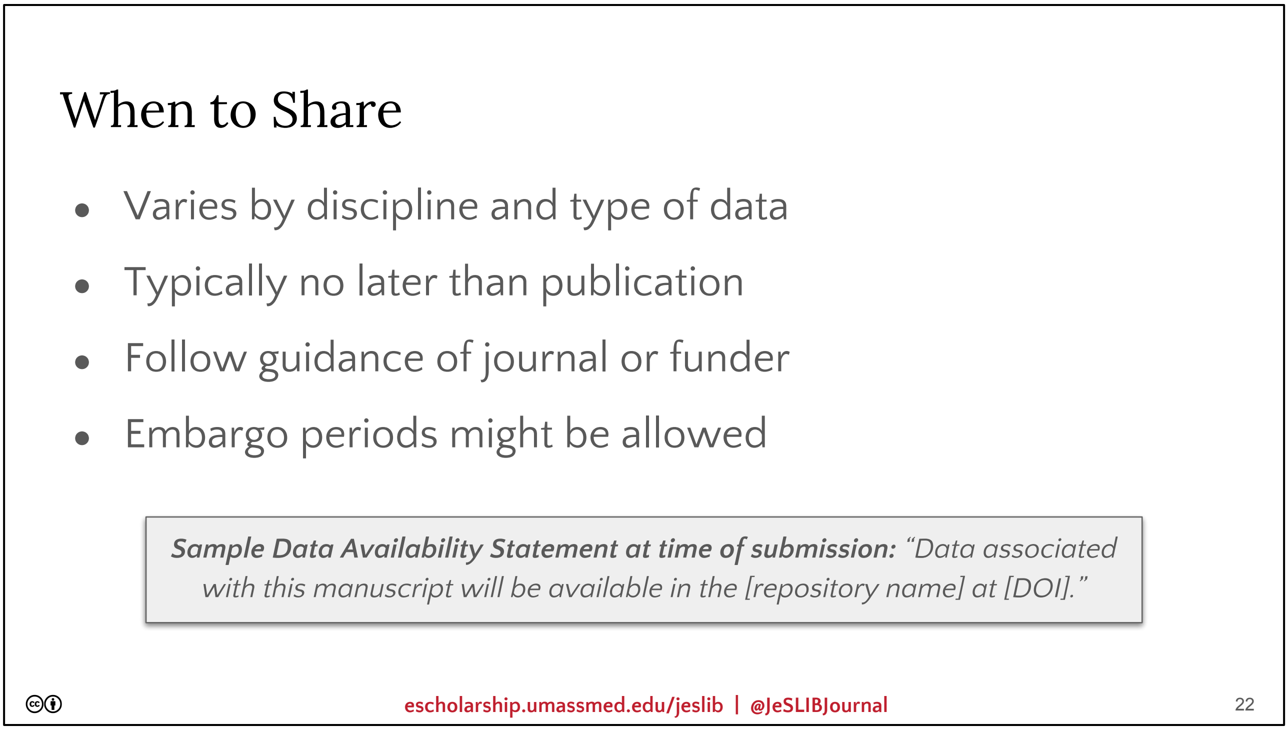
Are there any limits to data sharing because of privacy or confidentiality?

Most journal and funder data policies stipulate that you must share publicly at least the minimum amount of data needed to independently reproduce the results and conclusions described in the manuscript, along with any supporting documentation such as a data dictionary, codebook, or readme files. We’ll go into more detail about data documentation later in the presentation.

Authors generally have some discretion as to what constitutes the “minimal data set”. For example, you usually don’t need to submit your entire data set if only a portion of the data was used in the study. In some cases it will make sense to share the raw data, while in other cases it might make sense to share manipulated or processed data.

If sharing your data is not ethical or legal -- for example, because of privacy, confidentiality, or because your Institutional Review Board has dictated that the data can’t be shared -- consult with the journal and/or funder to find out acceptable alternatives.

PLOS ONE data availability: https://journals.plos.org/plosone/s/data-availability
Taylor & Francis data sharing FAQs:
https://authorservices.taylorandfrancis.com/data-sharing-faqs/
Journal of the Medical Library Association data sharing policy:
http://jmla.mlanet.org/ojs/jmla/about/editorialPolicies#custom-0



When to Share

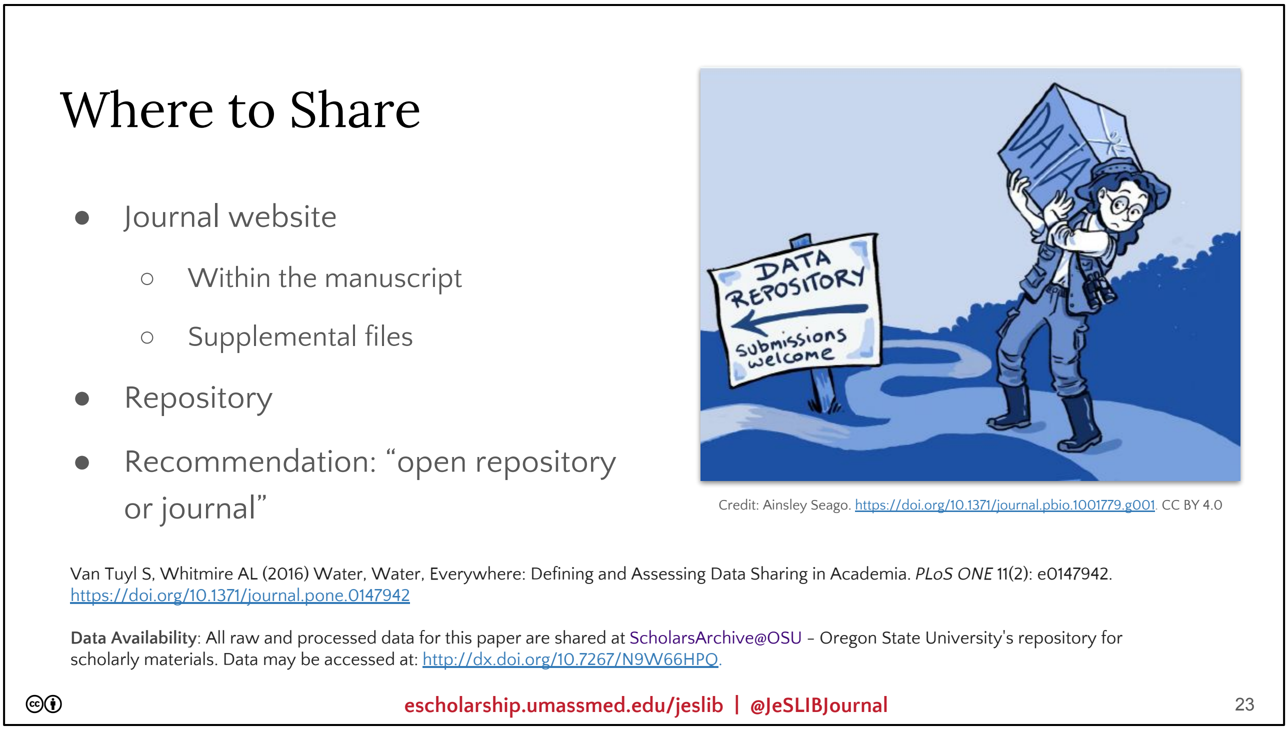
- Varies by discipline and type of data
- Typically no later than publication
- Follow guidance of journal or funder
- Embargo periods might be allowed

> ***Sample Data Availability Statement at time of submission:*** *"Data associated with this manuscript will be available in the [repository name] at [DOI]."*

When you will be required to share your data will vary by discipline, type of data, and policies of the journal and funder. Typically you'll need to provide information about where and how the data can be accessed when your manuscript is accepted, so that a DOI or other persistent identifier can be included in the published manuscript.

If *permanent* closed access is unnecessary for your data or documentation, an embargo might be a good solution. An embargo period is a formal request by an author to restrict access to documents or data for a specified period of time where the data contains, for example, sensitive information, confidential government statistics, or needs to be delayed until after all results and findings from the data have been published.

If you're submitting a manuscript and are not yet sure exactly where you'll be sharing your data, or your data has not yet been officially deposited in a repository, the journal may accept a placeholder statement such as the one you see on the slide: "Data associated with this manuscript will be available in the [repository name] at [DOI]."



DATA
DATA REPOSITORY
submissions welcome

Where to Share

- Journal website
  - Within the manuscript
  - Supplemental files
- Repository
- Recommendation: “open repository or journal”

Credit: Ainsley Seago. https://doi.org/10.1371/journal.pbio.1001779.g001. CC BY 4.0

Van Tuyl S, Whitmire AL (2016) Water, Water, Everywhere: Defining and Assessing Data Sharing in Academia. *PLoS ONE* 11(2): e0147942. https://doi.org/10.1371/journal.pone.0147942

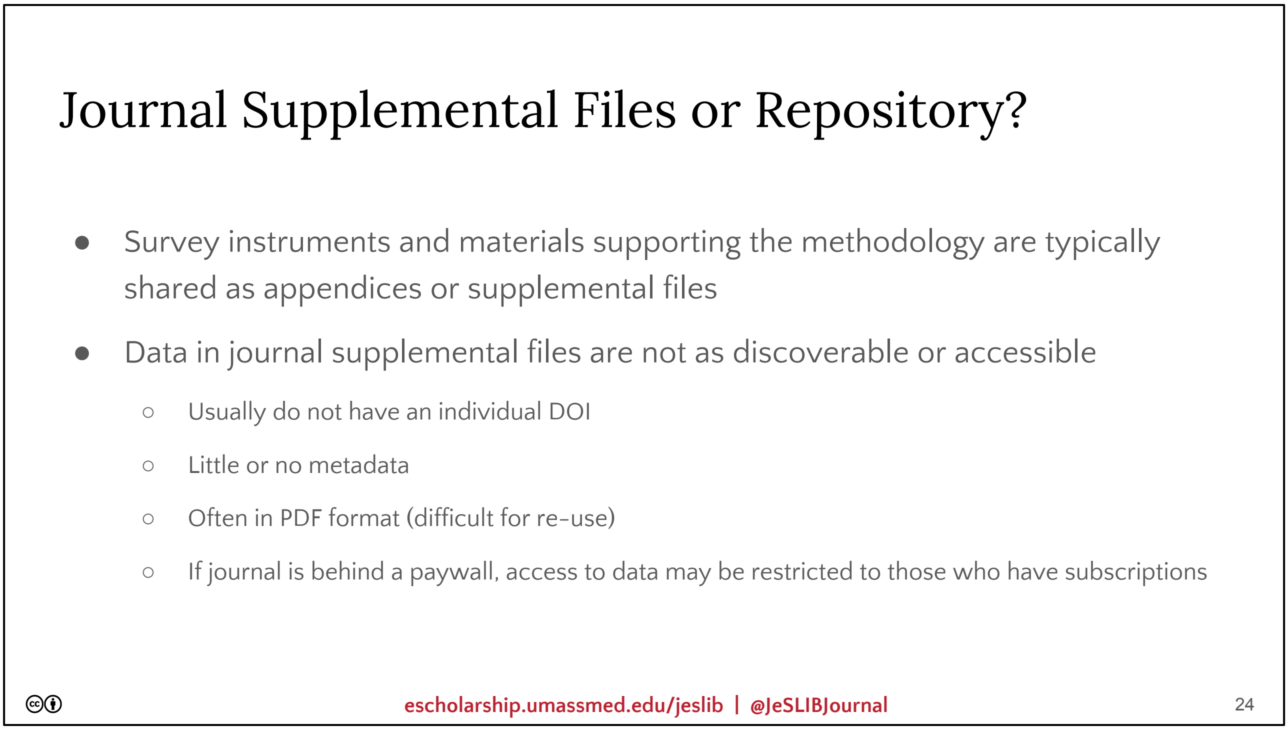
Data Availability: All raw and processed data for this paper are shared at ScholarsArchive@OSU - Oregon State University's repository for scholarly materials. Data may be accessed at: http://dx.doi.org/10.7267/N9W66HPQ.

Most data sharing policies require that you share your data in an appropriate public repository, unless it is already available in the submitted article.

We recommend sharing through an open repository or open access journal, for maximum access to the data and for ease of compliance with data sharing requirements. Non-repository websites -- such as departmental, lab, or personal web pages -- don’t offer the same level of visibility and aren’t as likely to sustain long-term access.

I’ll quote a sentence from the study cited on the slide, which evaluated data sharing practices of researchers at their institution: “Many of the Accessibility issues we experienced when trying to find data for this project were related to the content being shared through a closed access journal or repository.”

Van Tuyl S, Whitmire AL (2016) Water, Water, Everywhere: Defining and Assessing Data Sharing in Academia. *PLoS ONE* 11(2): e0147942. https://doi.org/10.1371/journal.pone.0147942



Journal Supplemental Files or Repository?

- Survey instruments and materials supporting the methodology are typically shared as appendices or supplemental files
- Data in journal supplemental files are not as discoverable or accessible
  - Usually do not have an individual DOI
  - Little or no metadata
  - Often in PDF format (difficult for re-use)
  - If journal is behind a paywall, access to data may be restricted to those who have subscriptions

Currently the most popular way to share data is within the journal article and supplemental files, but this is starting to change as more data repositories have become available, and journals recommend using them. For most journal data policies, the journal editor has the final decision as to where data is made available, following the norms the journal has established.

Materials supporting the methodology -- such as survey instruments, questionnaires, rubrics, assessment instruments, scales, and figures -- are typically shared as appendices or supplemental files on the journal website. This is reflected in Journal of the Medical Library Association's policy, as Scott explained earlier.

It's important to point out that journal supplemental files often do not have their own DOIs or much metadata, making them more difficult to find and cite. This is especially true if you are publishing in a subscription journal, where access to data may be restricted and closed to those without subscriptions.

Repository Options

- Disciplinary or domain
- Institutional
- General (e.g. Figshare, Open Science Framework, Zenodo, Dryad, Harvard Dataverse, OpenICPSR)

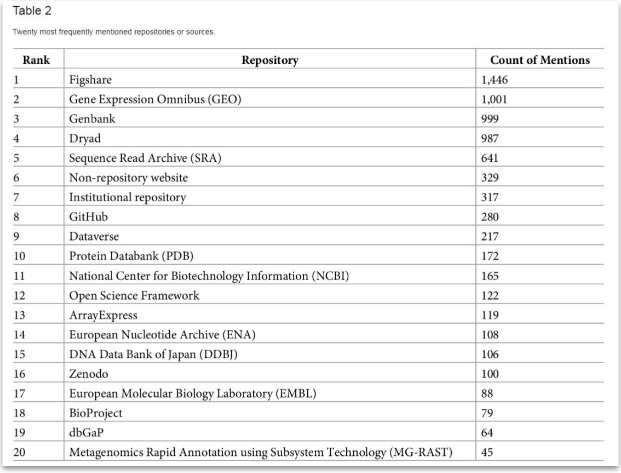


Table 2

Twenty most frequently mentioned repositories or sources.

Rank	Repository	Count of Mentions
1	Figshare	1,446
2	Gene Expression Omnibus (GEO)	1,001
3	Genbank	999
4	Dryad	987
5	Sequence Read Archive (SRA)	641
6	Non-repository website	329
7	Institutional repository	317
8	GitHub	280
9	Dataverse	217
10	Protein Databank (PDB)	172
11	National Center for Biotechnology Information (NCBI)	165
12	Open Science Framework	122
13	ArrayExpress	119
14	European Nucleotide Archive (ENA)	108
15	DNA Data Bank of Japan (DDBJ)	106
16	Zenodo	100
17	European Molecular Biology Laboratory (EMBL)	88
18	BioProject	79
19	dbGaP	64
20	Metagenomics Rapid Annotation using Subsystem Technology (MG-RAST)	45

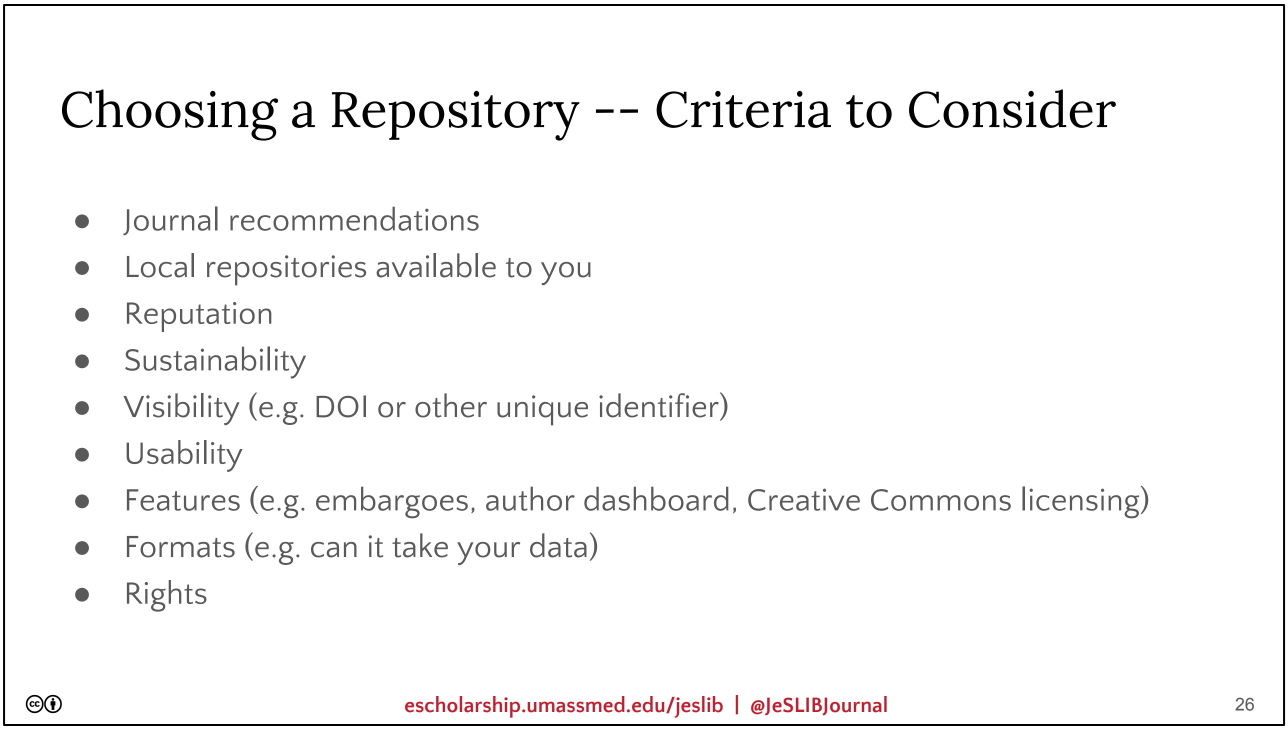
Federer LM, Belter CW, Joubert DJ, Livinski A, Lu Y-L, Snyders LN, et al. 2018. "Data sharing in PLOS ONE: An analysis of Data Availability Statements." *PLoS ONE* 13(5): e0194768. https://doi.org/10.1371/journal.pone.0194768

Data Availability: The final dataset and accompanying code are available on the Open Science Framework, https://doi.org/10.17605/OSF.IO/UN8JX

As a general rule, a domain repository is likely to offer the best home for data than can be publicly shared, because data is available to your peers in a known, authoritative place. For library science, there really isn't a disciplinary data repository at this point, although your data may be appropriate for one in another discipline. But institutional or general repositories are great solutions for library research data.

The table on the slide is from a paper by Lisa Federer and her colleagues that analyzed the Data Availability Statements from more than 47K papers published in PLOS ONE between March 2014 and May 2016. This table lists the 20 most frequently mentioned repositories or sources. Many of these sources are ones that librarians might utilize -- and are utilizing -- such as Figshare, Dataverse, the Open Science Framework, Zenodo, and of course, institutional repositories.

Federer LM, Belter CW, Joubert DJ, Livinski A, Lu Y-L, Snyders LN, et al. 2018. "Data sharing in PLOS ONE: An analysis of Data Availability Statements." *PLoS ONE* 13(5): e0194768. https://doi.org/10.1371/journal.pone.0194768



Choosing a Repository -- Criteria to Consider

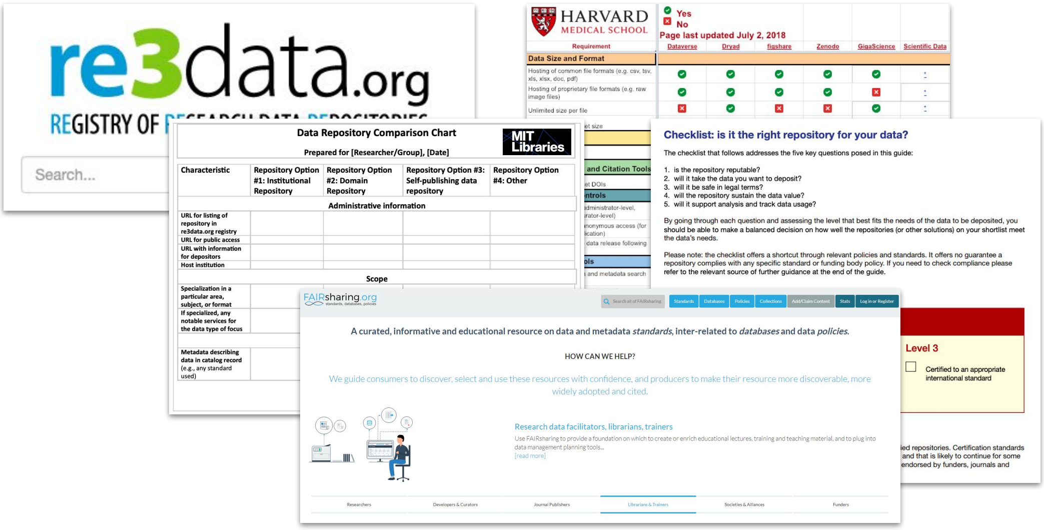
- Journal recommendations
- Local repositories available to you
- Reputation
- Sustainability
- Visibility (e.g. DOI or other unique identifier)
- Usability
- Features (e.g. embargoes, author dashboard, Creative Commons licensing)
- Formats (e.g. can it take your data)
- Rights

How do you choose a data repository? Here are some criteria to consider. Make sure to consult any recommendations that the journal makes for data deposit and look into any local repositories that might be at your disposal, such as at your institution or available through an institutional membership. You want the repository to be reputable and to be backed by an institution, community, or funder. Make sure you can obtain a DOI or other unique identifier, which will make it easier for your data to be discovered and cited. Features to look for include the ability to set an embargo, track usage data, and apply a Creative Commons license. Make sure the repository can accommodate the data you've generated and gives you sufficient rights.

Whyte, A. (2015). 'Where to keep research data: DCC checklist for evaluating data repositories' v.1.1 Edinburgh: Digital Curation Centre. http://www.dcc.ac.uk/resources/how-guides-checklists/where-keep-research-data/where-keep-research-data

Boston University Data Services: What is a Data Repository?
https://www.bu.edu/data/share/selecting-a-data-repository

Choosing a Repository -- Resources



re3data.org

REGISTRY OF RESEARCH DATA REPOSITORIES

Search...

HARVARD MEDICAL SCHOOL

✓ Yes
☒ No
Page last updated July 2, 2018

Requirement	Dataverse	Dryad	figshare	Zenodo	GigaScience	Scientific Data
Data Size and Format						
Hosting of common file formats (e.g. csv, tsv, xls, xlsx, doc, pdf)	✓	✓	✓	✓	✓	*
Hosting of proprietary file formats (e.g. raw image files)	✓	✓	✓	✓	☒	*
Unlimited size per file	☒	✓	☒	☒	✓	*

et size

and Citation Tools

et DOIs

ntrols

dministrator-level, rator-level)

nonymous access (for ication)

data release following

ols

and metadata search

Data Repository Comparison Chart

Prepared for [Researcher/Group], [Date]

MIT Libraries

Characteristic	Repository Option #1: Institutional Repository	Repository Option #2: Domain Repository	Repository Option #3: Self-publishing data repository	Repository Option #4: Other
Administrative information				
URL for listing of repository in re3data.org registry				
URL for public access				
URL with information for depositors				
Host institution				
Scope				
Specialization in a particular area, subject, or format				
If specialized, any notable services for the data type of focus				
Metadata describing data in catalog record (e.g., any standard used)				

Checklist: is it the right repository for your data?

The checklist that follows addresses the five key questions posed in this guide:

1. is the repository reputable?
2. will it take the data you want to deposit?
3. will it be safe in legal terms?
4. will the repository sustain the data value?
5. will it support analysis and track data usage?

By going through each question and assessing the level that best fits the needs of the data to be deposited, you should be able to make a balanced decision on how well the repositories (or other solutions) on your shortlist meet the data's needs.

Please note: the checklist offers a shortcut through relevant policies and standards. It offers no guarantee a repository complies with any specific standard or funding body policy. If you need to check compliance please refer to the relevant source of further guidance at the end of the guide.

Level 3

☐ Certified to an appropriate international standard

ied repositories. Certification standards and that is likely to continue for some endorsed by funders, journals and

FAIRsharing.org standards, databases, policies

A curated, informative and educational resource on data and metadata *standards*, inter-related to *databases* and data *policies*.

HOW CAN WE HELP?

We guide consumers to discover, select and use these resources with confidence, and producers to make their resource more discoverable, more widely adopted and cited.

Research data facilitators, librarians, trainers

Use FAIRsharing to provide a foundation on which to create or enrich educational lectures, training and teaching material, and to plug into data management planning tools...
[read more]

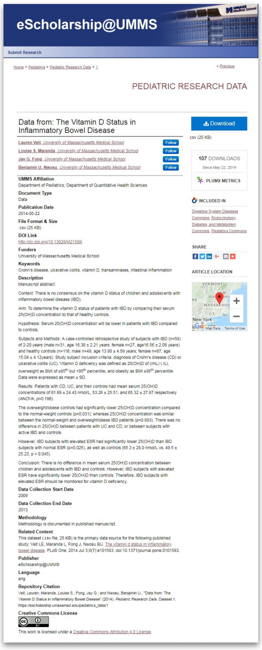
There are several registries and other sources you can consult to help you address those considerations and select an appropriate data repository. R3data.org and FAIRsharing.org help you discover a reputable repository, as does the new Repository Finder tool hosted by DataCite. The Digital Curation Centre has a checklist that addresses important criteria, and MIT provides a template to compare repositories. Harvard Medical School's matrix compares several of the general data repositories.

- Registry of Research Data Repositories: https://www.re3data.org
- FAIRsharing: https://fairsharing.org/
- Repository Finder: https://repositoryfinder.datacite.org/
- MIT Libraries Repository Comparison Template: https://libraries.mit.edu/data-management/share/find-repository
- HMS Biomedical Data Repositories Matrix: https://datamanagement.hms.harvard.edu/repositories
- Whyte, A. (2015). 'Where to keep research data: DCC checklist for evaluating data repositories' v.1.1 Edinburgh: Digital Curation Centre. http://www.dcc.ac.uk/resources/how-guides-checklists/where-keep-research-data/where-keep-research-data

Institutional

eScholarship@UMMS can serve as a home for data files that support scholarly publications, including Journal of eScience Librarianship (JeSLIB) journal articles that must meet requirements for the preservation and dissemination of data.

https://escholarship.umassmed.edu



eScholarship@UMMS

Submit Research

Home > Pediatrics > Pediatric Research Data > 1

< Previous

PEDIATRIC RESEARCH DATA

Data from: The Vitamin D Status in Inflammatory Bowel Disease

Lauren Veit, University of Massachusetts Medical School — Follow
Louise S. Maranda, University of Massachusetts Medical School — Follow
Jay G. Fong, University of Massachusetts Medical School — Follow
Benjamin U. Nwosu, University of Massachusetts Medical School — Follow

UMMS Affiliation
Department of Pediatrics; Department of Quantitative Health Sciences

Document Type
Data

Publication Date
2014-05-22

File Format & Size
.csv (25 KB)

DOI Link
http://dx.doi.org/10.13028/M2159K

Funders
University of Massachusetts Medical School

Keywords
Crohn's disease, ulcerative colitis, vitamin D, transaminases, intestinal inflammation

Description
Manuscript abstract:

Context: There is no consensus on the vitamin D status of children and adolescents with inflammatory bowel disease (IBD).

Aim: To determine the vitamin D status of patients with IBD by comparing their serum 25(OH)D concentration to that of healthy controls.

Hypothesis: Serum 25(OH)D concentration will be lower in patients with IBD compared to controls.

Subjects and Methods: A case-controlled retrospective study of subjects with IBD (n=58) of 2-20 years (male n=31, age 16.38 ± 2.21 years; female n=27, age16.56 ± 2.08 years) and healthy controls (n=116; male n=49, age 13.90 ± 4.59 years; female n=67, age 15.04 ± 4.12years). Study subject inclusion criteria: diagnosis of Crohn's disease (CD) or ulcerative colitis (UC). Vitamin D deficiency was defined as 25(OH)D of (/mL) (/L), overweight as BMI of ≥85th but <95th percentile, and obesity as BMI ≥95th percentile. Data were expressed as mean ± SD.

Results: Patients with CD, UC, and their controls had mean serum 25(OH)D concentrations of 61.69 ± 24.43 nmol/L, 53.26 ± 25.51, and 65.32 ± 27.97 respectively (ANOVA, *p*=0.196).

The overweight/obese controls had significantly lower 25(OH)D concentration compared to the normal-weight controls (p=0.031); whereas 25(OH)D concentration was similar between the normal-weight and overweight/obese IBD patients (p=0.883). There was no difference in 25(OH)D between patients with UC and CD, or between subjects with active IBD and controls.

However, IBD subjects with elevated ESR had significantly lower 25(OH)D than IBD subjects with normal ESR (p=0.025), as well as controls (65.3 ± 28.0 nmol/L vs. 49.5 ± 25.23, p = 0.045).

Conclusion: There is no difference in mean serum 25(OH)D concentration between children and adolescents with IBD and controls. However, IBD subjects with elevated ESR have significantly lower 25(OH)D than controls. Therefore, IBD subjects with elevated ESR should be monitored for vitamin D deficiency.

Data Collection Start Date
2009

Data Collection End Date
2013

Methodology
Methodology is documented in published manuscript.

Related Content
This dataset (.csv file, 25 KB) is the primary data source for the following published study: Veit LE, Maranda L, Fong J, Nwosu BU. The vitamin d status in inflammatory bowel disease. PLoS One. 2014 Jul 3;9(7):e101583. doi:10.1371/journal.pone.0101583.

Publisher
eScholarship@UMMS

Language
eng

Repository Citation
Veit, Lauren; Maranda, Louise S.; Fong, Jay G.; and Nwosu, Benjamin U., "Data from: The Vitamin D Status in Inflammatory Bowel Disease" (2014). *Pediatric Research Data*. Dataset 1. https://escholarship.umassmed.edu/pediatrics_data/1

Creative Commons License

This work is licensed under a Creative Commons Attribution 4.0 License.

Download

.csv (25 KB)

107 DOWNLOADS
Since May 22, 2014

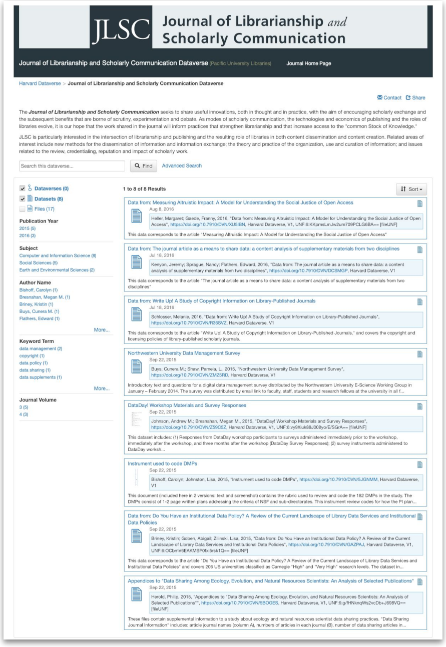
PLUMX METRICS

INCLUDED IN

Digestive System Diseases Commons, Endocrinology, Diabetes, and Metabolism Commons, Pediatrics Commons

SHARE

ARTICLE LOCATION



JLSC Journal of Librarianship *and* Scholarly Communication

Journal of Librarianship and Scholarly Communication Dataverse (Pacific University Libraries) Journal Home Page

Harvard Dataverse > Journal of Librarianship and Scholarly Communication Dataverse

Contact Share

The ***Journal of Librarianship and Scholarly Communication*** seeks to share useful innovations, both in thought and in practice, with the aim of encouraging scholarly exchange and the subsequent benefits that are borne of scrutiny, experimentation and debate. As modes of scholarly communication, the technologies and economics of publishing and the roles of libraries evolve, it is our hope that the work shared in the journal will inform practices that strengthen librarianship and that increase access to the "common Stock of Knowledge."

JLSC is particularly interested in the intersection of librarianship and publishing and the resulting role of libraries in both content dissemination and content creation. Related areas of interest include new methods for the dissemination of information and information exchange; the theory and practice of the organization, use and curation of information; and issues related to the review, credentialing, reputation and impact of scholarly work.

Search this dataverse... Find Advanced Search

☑ Dataverses (0)
☑ Datasets (8)
□ Files (17)

Publication Year
2015 (5)
2016 (3)

Subject
Computer and Information Science (8)
Social Sciences (5)
Earth and Environmental Sciences (2)

Author Name
Bishoff, Carolyn (1)
Bresnahan, Megan M. (1)
Briney, Kristin (1)
Buys, Cunera M. (1)
Flathers, Edward (1)
More...

Keyword Term
data management (2)
copyright (1)
data policy (1)
data sharing (1)
data supplements (1)
More...

Journal Volume
3 (5)
4 (3)

1 to 8 of 8 Results

Sort

Data from: Measuring Altruistic Impact: A Model for Understanding the Social Justice of Open Access
Aug 8, 2016
Heller, Margaret; Gaede, Franny, 2016, "Data from: Measuring Altruistic Impact: A Model for Understanding the Social Justice of Open Access", https://doi.org/10.7910/DVN/XU5IBN, Harvard Dataverse, V1, UNF:6:KKpmsLmJw2um709PCLG6BA== [fileUNF]
This data corresponds to the article "Measuring Altruistic Impact: A Model for Understanding the Social Justice of Open Access"

Data from: The journal article as a means to share data: a content analysis of supplementary materials from two disciplines
Jul 18, 2016
Kenyon, Jeremy; Sprague, Nancy; Flathers, Edward, 2016, "Data from: The journal article as a means to share data: a content analysis of supplementary materials from two disciplines", https://doi.org/10.7910/DVN/DCSMGP, Harvard Dataverse, V1
This data corresponds to the article "The journal article as a means to share data: a content analysis of supplementary materials from two disciplines"

Data from: Write Up! A Study of Copyright Information on Library-Published Journals
Jul 18, 2016
Schlosser, Melanie, 2016, "Data from: Write Up! A Study of Copyright Information on Library-Published Journals", https://doi.org/10.7910/DVN/R36SVZ, Harvard Dataverse, V1
This data corresponds to the article "Write Up! A Studiy of Copyright Information on Library-Published Journals," and covers the copyright and licensing policies of library-published scholarly journals.

Northwestern University Data Management Survey
Sep 22, 2015
Buys, Cunera M.; Shaw, Pamela, L., 2015, "Northwestern University Data Management Survey", https://doi.org/10.7910/DVN/ZMZ5RD, Harvard Dataverse, V1
Introductory text and questions for a digital data management survey distributed by the Northwestern University E-Science Working Group in January – February 2014. The survey was distributed by email link to faculty, staff, students and research fellows at the university in all f...

DataDay! Workshop Materials and Survey Responses
Sep 22, 2015
Johnson, Andrew M.; Bresnahan, Megan M., 2015, "DataDay! Workshop Materials and Survey Responses", https://doi.org/10.7910/DVN/Z59C5Z, Harvard Dataverse, V1, UNF:6:vy9Xiuk88J008yo/E/SGrA== [fileUNF]
This dataset includes: (1) Responses from DataDay workshop participants to surveys administered immediately prior to the workshop, immediately after the workshop, and three months after the workshop (DataDay Survey Responses); (2) survey instruments administered to DataDay worksh...

Instrument used to code DMPs
Sep 22, 2015
Bishoff, Carolyn; Johnston, Lisa, 2015, "Instrument used to code DMPs", https://doi.org/10.7910/DVN/5JGNMM, Harvard Dataverse, V1
This document (included here in 2 versions: text and screenshot) contains the rubric used to review and code the 182 DMPs in the study. The DMPs consist of 1-2 page written plans addressing the criteria of NSF and sub-directorates. This instrument review codes for how the PI plan...

Data from: Do You Have an Institutional Data Policy? A Review of the Current Landscape of Library Data Services and Institutional Data Policies
Sep 22, 2015
Briney, Kristin; Goben, Abigail; Zilinski, Lisa, 2015, "Data from: Do You Have an Institutional Data Policy? A Review of the Current Landscape of Library Data Services and Institutional Data Policies", https://doi.org/10.7910/DVN/GAZPAJ, Harvard Dataverse, V1, UNF:6:OCbmV6EAKMSP0fxi5rsk1Q== [fileUNF]
This data corresponds to the article "Do You Have an Institutional Data Policy? A Review of the Current Landscape of Library Data Services and Institutional Data Policies" and covers 206 US universities classified as Carnegie "High" and "Very High" research levels. The dataset in...

Appendices to "Data Sharing Among Ecology, Evolution, and Natural Resources Scientists: An Analysis of Selected Publications"
Sep 22, 2015
Herold, Philip, 2015, "Appendices to "Data Sharing Among Ecology, Evolution, and Natural Resources Scientists: An Analysis of Selected Publications"", https://doi.org/10.7910/DVN/5BOGES, Harvard Dataverse, V1, UNF:6:g/fHNknqWs2vcDb+J698VQ== [fileUNF]
These files contain supplemental information to a study about ecology and natural resources scientist data sharing practices. "Data Sharing Journal Information" includes: article journal names (column A), numbers of articles in each journal (B), number of data sharing articles in...

General

Journal of Librarianship and Scholarly Communication (JLSC) does not require data publication, however does encourage authors who wish to share their data as a supplement to their articles to either deposit their data in an external repository or in the JLSC Dataverse.

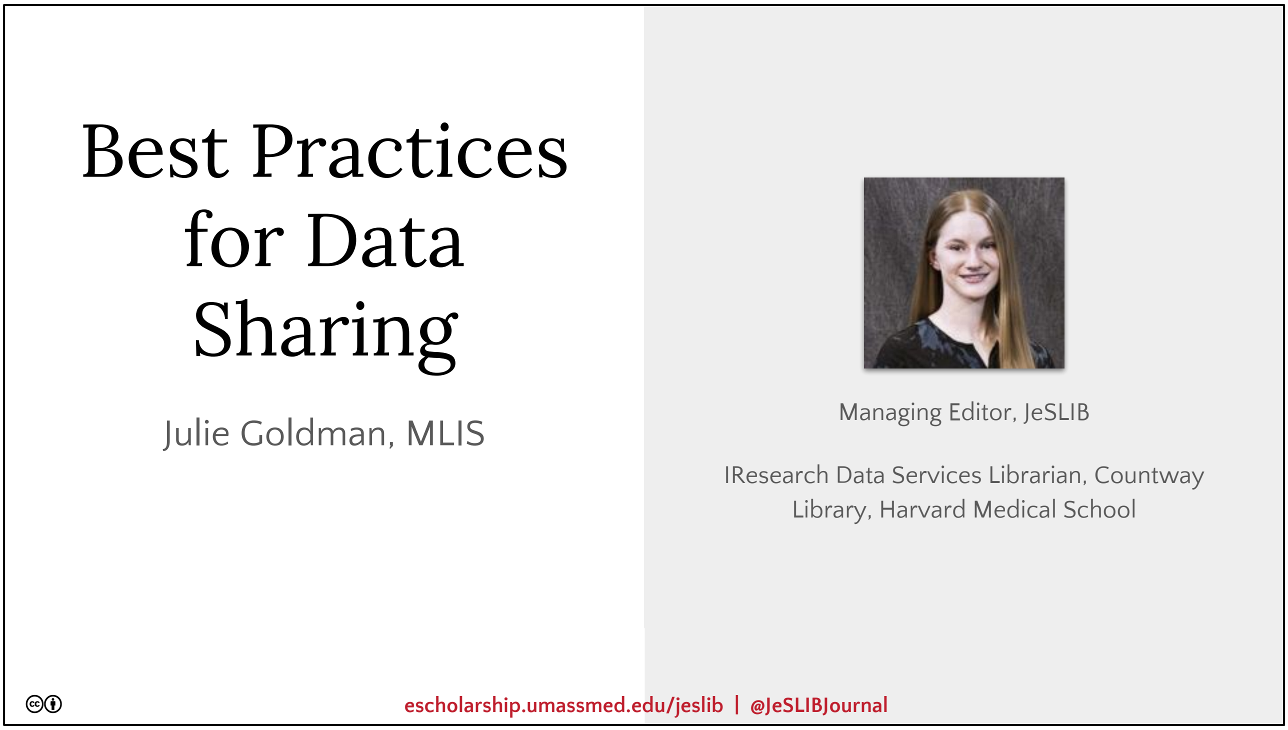
https://dataverse.harvard.edu/dataverse/jlsc

A journal may also make a data repository available to its authors. For example, the Journal of Librarianship and Scholarly Communication has its own instance of Dataverse. The Journal of eScience Librarianship's data policy is still evolving, but because the journal platform is also a repository, we plan to offer authors the option to deposit their data with the journal. Benefits include the ability to provide a DOI for the dataset (separate from the article), Creative Commons licenses, usage metrics, and embargoes.

Now I'll turn the presentation over to Julie Goldman to highlight some best practices.

Open data in eScholarship@UMMS:
https://libraryguides.umassmed.edu/openaccess/open_data#s-lg-box-10569427

Journal of Librarianship and Scholarly Communication data sharing policy:
https://jlsc-pub.org/about/editorialpolicies/#data-sharing
Author guidelines: https://jlsc-pub.org/about/submissions/#data



Best Practices for Data Sharing

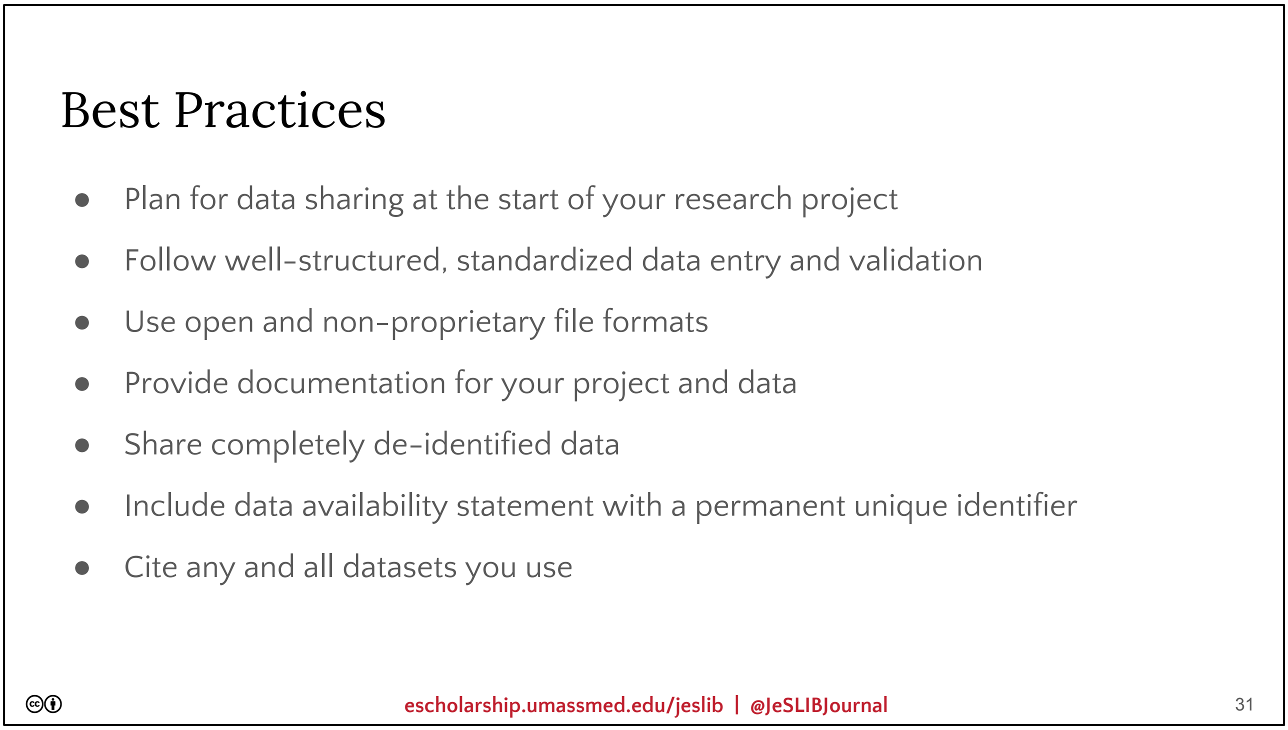
Julie Goldman, MLIS

Managing Editor, JeSLIB

IResearch Data Services Librarian, Countway Library, Harvard Medical School

escholarship.umassmed.edu/jeslib | @JeSLIBJournal

Best Practices for Data Sharing: Julie Goldman



Best Practices

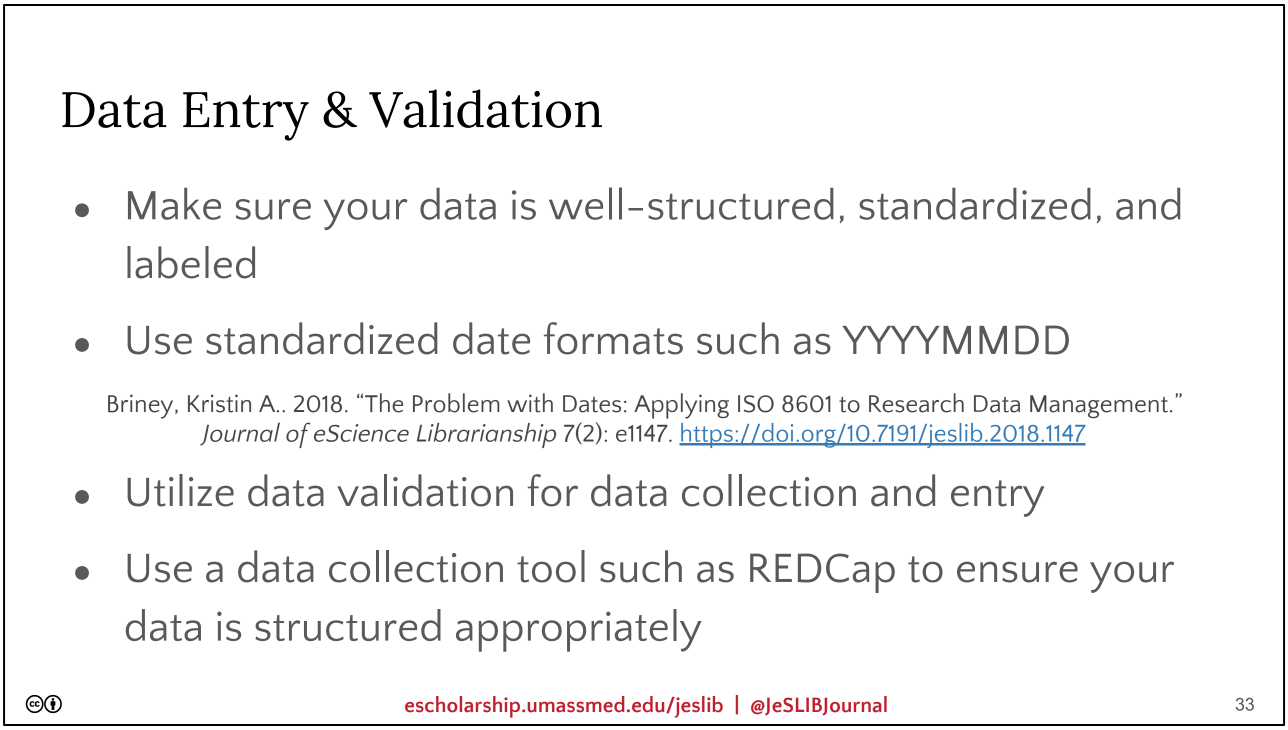
- Plan for data sharing at the start of your research project
- Follow well-structured, standardized data entry and validation
- Use open and non-proprietary file formats
- Provide documentation for your project and data
- Share completely de-identified data
- Include data availability statement with a permanent unique identifier
- Cite any and all datasets you use

So now that we have looked at the what, when and where to share, let's discuss how to approach data sharing. Here are some best practices for preparing your data for sharing. These are just highlights, and we are taking a “practice what we preach” mentality.

- Again, we want to stress that planning ahead from the start of your research project helps in meeting these data sharing challenges. You should include what data you plan to share in your original research process - it is always a good idea to create a data management plan!
- You should follow well-structured, standardized data entry and validation. Much of the research process is spent on the ‘data wrangling’ stage, but some of it can be prevented with good strategies for data collection up front.
- Use open and non-proprietary file formats. In order to maximize accessibility and ensure long term preservation, we encourage the use of non-proprietary formats. Spreadsheets are preferable to PDF for tabular data, so using comma-delimited text files (.csv) or plain text files (.txt) rather than Excel (.xls, .xlsx) and Word (.doc, .docx) documents are the best choices.
- Provide documentation for your project and data. It is a good practice to begin to document your data at the very beginning of your research project and continue to add information as the project progresses.
- Share completely de-identified data. This refers to the process of removing or

- obscuring any personally identifiable information from individual records in a way that minimizes the risk of unintended disclosure of the identity of individuals and information about them.
- Include a data availability statement with a permanent unique identifier. This is also sometimes referred to as a 'data access statement' iand s crucial in signposting where the data associated with a paper is available, and under what conditions the data can be accessed, including links to the data set.
- Finally, cite any and all datasets you use, including your own. While citation standards for data sets are still evolving, it is only best scholarly practice to cite any and all sources you use, and this includes data. These last two points help enhance discoverability of your and other’s work.

Now we’ll go into more detail on a few of these and share some helpful resources.



Data Entry & Validation

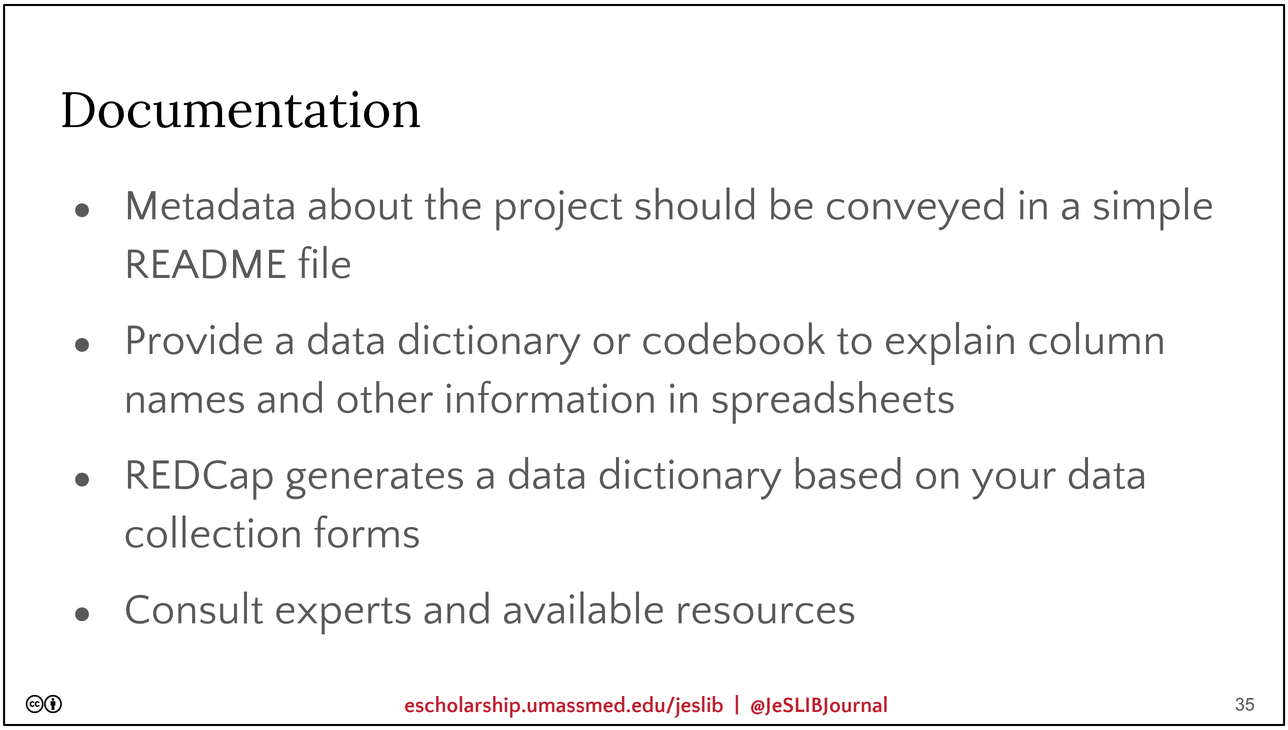
- Make sure your data is well-structured, standardized, and labeled
- Use standardized date formats such as YYYYMMDD

Briney, Kristin A.. 2018. “The Problem with Dates: Applying ISO 8601 to Research Data Management.” *Journal of eScience Librarianship* 7(2): e1147. https://doi.org/10.7191/jeslib.2018.1147

- Utilize data validation for data collection and entry
- Use a data collection tool such as REDCap to ensure your data is structured appropriately

- Structured data has been organized so that its elements can be made accessible for more effective analysis. I highly suggest the Library Carpentry lesson ‘Tidy Data for Librarians’, which provides tips and tricks for working in Excel and OpenRefine, and ensuring you are working with tidy, well-structured and labeled data.
  - Library Carpentry, Tidy Data for Librarians: https://librarycarpentry.org/lc-spreadsheets
- Apply standardized date formats such as ISO 8601. This standard provides needed consistency in date formatting, allows for inclusion of several types of date-time information, and can sort dates chronologically. If you have not read Kristin Briney’s JeSLIB commentary on applying this standard to your everyday work, we highly suggest it. This is a simple action you can start applying to your daily practice today!
  - Briney, Kristin A.. 2018. "The Problem with Dates: Applying ISO 8601 to Research Data Management." *Journal of eScience Librarianship* 7(2): e1147. https://doi.org/10.7191/jeslib.2018.1147
- When you have a well-structured data table, you can use several simple techniques within your spreadsheet to ensure the data you enter is free of errors. These approaches include techniques that are implemented prior to entering data (quality assurance) and techniques that are used after entering data to check for errors (quality control). Again, there are many tips in the

- Library Carpentry lessons you can explore here.
  - Library Carpentry, Tidy Data for Librarians: https://librarycarpentry.org/lc-spreadsheets
- We recommend using REDCap for data collection over other survey collection tools. REDCap is a free, web-based, and user-friendly electronic data capture tools, useful for collecting and tracking information and data from research studies. It really is a data management and survey tool in one, because, just to name a few features, it is fully customizable, has multi site access, data import functions, you can export survey results to common data analysis packages, and is a secure, HIPAA-compliant option.
  - REDCap: https://projectredcap.org



Documentation

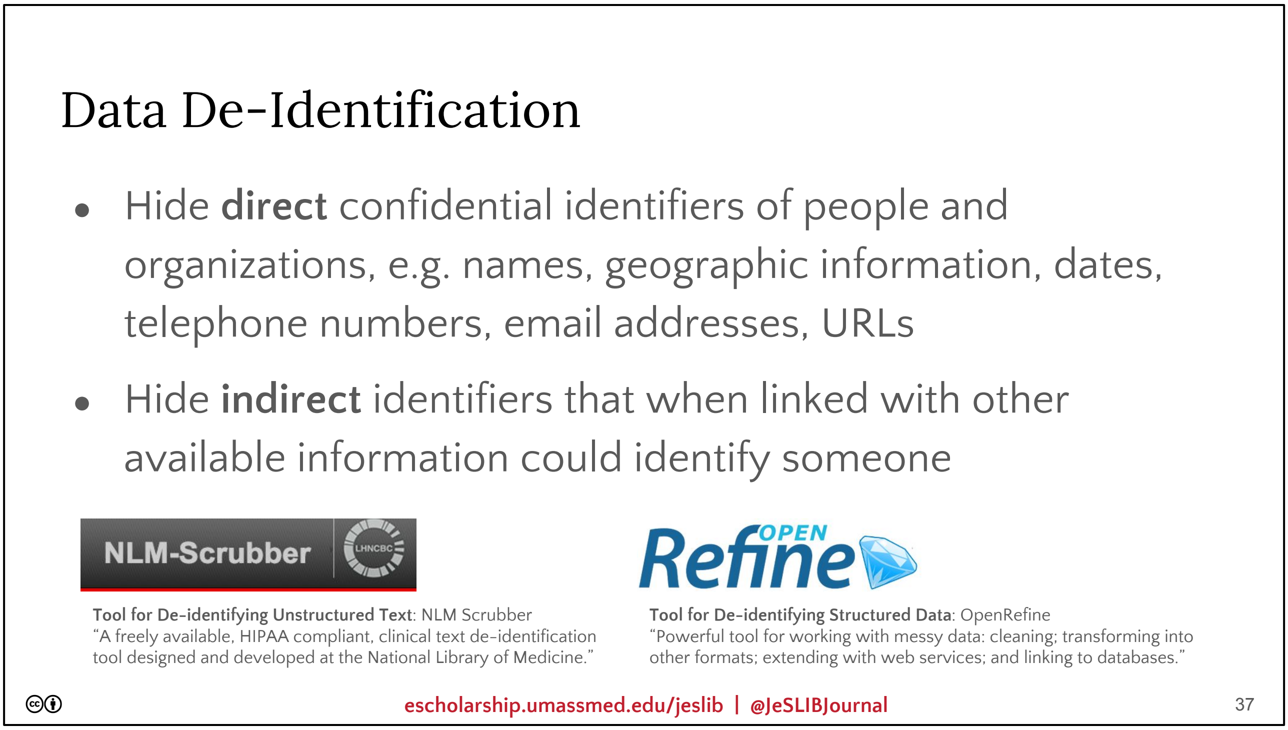
- Metadata about the project should be conveyed in a simple README file
- Provide a data dictionary or codebook to explain column names and other information in spreadsheets
- REDCap generates a data dictionary based on your data collection forms
- Consult experts and available resources

- Data documentation is the key for future understanding of research data. Along with your data, you should provide documentation (“metadata”) that helps you and everyone on your project team understand your data, and also helps other researchers find, use, and properly cite your data. Metadata “completes” a dataset.
  - At a minimum, create a readme text file that includes basic documentation such as title, creator, identifier, rights and access information, dates, location, methodology, etc.
  - Ideally, you want to use standard terminology to enable others to find and use your data. Identify keywords and ontologies for your area of research, if they are available.
- In REDCap, a Data Dictionary is created in CSV format representing the structure of your database. It contains the metadata used to construct your data collection instruments. In addition a human-readable version is accessible, known as the "Codebook", which serves as a quick reference that lets you view the attributes of any given field in the project without having to download and interpret the Data Dictionary. Therefore eliminating steps for you to create these items yourself!
- We acknowledge that we have already covered a lot, and even following the minimum standards for sharing can be difficult and time consuming! So obviously we encourage you to consult experts and resources for assistance

- and advice, and we have included many resources throughout the slides and a full list at the end.
    - Van Tuyl S, Whitmire AL (2016) Water, Water, Everywhere: Defining and Assessing Data Sharing in Academia. *PLoS ONE* 11(2): e0147942. https://doi.org/10.1371/journal.pone.0147942
        - "We recognize that data sharing and meeting even the minimum standards for sharing [...] can be difficult and time consuming, and we encourage authors to engage with research data service providers either at the local level (often part of their academic library) or at the national level (e.g. DataONE — www.dataone.org DataQ — www.researchdataq.org etc.) for assistance and advice."

Additional documentation resources:

- Kristin Briney provides an excellent overview of data documentation in her book, which is listed at the end on the Resources slide.
- Guide to writing "readme" style metadata (Cornell University): https://data.research.cornell.edu/content/readme (also available as a pdf (http://data.research.cornell.edu/sites/default/files/SciMD_ReadMe_Guidelines_v4_1_0.pdf) with example readme files) and readme file template (https://cornell.app.box.com/v/ReadmeTemplate)
- DataONE Best Practices Primer: https://www.dataone.org/sites/all/documents/DataONE_BP_Primer_020212.pdf



Data De-Identification

- Hide **direct** confidential identifiers of people and organizations, e.g. names, geographic information, dates, telephone numbers, email addresses, URLs
- Hide **indirect** identifiers that when linked with other available information could identify someone

Tool for De-identifying Unstructured Text: NLM Scrubber
"A freely available, HIPAA compliant, clinical text de-identification tool designed and developed at the National Library of Medicine."

Tool for De-identifying Structured Data: OpenRefine
"Powerful tool for working with messy data: cleaning; transforming into other formats; extending with web services; and linking to databases."

escholarship.umassmed.edu/jeslib | @JeSLIBJournal

37

De-identification of data refers to the process of removing or obscuring any personally identifiable information from individual records in a way that minimizes the risk of unintended disclosure of the identity of individuals and information about them. It is considered successful when there is no reasonable basis to believe that the remaining information in the records can be used to identify an individual.

- You should hide **direct** confidential identifiers of people and organizations, e.g. names, geographic information, dates, telephone numbers, email addresses, URLs — any unique identifying number, characteristic, or code.
- Also, hide indirect **identifiers** that when linked with other available information could identify someone or their institution.
- Approaches include replacing personal information with surrogates that can later be used to look up the real values, dropping the columns or recoding the variables.

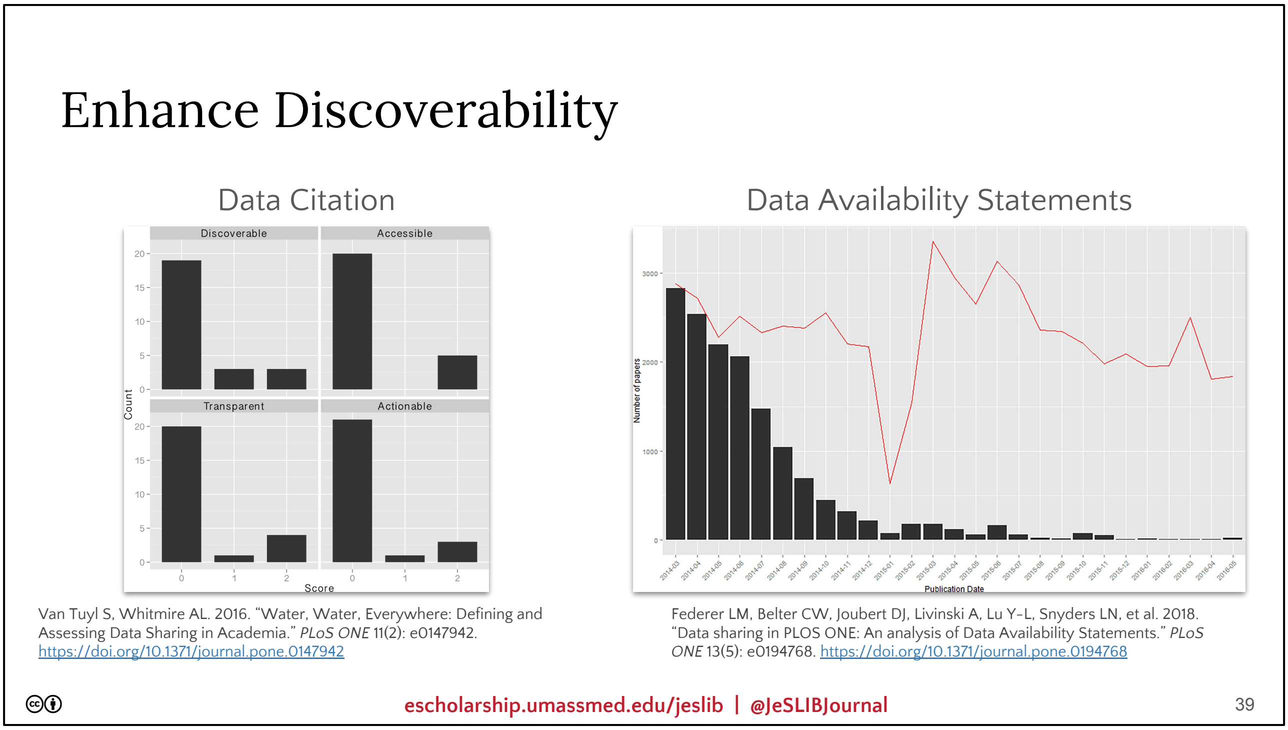
See these two tools as options for de-identifying data:

- NLM-Scrubber for unstructured text - which is out of the National Library of Medicine.
  - NLM-Scrubber: https://scrubber.nlm.nih.gov
- OpenRefine for structured data - which is open source and I highly suggest librarians get familiar with!
  - OpenRefine: http://openrefine.org

I’ll also mention REDCap again as it is more secure option than Microsoft Excel. It is also HIPAA-compliant, as fields in REDCap can be marked as identifiable; and the user has the option of de-identifying their data during export. REDCap also offers daily backups, basic support, and an audit trail feature for even more security.

Additional resources:

- “Preparing data for deposit” from UK Data Archive: https://ukdataservice.ac.uk/deposit-data/preparing-data.aspx and “Depositing shareable survey data” https://ukdataservice.ac.uk/media/440320/depositsurvey.pdf
- “Preparing data for sharing” from ICPSR: https://www.icpsr.umich.edu/icpsrweb/content/deposit/guide/chapter5.html
- De-Identifying Human Subjects Data (Johns Hopkins University): https://jh.app.box.com/v/de-identificationtips
- How to anonymize quantitative data (UK Data Archive): https://www.ukdataservice.ac.uk/manage-data/legal-ethical/anonymisation/quantitative
- How to anonymize qualitative data (UK Data Archive): https://www.ukdataservice.ac.uk/manage-data/legal-ethical/anonymisation/qualitative



Enhance Discoverability

Data Citation

Van Tuyl S, Whitmire AL. 2016. “Water, Water, Everywhere: Defining and Assessing Data Sharing in Academia.” *PLoS ONE* 11(2): e0147942. https://doi.org/10.1371/journal.pone.0147942

Data Availability Statements

Federer LM, Belter CW, Joubert DJ, Livinski A, Lu Y-L, Snyders LN, et al. 2018. “Data sharing in PLOS ONE: An analysis of Data Availability Statements.” *PLoS ONE* 13(5): e0194768. https://doi.org/10.1371/journal.pone.0194768

escholarship.umassmed.edu/jeslib | @JeSLIBJournal

39

Enhancing discoverability to your data involves including a data availability statement with a permanent unique identifier, allowing others to successfully cite your work.

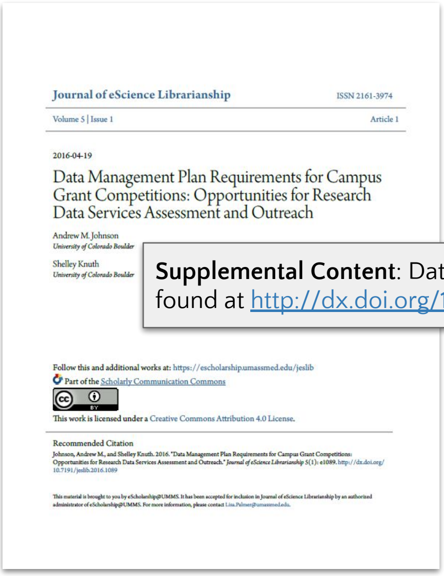
The first example on the slide is from Steven Van Tuyl and Amanda Whitmire’s 2016 evaluation of data sharing practices by OSU researchers funded by the NSF. The figure shows low scores in discoverability and accessibility to data. These low scores stem from issues surrounding standards in data citation styles and methods -- issues in the fact that they varied widely -- making it difficult to find and access data that was meant to be shared.

- Van Tuyl S, Whitmire AL (2016) Water, Water, Everywhere: Defining and Assessing Data Sharing in Academia. *PLoS ONE* 11(2): e0147942. https://doi.org/10.1371/journal.pone.0147942
  - “One of the biggest Discoverability problems we saw with data sharing through journal articles was that citation styles and methods varied widely, making it difficult to find data that was meant to be shared.”
  - Also, the authors followed best practices for data citation by citing their dataset in this paper.

The second example is from Lisa Federer et al.’s 2018 research of data availability statements from over 47,000 papers published in PLOS ONE between March 2014 and May 2016. The figure shows ‘articles missing a Data Availability Statement over time' (the red line indicates total published articles, while the bars indicate articles with no Data Availability Statement). While we can see that once PLOS' data sharing policy was implemented in March 2014 the amount of articles missing a data availability statement decreased, within their study the authors still found some inconsistency with the statements that were included. For example, where a direct URL, DOI, or accession number was not included, basically making it very hard for others to actually directly navigate to or find the data tied to the article. I'll also point out that Lisa's article also has an image which shows sample statements, which might be helpful for people.

- Federer LM, Belter CW, Joubert DJ, Livinski A, Lu Y-L, Snyders LN, et al. (2018) Data sharing in PLOS ONE: An analysis of Data Availability Statements. *PLoS ONE* 13(5): e0194768. https://doi.org/10.1371/journal.pone.0194768
    - "18.2% of PLOS papers named a specific repository or source where data were available. Most Data Accessibility Statements direct the reader to the paper itself or supplementary information. Even among the data repository articles, some Data Accessibility Statements indicated a repository but failed to include a URL, DOI, or accession number — basically sending readers on a wild goose chase to locate their data within the repository."

Data availability statement templates from from Taylor and Francis: https://authorservices.taylorandfrancis.com/data-availability-statement-templates



Journal of eScience Librarianship ISSN 2161-3974

Volume 5 | Issue 1 Article 1

2016-04-19

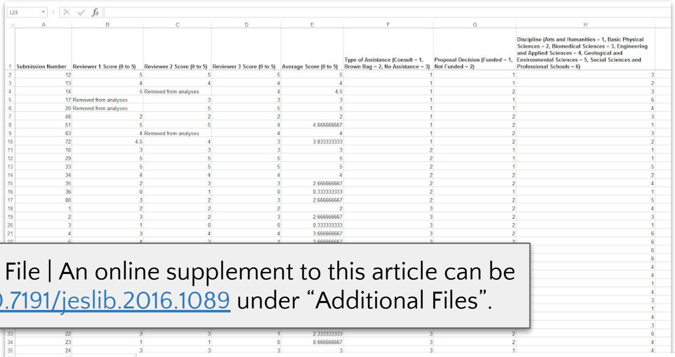
Data Management Plan Requirements for Campus Grant Competitions: Opportunities for Research Data Services Assessment and Outreach

Andrew M. Johnson
University of Colorado Boulder

Shelley Knuth
University of Colorado Boulder

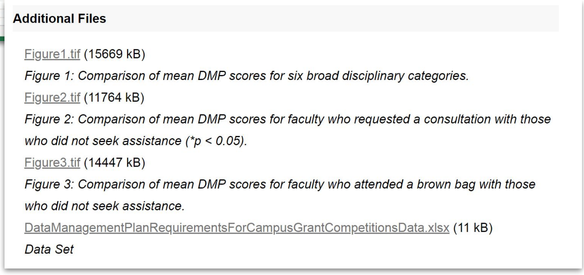
Recommended Citation
Johnson, Andrew M., and Shelley Knuth. 2016. "Data Management Plan Requirements for Campus Grant Competitions: Opportunities for Research Data Services Assessment and Outreach." *Journal of eScience Librarianship* 5(1): e1089. http://dx.doi.org/10.7191/jeslib.2016.1089

Supplemental Content: Data File | An online supplement to this article can be found at http://dx.doi.org/10.7191/jeslib.2016.1089 under “Additional Files”.



L23

	A	B	C	D	E	F	G	H
1	Submission Number	Reviewer 1 Score (0 to 5)	Reviewer 2 Score (0 to 5)	Reviewer 3 Score (0 to 5)	Average Score (0 to 5)	Type of Assistance (Consult = 1, Brown Bag = 2, No Assistance = 3)	Proposal Decision (Funded = 1, Not Funded = 2)	Discipline (Arts and Humanities = 1, Basic Physical Sciences = 2, Biomedical Sciences = 3, Engineering and Applied Sciences = 4, Geological and Environmental Sciences = 5, Social Sciences and Professional Schools = 6)
2	12	5	5	5	5	1	1	3
3	13	4	4	4	4	1	1	2
4	14	5	Removed from analyses	4	4.5	1	2	3
5	17	Removed from analyses	3	3	3	1	1	6
6	20	Removed from analyses	5	5	5	1	1	4
7	48	2	2	2	2	1	2	3
8	51	5	5	4	4.666666667	1	2	1
9	63	4	Removed from analyses	4	4	1	2	3
10	72	4.5	4	3	3.833333333	1	2	2
11	10	3	3	3	3	2	1	1
12	29	5	5	5	5	2	1	1
13	33	5	5	5	5	2	1	5
14	34	4	4	4	4	2	2	2
15	35	2	3	3	2.666666667	2	2	4
16	36	0	1	0	0.333333333	2	1	1
17	80	3	2	3	2.666666667	2	2	5
18	1	2	2	2	2	3	2	4
19	2	3	2	3	2.666666667	3	2	3
20	3	1	0	0	0.333333333	3	2	1
21	4	3	4	4	3.666666667	3	2	6
22	5	4	3	1	2.666666667	3	2	6
23								6
24								6
25								4
26								4
27								1
28								4
29								3
30								1
31								4
32								3
33	22	3	3	1	2.333333333	3	2	6
34	23	1	1	0	0.666666667	3	2	4
35	24	3	3	3	3	3	1	4
36	25							
37	26							
38	27							



Additional Files

Figure1.tif (15669 kB)

Figure 1: Comparison of mean DMP scores for six broad disciplinary categories.

Figure2.tif (11764 kB)

*Figure 2: Comparison of mean DMP scores for faculty who requested a consultation with those who did not seek assistance (*p < 0.05).*

Figure3.tif (14447 kB)

Figure 3: Comparison of mean DMP scores for faculty who attended a brown bag with those who did not seek assistance.

DataManagementPlanRequirementsForCampusGrantCompetitionsData.xlsx (11 kB)

Data Set

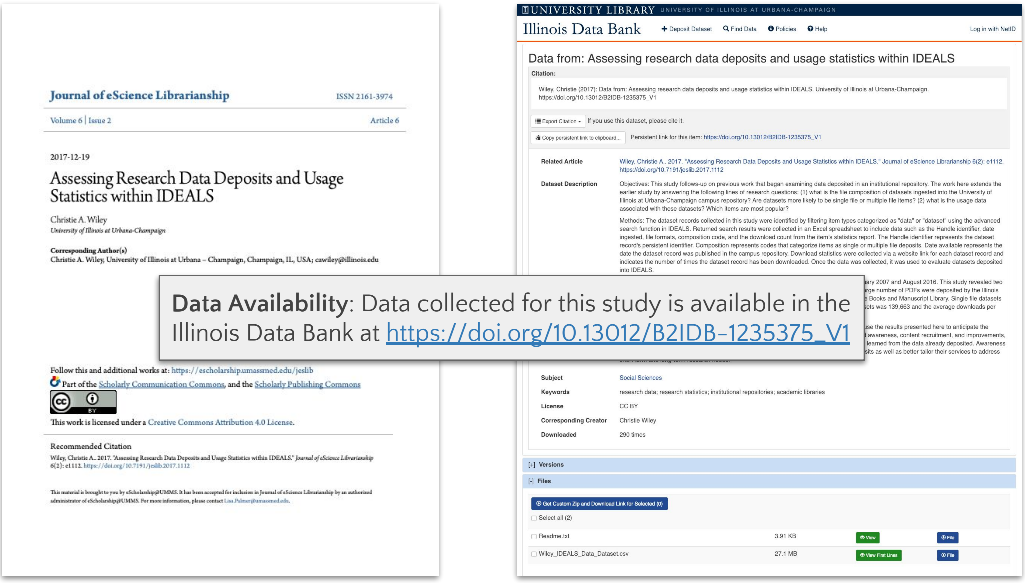
JeSLIB Example: Supplementary Materials

We’ll finish up by exploring some examples of data deposit from JeSLIB & JMLA.

This first example is from JeSLIB where a data file was submitted along with the article as supplemental material. There is no data availability statement, since the data file has been submitted as a supplementary file hosted as part of the article record on our eScholarship@UMMS platform. We are sharing this example, because as as we think about revamping our own data policy at JeSLIB, we probably want to state in our policy that specific file formats should be submitted, such as csv over those proprietary file formats we discussed earlier.

Johnson, Andrew M., and Shelley Knuth. 2016. "Data Management Plan Requirements for Campus Grant Competitions: Opportunities for Research Data Services Assessment and Outreach." Journal of eScience Librarianship 5(1): e1089. http://dx.doi.org/10.7191/jeslib.2016.1089

Supplemental Content: Data File **|** An online supplement to this article can be found at http://dx.doi.org/10.7191/jeslib.2016.1089 under “Additional Files”.



Journal of eScience Librarianship ISSN 2161-3974

Volume 6 | Issue 2 Article 6

2017-12-19

Assessing Research Data Deposits and Usage Statistics within IDEALS

Christie A. Wiley
University of Illinois at Urbana-Champaign

Corresponding Author(s)
Christie A. Wiley, University of Illinois at Urbana – Champaign, Champaign, IL, USA; cawiley@illinois.edu

Follow this and additional works at: https://escholarship.umassmed.edu/jeslib
Part of the Scholarly Communication Commons, and the Scholarly Publishing Commons

This work is licensed under a Creative Commons Attribution 4.0 License.

Recommended Citation
Wiley, Christie A.. 2017. "Assessing Research Data Deposits and Usage Statistics within IDEALS." *Journal of eScience Librarianship* 6(2): e1112. https://doi.org/10.7191/jeslib.2017.1112

This material is brought to you by eScholarship@UMMS. It has been accepted for inclusion in Journal of eScience Librarianship by an authorized administrator of eScholarship@UMMS. For more information, please contact Lisa.Palmer@umassmed.edu.

UNIVERSITY LIBRARY UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

Illinois Data Bank

Data from: Assessing research data deposits and usage statistics within IDEALS

Citation:

Wiley, Christie (2017): Data from: Assessing research data deposits and usage statistics within IDEALS. University of Illinois at Urbana-Champaign. https://doi.org/10.13012/B2IDB-1235375_V1

Export Citation ▾ If you use this dataset, please cite it.

Copy persistent link to clipboard... Persistent link for this item: https://doi.org/10.13012/B2IDB-1235375_V1

Related Article	Wiley, Christie A.. 2017. "Assessing Research Data Deposits and Usage Statistics within IDEALS." Journal of eScience Librarianship 6(2): e1112. https://doi.org/10.7191/jeslib.2017.1112
Dataset Description	Objectives: This study follows-up on previous work that began examining data deposited in an institutional repository. The work here extends the earlier study by answering the following lines of research questions: (1) what is the file composition of datasets ingested into the University of Illinois at Urbana-Champaign campus repository? Are datasets more likely to be single file or multiple file items? (2) what is the usage data associated with these datasets? Which items are most popular? Methods: The dataset records collected in this study were identified by filtering item types categorized as "data" or "dataset" using the advanced search function in IDEALS. Returned search results were collected in an Excel spreadsheet to include data such as the Handle identifier, date ingested, file formats, composition code, and the download count from the item's statistics report. The Handle identifier represents the dataset record's persistent identifier. Composition represents codes that categorize items as single or multiple file deposits. Date available represents the date the dataset record was published in the campus repository. Download statistics were collected via a website link for each dataset record and indicates the number of times the dataset record has been downloaded. Once the data was collected, it was used to evaluate datasets deposited into IDEALS. uary 2007 and August 2016. This study revealed two rge number of PDFs were deposited by the Illinois e Books and Manuscript Library. Single file datasets sets was 139,663 and the average downloads per use the results presented here to anticipate the awareness, content recruitment, and improvements, learned from the data already deposited. Awareness sits as well as better tailor their services to address
Subject	Social Sciences
Keywords	research data; research statistics; institutional repositories; academic libraries
License	CC BY
Corresponding Creator	Christie Wiley
Downloaded	290 times

[+] Versions

[-] Files

Get Custom Zip and Download Link for Selected (0)

□ Select all (2)

□ Readme.txt	3.91 KB	View	File
□ Wiley_IDEALS_Data_Dataset.csv	27.1 MB	View First Lines	File

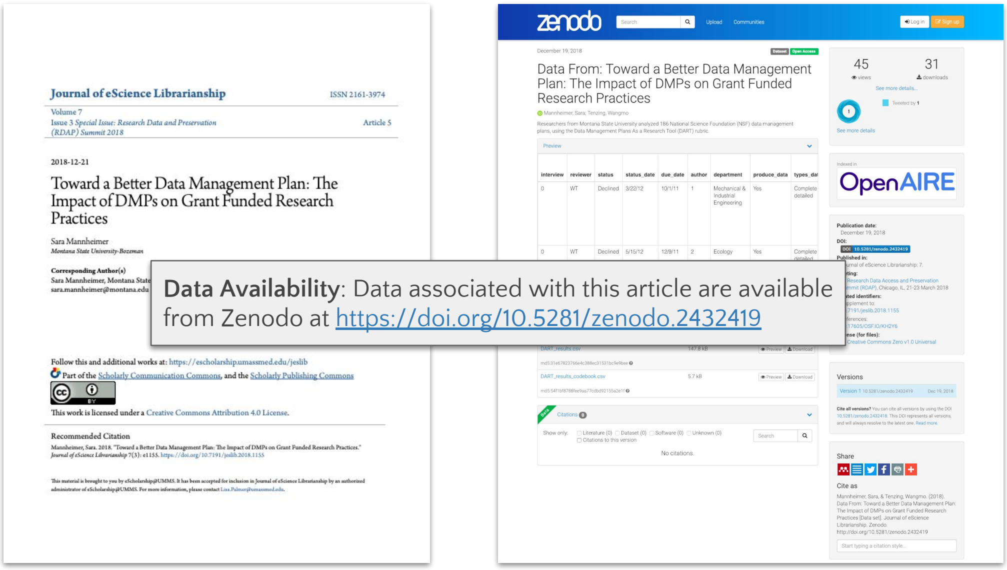
Data Availability: Data collected for this study is available in the Illinois Data Bank at https://doi.org/10.13012/B2IDB-1235375_V1

JeSLIB Example: Data Availability

This is a second example in JeSLIB, where data supporting a submitted article was made available in an institutional repository. So we have included a nice Data Availability statement that points to the DOI of the dataset deposited in the Illinois Data Bank. The statement is found both within the article text, as well as within the metadata on the article landing page.

Wiley, Christie A.. 2017. "Assessing Research Data Deposits and Usage Statistics within IDEALS." *Journal of eScience Librarianship* 6(2): e1112. https://doi.org/10.7191/jeslib.2017.1112

Data Availability: Data collected for this study is available in the Illinois Data Bank at https://doi.org/10.13012/B2IDB-1235375_V1



Journal of eScience Librarianship ISSN 2161-3974

Volume 7
Issue 3 *Special Issue: Research Data and Preservation (RDAP) Summit 2018* Article 5

2018-12-21

Toward a Better Data Management Plan: The Impact of DMPs on Grant Funded Research Practices

Sara Mannheimer
Montana State University-Bozeman

Corresponding Author(s)
Sara Mannheimer, Montana State
sara.mannheimer@montana.edu

Follow this and additional works at: https://escholarship.umassmed.edu/jeslib
Part of the Scholarly Communication Commons, and the Scholarly Publishing Commons

This work is licensed under a Creative Commons Attribution 4.0 License.

Recommended Citation
Mannheimer, Sara. 2018. "Toward a Better Data Management Plan: The Impact of DMPs on Grant Funded Research Practices." *Journal of eScience Librarianship* 7(3): e1155. https://doi.org/10.7191/jeslib.2018.1155

This material is brought to you by eScholarship@UMMS. It has been accepted for inclusion in Journal of eScience Librarianship by an authorized administrator of eScholarship@UMMS. For more information, please contact Lisa.Palmer@umassmed.edu.

December 19, 2018

Data From: Toward a Better Data Management Plan: The Impact of DMPs on Grant Funded Research Practices

Mannheimer, Sara; Tenzing, Wangmo

Researchers from Montana State University analyzed 186 National Science Foundation (NSF) data management plans, using the Data Management Plans As a Research Tool (DART) rubric.

Preview

interview	reviewer	status	status_date	due_date	author	department	produce_data	types_dat
0	WT	Declined	3/22/12	10/1/11	1	Mechanical & Industrial Engineering	Yes	Complete detailed
0	WT	Declined	5/15/12	12/9/11	2	Ecology	Yes	Complete detailed

DART_results.csv 147.8 kB Preview Download
md5:31e67823766e4c388ec31531bc9e9bee

DART_results_codebook.csv 5.7 kB Preview Download
md5:54f1bf8788fee9aa77cdbd92155a2e1f

Citations 0

Show only: ☐ Literature (0) ☐ Dataset (0) ☐ Software (0) ☐ Unknown (0)
☐ Citations to this version

No citations.

45 views 31 downloads
See more details...
Tweeted by 1
See more details

Indexed in

Publication date:
December 19, 2018
DOI:
DOI 10.5281/zenodo.2432419
Published in:
urnal of eScience Librarianship: 7.
ting:
Research Data Access and Preservation
mmit (RDAP), Chicago, IL, 21-23 March 2018
ted identifiers:
pplement to:
.7191/jeslib.2018.1155
ferences:
.17605/OSF.IO/KH2Y6
nse (for files):
Creative Commons Zero v1.0 Universal

Versions
Version 1 10.5281/zenodo.2432419 Dec 19, 2018

Cite all versions? You can cite all versions by using the DOI 10.5281/zenodo.2432418. This DOI represents all versions, and will always resolve to the latest one. Read more.

Share

Cite as
Mannheimer, Sara, & Tenzing, Wangmo. (2018). Data From: Toward a Better Data Management Plan: The Impact of DMPs on Grant Funded Research Practices [Data set]. Journal of eScience Librarianship. Zenodo. http://doi.org/10.5281/zenodo.2432419

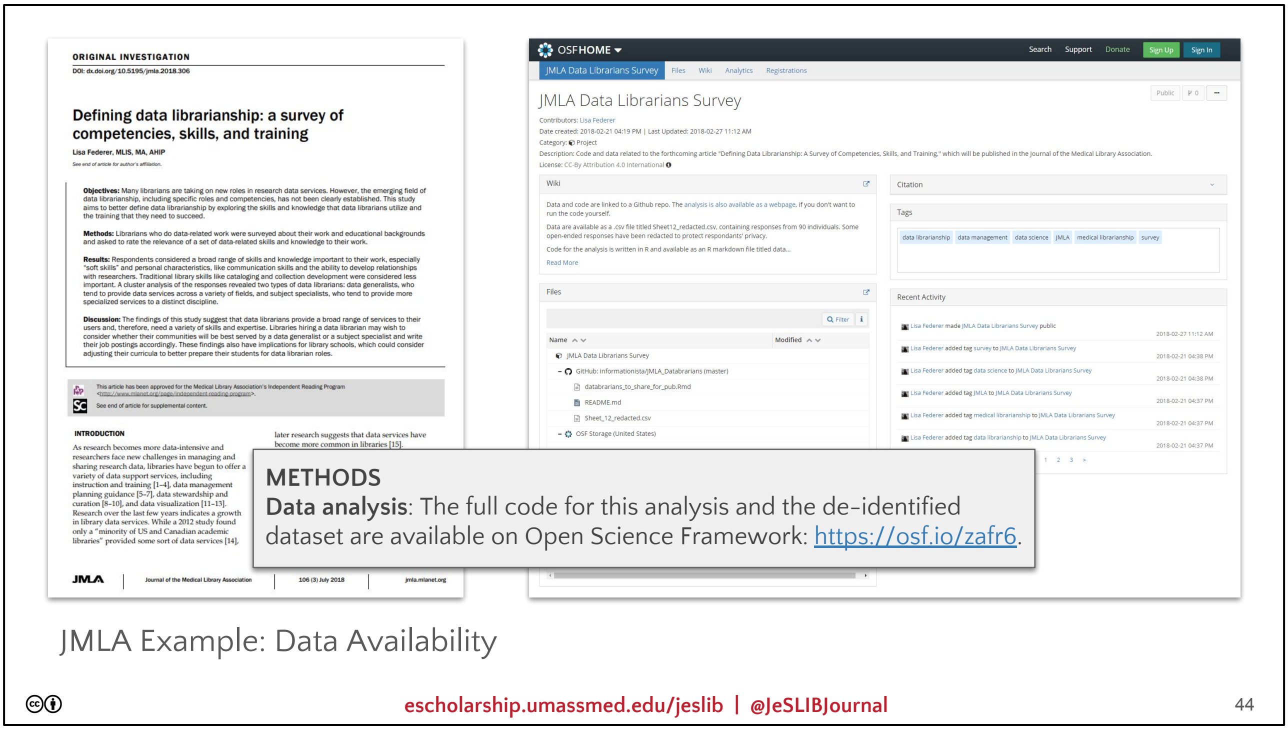
Data Availability: Data associated with this article are available from Zenodo at https://doi.org/10.5281/zenodo.2432419

JeSLIB Example: Data Availability

The most recent example of data deposit we can share from JeSLIB, shows where data supporting a submitted article was made available in the general repository, Zenodo. Again, we have included a nice Data Availability statement that points to the DOI of the dataset in Zenodo. The author included a data citation for this dataset in the manuscript.

Mannheimer, Sara. 2018. "Toward a Better Data Management Plan: The Impact of DMPs on Grant Funded Research Practices." *Journal of eScience Librarianship* 7(3): e1155. https://doi.org/10.7191/jeslib.2018.1155

Data Availability: Data associated with this article are available from Zenodo at https://doi.org/10.5281/zenodo.2432419



ORIGINAL INVESTIGATION

DOI: dx.doi.org/10.5195/jmla.2018.306

Defining data librarianship: a survey of competencies, skills, and training

Lisa Federer, MLIS, MA, AHIP

See end of article for author's affiliation.

Objectives: Many librarians are taking on new roles in research data services. However, the emerging field of data librarianship, including specific roles and competencies, has not been clearly established. This study aims to better define data librarianship by exploring the skills and knowledge that data librarians utilize and the training that they need to succeed.

Methods: Librarians who do data-related work were surveyed about their work and educational backgrounds and asked to rate the relevance of a set of data-related skills and knowledge to their work.

Results: Respondents considered a broad range of skills and knowledge important to their work, especially "soft skills" and personal characteristics, like communication skills and the ability to develop relationships with researchers. Traditional library skills like cataloging and collection development were considered less important. A cluster analysis of the responses revealed two types of data librarians: data generalists, who tend to provide data services across a variety of fields, and subject specialists, who tend to provide more specialized services to a distinct discipline.

Discussion: The findings of this study suggest that data librarians provide a broad range of services to their users and, therefore, need a variety of skills and expertise. Libraries hiring a data librarian may wish to consider whether their communities will be best served by a data generalist or a subject specialist and write their job postings accordingly. These findings also have implications for library schools, which could consider adjusting their curricula to better prepare their students for data librarian roles.

This article has been approved for the Medical Library Association's Independent Reading Program <http://www.mlanet.org/page/independent-reading-program>.

See end of article for supplemental content.

INTRODUCTION

As research becomes more data-intensive and researchers face new challenges in managing and sharing research data, libraries have begun to offer a variety of data support services, including instruction and training [1–4], data management planning guidance [5–7], data stewardship and curation [8–10], and data visualization [11–13]. Research over the last few years indicates a growth in library data services. While a 2012 study found only a "minority of US and Canadian academic libraries" provided some sort of data services [14], later research suggests that data services have become more common in libraries [15].

JMLA | Journal of the Medical Library Association | 106 (3) July 2018 | jmla.mlanet.org

JMLA Data Librarians Survey

Contributors: Lisa Federer

Date created: 2018-02-21 04:19 PM | Last Updated: 2018-02-27 11:12 AM

Category: Project

Description: Code and data related to the forthcoming article "Defining Data Librarianship: A Survey of Competencies, Skills, and Training," which will be published in the Journal of the Medical Library Association.

License: CC-By Attribution 4.0 International

Wiki

Data and code are linked to a Github repo. The analysis is also available as a webpage, if you don't want to run the code yourself.

Data are available as a .csv file titled Sheet12_redacted.csv, containing responses from 90 individuals. Some open-ended responses have been redacted to protect respondants' privacy.

Code for the analysis is written in R and available as an R markdown file titled data...

Read More

Files

Name	Modified
JMLA Data Librarians Survey	
− GitHub: informationista/JMLA_Databrarians (master)	
databrarians_to_share_for_pub.Rmd	
README.md	
Sheet_12_redacted.csv	
− OSF Storage (United States)	

Citation

Tags

data librarianship | data management | data science | JMLA | medical librarianship | survey

Recent Activity

Lisa Federer made JMLA Data Librarians Survey public	2018-02-27 11:12 AM
Lisa Federer added tag survey to JMLA Data Librarians Survey	2018-02-21 04:38 PM
Lisa Federer added tag data science to JMLA Data Librarians Survey	2018-02-21 04:38 PM
Lisa Federer added tag JMLA to JMLA Data Librarians Survey	2018-02-21 04:37 PM
Lisa Federer added tag medical librarianship to JMLA Data Librarians Survey	2018-02-21 04:37 PM
Lisa Federer added tag data librarianship to JMLA Data Librarians Survey	2018-02-21 04:37 PM

1 2 3 >

METHODS

Data analysis: The full code for this analysis and the de-identified dataset are available on Open Science Framework: https://osf.io/zafr6.

JMLA Example: Data Availability

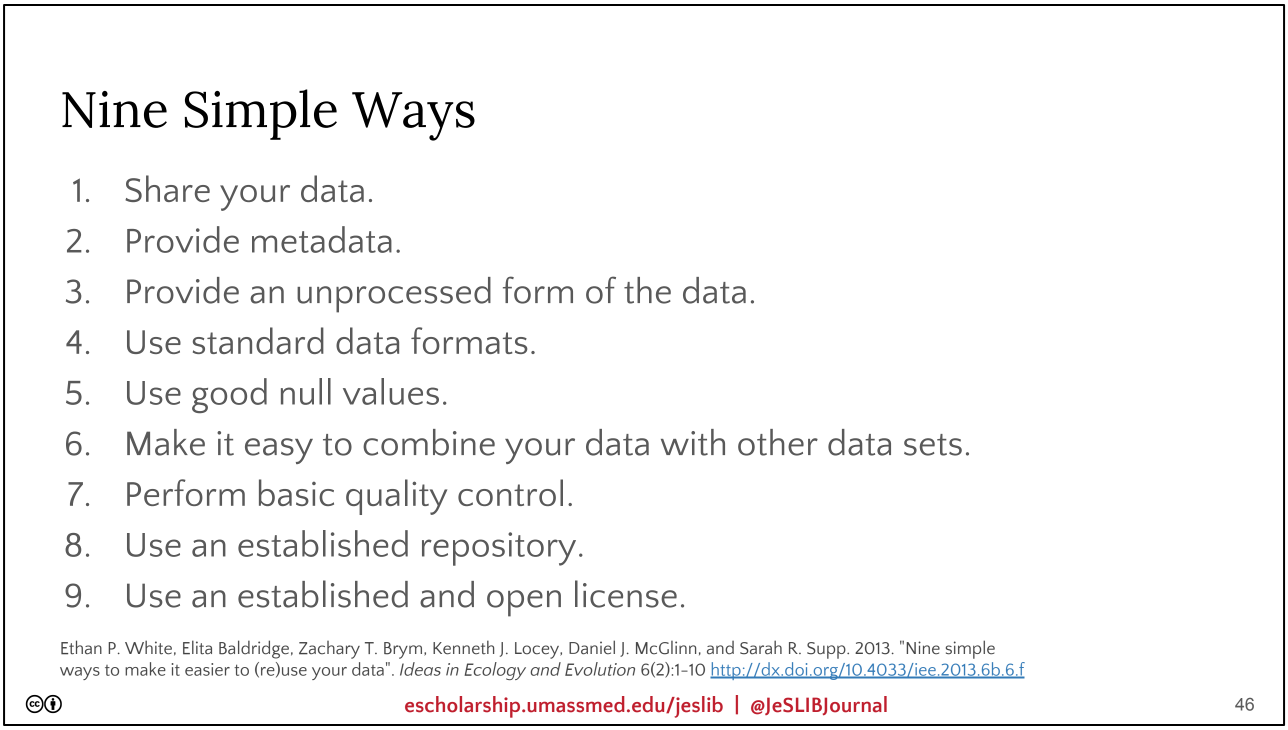
This final example is from JMLA where data, including code and analysis, have been made available via the Open Science Framework. We note here that no 'Data Availability' statement is provided, instead a sentence is included in the 'Data analysis' section of the Methods, pointing to the globally unique identifier (GUID) for that OSF project. This is because the JMLA Data Policy has not gone into effect yet; if this article were to be published come October, 2019, as Scott described earlier, a structured 'Data Availability' statement would be included.

Federer L. Defining data librarianship: a survey of competencies, skills, and training. J Med Libr Assoc. 2018 Jul;106(3):294-303. doi: 10.5195/jmla.2018.306. Epub 2018 Jul 1. PubMed PMID: 29962907; PubMed Central PMCID: PMC6013124. https://doi.org/10.5195/jmla.2018.306

METHODS

Data analysis: The full code for this analysis and the de-identified dataset are available on Open Science Framework: https://osf.io/zafr6.

- Dataset includes: README file; Data, in csv file (illustrates how you can make your survey data available and also redact survey comments to protect privacy of respondents); R code used for analysis; article also has 2 PDF appendices as supplementary materials (the survey instrument and the taxonomy)



Nine Simple Ways

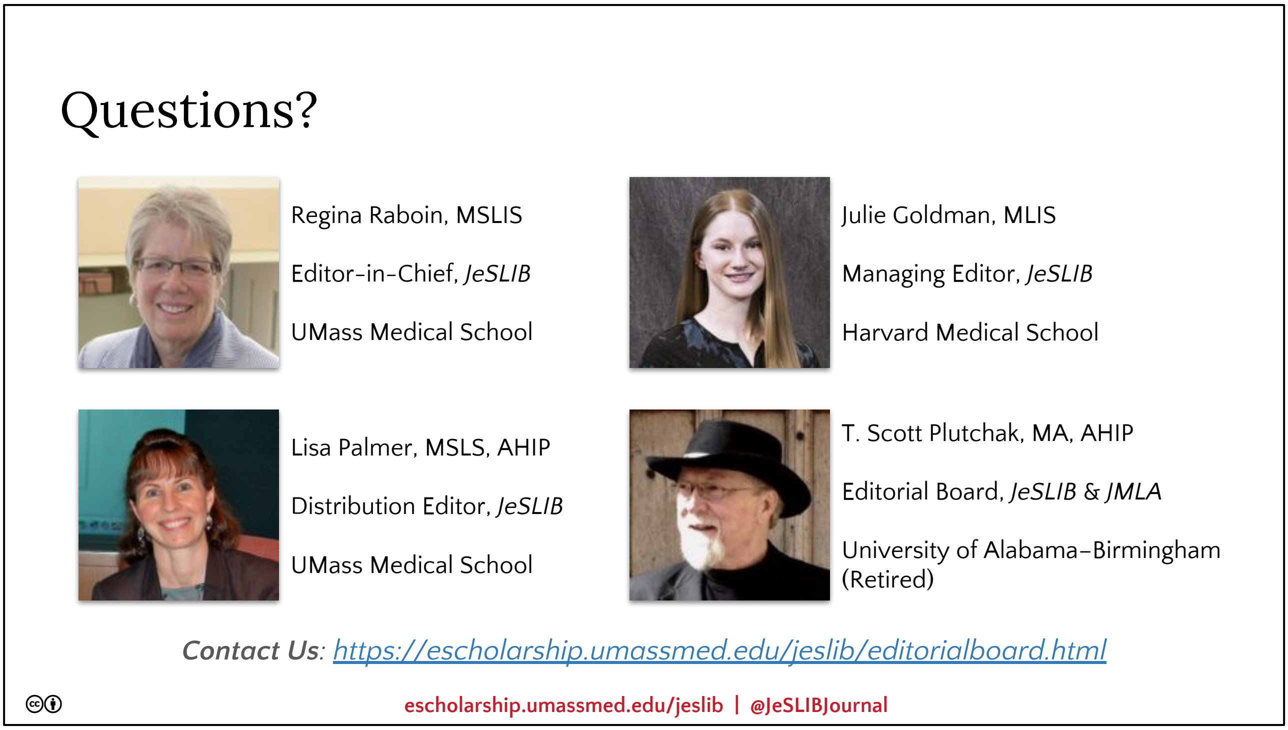
1. Share your data.
2. Provide metadata.
3. Provide an unprocessed form of the data.
4. Use standard data formats.
5. Use good null values.
6. Make it easy to combine your data with other data sets.
7. Perform basic quality control.
8. Use an established repository.
9. Use an established and open license.

Ethan P. White, Elita Baldridge, Zachary T. Brym, Kenneth J. Locey, Daniel J. McGlinn, and Sarah R. Supp. 2013. "Nine simple ways to make it easier to (re)use your data". *Ideas in Ecology and Evolution* 6(2):1–10 http://dx.doi.org/10.4033/iee.2013.6b.6.f

To close, here are “Nine Simple Ways” to make your data understandable, easy to analyze, and readily available to the wider community, applicable to both scientists and librarians.

1. Share your data.
2. Provide metadata. -- we discussed documentation
3. Provide an unprocessed form of the data. -- sharing the raw data, unanalyzed or code
4. Use standard data formats. -- we mentioned a few standards such as ISO 8601 for dates
5. Use good null values.
6. Make it easy to combine your data with other data sets. -- this is where tidy, well-structured data will be easy to work with
7. Perform basic quality control.
8. Use an established repository. -- Lisa described qualities to look for when evaluating a repository to choose
9. Use an established and open license.

Ethan P. White, Elita Baldridge, Zachary T. Brym, Kenneth J. Locey, Daniel J. McGlinn, and Sarah R. Supp: "Nine simple ways to make it easier to (re)use your data". *Ideas in Ecology and Evolution* 6(2):1-10 http://dx.doi.org/10.4033/iee.2013.6b.6.f



Questions?

Regina Raboin, MSLIS

Editor-in-Chief, *JeSLIB*

UMass Medical School

Julie Goldman, MLIS

Managing Editor, *JeSLIB*

Harvard Medical School

Lisa Palmer, MSLS, AHIP

Distribution Editor, *JeSLIB*

UMass Medical School

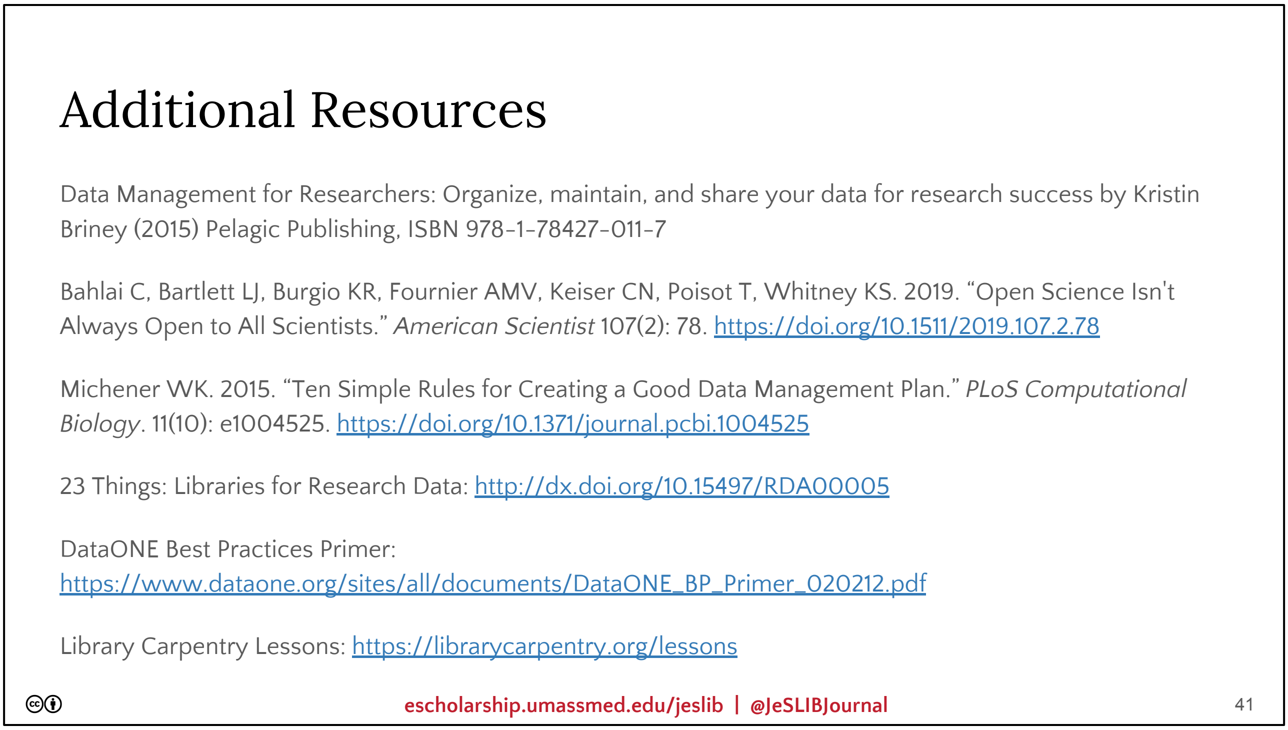
T. Scott Plutchak, MA, AHIP

Editorial Board, *JeSLIB* & *JMLA*

University of Alabama–Birmingham (Retired)

Contact Us: *https://escholarship.umassmed.edu/jeslib/editorialboard.html*

Contact Us: https://escholarship.umassmed.edu/jeslib/editorialboard.html



Additional Resources

Data Management for Researchers: Organize, maintain, and share your data for research success by Kristin Briney (2015) Pelagic Publishing, ISBN 978-1-78427-011-7

Bahlai C, Bartlett LJ, Burgio KR, Fournier AMV, Keiser CN, Poisot T, Whitney KS. 2019. "Open Science Isn't Always Open to All Scientists." *American Scientist* 107(2): 78. https://doi.org/10.1511/2019.107.2.78

Michener WK. 2015. "Ten Simple Rules for Creating a Good Data Management Plan." *PLoS Computational Biology*. 11(10): e1004525. https://doi.org/10.1371/journal.pcbi.1004525

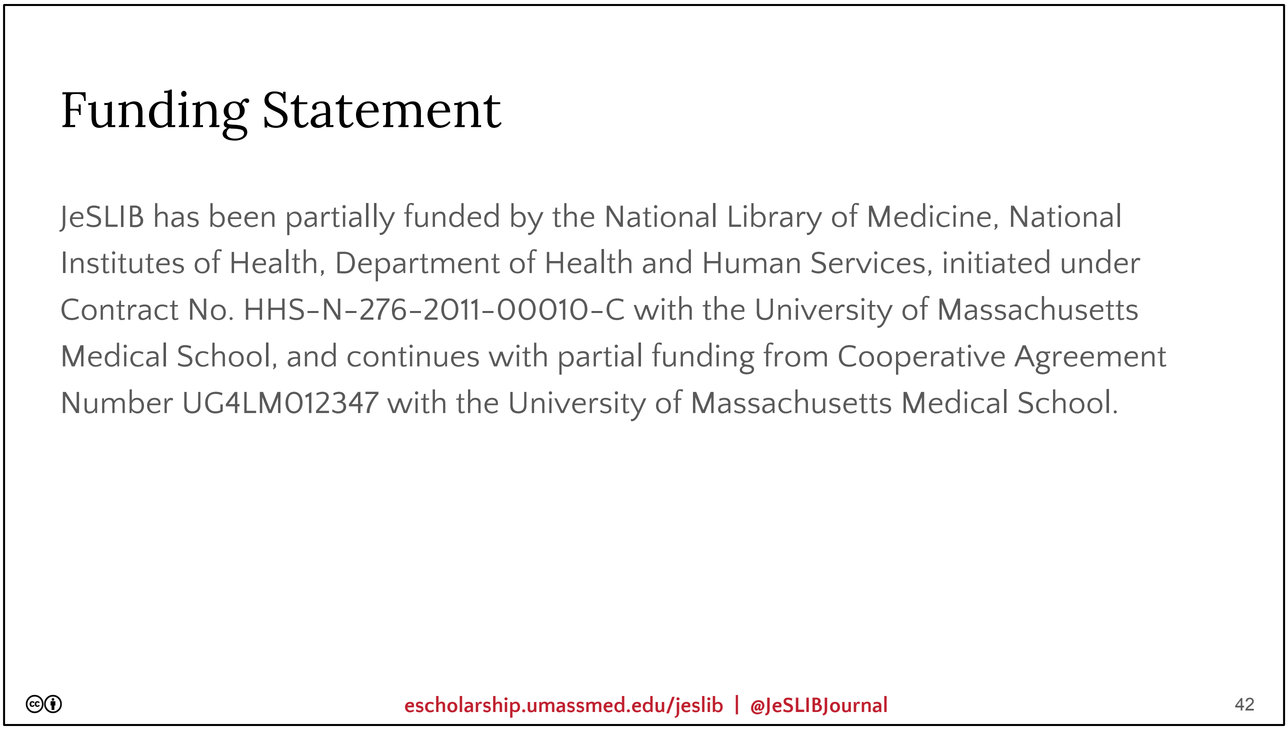
23 Things: Libraries for Research Data: http://dx.doi.org/10.15497/RDA00005

DataONE Best Practices Primer: https://www.dataone.org/sites/all/documents/DataONE_BP_Primer_020212.pdf

Library Carpentry Lessons: https://librarycarpentry.org/lessons

Additional Resources - See also the slide notes for more specific resources

- Data Management for Researchers: Organize, maintain, and share your data for research success by Kristin Briney (2015) Pelagic Publishing, ISBN 978-1-78427-011-7
- Bahlai C, Bartlett LJ, Burgio KR, Fournier AMV, Keiser CN, Poisot T, Whitney KS. 2019. "Open Science Isn't Always Open to All Scientists." *American Scientist* 107(2): 78. https://doi.org/10.1511/2019.107.2.78
- Michener WK. 2015. "Ten Simple Rules for Creating a Good Data Management Plan." *PLoS Computational Biology*. 11(10): e1004525. https://doi.org/10.1371/journal.pcbi.1004525
- 23 Things: Libraries for Research Data: http://dx.doi.org/10.15497/RDA00005
- DataONE Best Practices Primer: https://www.dataone.org/sites/all/documents/DataONE_BP_Primer_020212.pdf
- Library Carpentry Lessons: https://librarycarpentry.org/lessons



Funding Statement

JeSLIB has been partially funded by the National Library of Medicine, National Institutes of Health, Department of Health and Human Services, initiated under Contract No. HHS-N-276-2011-00010-C with the University of Massachusetts Medical School, and continues with partial funding from Cooperative Agreement Number UG4LM012347 with the University of Massachusetts Medical School.

JeSLIB Funding Statement: JeSLIB has been partially funded by the National Library of Medicine, National Institutes of Health, Department of Health and Human Services, initiated under Contract No. HHS-N-276-2011-00010-C with the University of Massachusetts Medical School, and continues with partial funding from Cooperative Agreement Number UG4LM012347 with the University of Massachusetts Medical School.