



Qualitative Data Sharing: Challenges and Opportunities

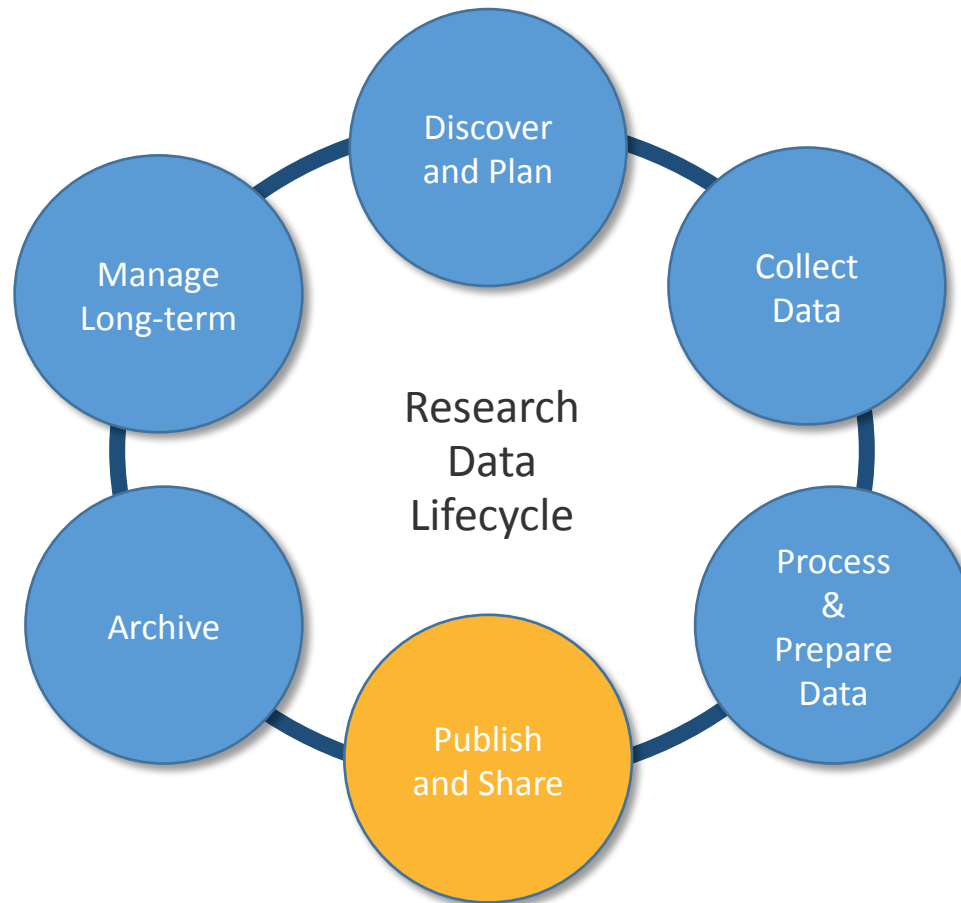
Jessica Trelogan
Research Data Services
j.trelogan@austin.utexas.edu

Janelle Hedstrom
Education Librarian
jhedstrom@austin.utexas.edu

University of Texas Libraries



What is Data Management?

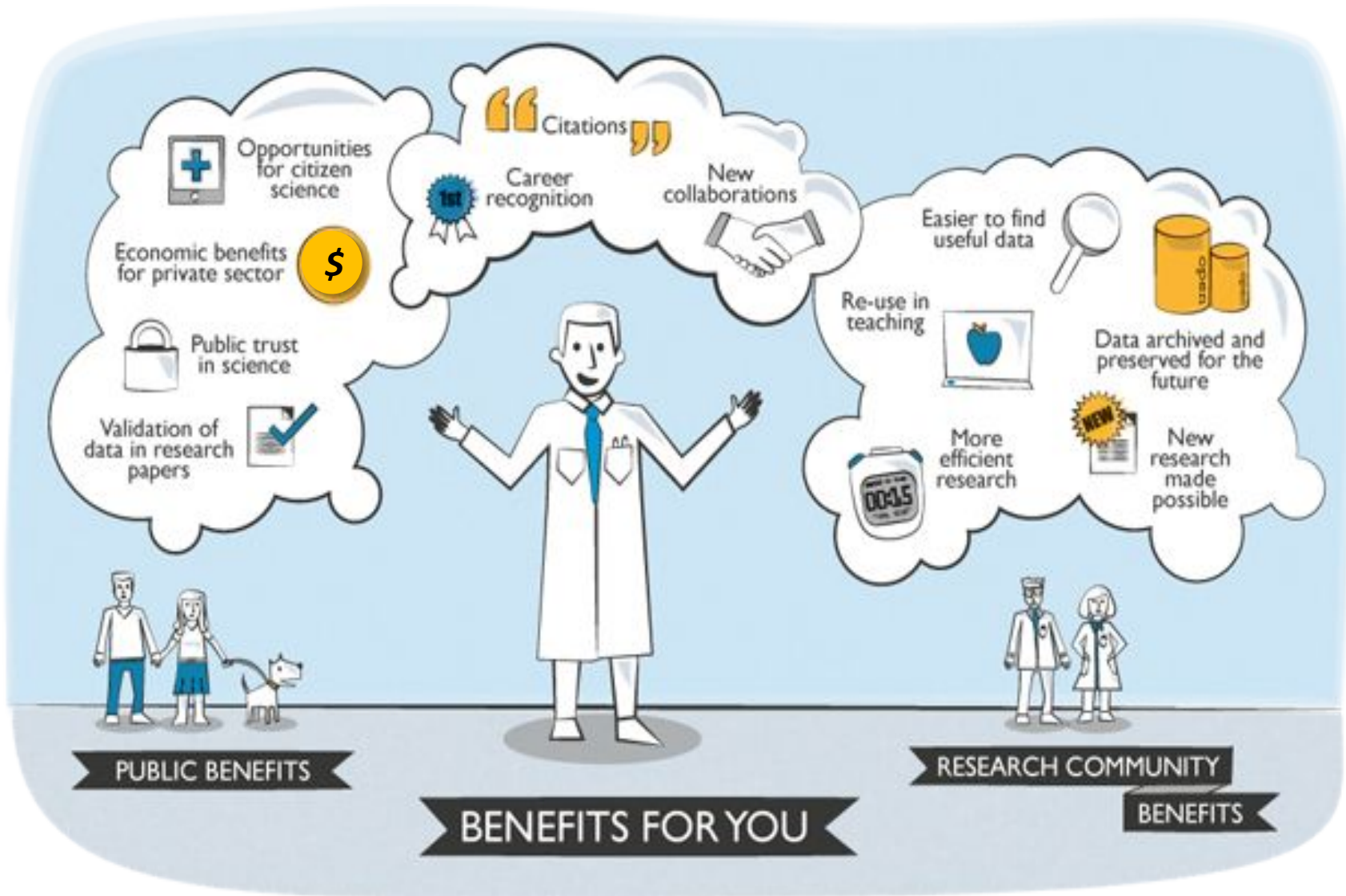


A collection of tasks practiced **throughout the lifecycle of research** that make it easier to find, understand, navigate, and use your data, as well as to *share it with others*.

What are qualitative data?

- In-depth/unstructured interviews, including video or audio
- Semi-structured interviews
- Structured interview questionnaires containing substantial open comments
- Maps/satellite imagery/geospatial data
- Unstructured or semi-structured diaries
- Observation field notes/technical fieldwork notes
- Case study notes
- Minutes of meetings
- Press clippings
- Court transcripts

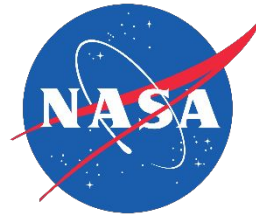
Why share?



Why share?

- Supports research transparency and reduces misconduct
- Fosters public trust
- Maximizes payoff of funder investment (especially taxpayer)
- Encourages new research with existing data
- Increases a scholar's visibility, citation rates and impact
- Advances qualitative meta-analytic research
- Provides data sets to your community for teaching and learning purposes

Why else?



NATIONAL ENDOWMENT FOR THE
Humanities

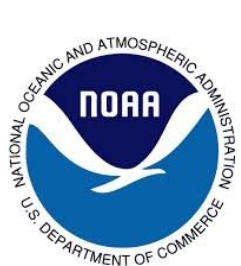


CENTERS FOR DISEASE
CONTROL AND PREVENTION



INSTITUTE of
Museum and Library
SERVICES

GORDON AND BETTY
MOORE
FOUNDATION



wellcome trust

It's required.

But...

I don't have the time or money.

I don't have the expertise.

My data are too sensitive.

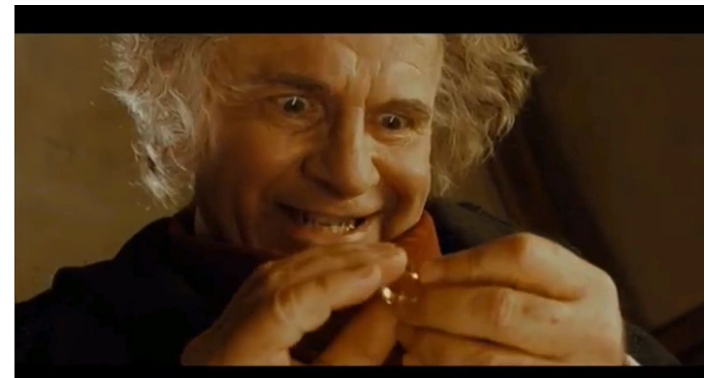
My data will be used incorrectly.

No one will understand my data.

I don't know where to start.

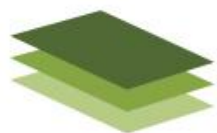
I'm not done publishing.

It's MINE!!!



Data Management Plans

A data management plan (DMP) is a **written document** describing the **nature** and **structure** of the data you will likely use or produce in the course of research, along with your **strategies** for dealing with it **throughout and after** your project.



DMPTool

Guidance and Resources for your Data Management Plan

<https://dmptool.org/>

Data Sharing & Ethics

Transparency is key!

In the project-planning phase, determine...

- Any legal constraints on sharing (FERPA/HIPAA)
- Types of data you'll need to de-identify
- Methods for de-identification of data

Include the possibility of sharing in...

- Consent forms
- IRB documentation
- Grant applications

Sample Informed Consent

“The information in this study will only be used in ways that will not reveal who you are. You will not be identified in any publication from this study or in any data files shared with other researchers. Your participation in this study is confidential. Federal or state laws may require us to show information to university or government officials [or sponsors], who are responsible for monitoring the safety of this study.”

Managing Sensitive Data

UT Information Security Office

https://security.utexas.edu/policies/sensitive_research

Office of Research Support: research.utexas.edu/ors/

Institutional Review Board (IRB)

Institutional Animal Care and Use Committee (IACUC)

Institutional Biosafety Committee (IBC)

Conflict of Interest section

Document, Document, Document

Data only useful if understandable!

Metadata

Readme.txt (use a [template](#))

Codebooks/Data Dictionaries

Lab books/field notes

Electronic Lab Notebooks



Photo courtesy of Institute of Classical Archaeology

Documenting Qualitative Data

- Research methods and practices (including the informed consent process) that are fully documented
- Blank copy of informed consent form with IRB approval number
- Details on setting of interviews
- Details on selection of interview subjects
- Instructions given to interviewers
- Data collection instruments such as interview questionnaires
- Steps taken to remove direct identifiers in the data (e.g., name, address, etc.)
- Any problems that arose during the selection and/or interview process and how they were handled
- Interview roster

Storage options at UT

UT Box

- UNLIMITED
- ISO-approved for Confidential data [HIPAA, FERPA, ITAR]
- Goes away when you do!

UTMail (Google Drive)

- UNLIMITED
- NOT approved for Confidential data
- Lifetime access

TACC: 5TB Free for PIs

ITS:

- Storage at UTDC at cost
- Storage on Austin Disk at \$1.10/GB/yr



UT's Data Classification Standard

Published

University data not otherwise identified as Confidential or Controlled data, and:

1. The data is publicly available, and
2. Such data have no requirement for confidentiality, integrity, or availability

Controlled

University data not otherwise identified as Confidential, and:

1. Data not publicly available, and
2. Data releasable in accordance with the Texas Public Information Act (e.g., contents of specific e-mail, date of birth, salary, etc.)

Confidential

University data that is:

1. Protected specifically by federal or state law or
2. Protected by University of Texas rules and regulations
3. Data not otherwise protected by a known civil statute or regulation, but which must be protected due to contractual agreements requiring confidentiality, integrity, or availability considerations

Keep Your Data Safe

Use strong passwords

- stache.security.utexas.edu

Do system updates

Have a backup strategy

- 3-2-1 rule: 3 copies; 2 media types; 1 off-site

Confidential data

<https://security.utexas.edu/iso-policies/cloud-services/decision-matrix>



The University of Texas at Austin
Information Security Office

Data Management & Personal Identifiers

Consider both direct (name, SSN, etc) and indirect identifiers (geography, race, institution, occupation, income, date of events, etc.)

Anonymization Goals:

- Add uncertainty to potential identity matches
- Balance risk against analytic utility
- Show due diligence

When Anonymization isn't possible: Establish sound restrictions for access and data sharing agreements.

Anonymization Tips

During planning: Create an anonymization protocol for use during transcription and analysis.

During collection and analysis : Do anonymization in this phase! Keep an anonymization log and store the encrypted log separate from data and other documents.

- Remove or use codes/pseudonyms/generalizations for direct identifiers
- Broaden indirect identifiers (zip code changed to city or state, map specific age to a range, etc.)
- Remove or recode indirect variables that shouldn't or can't be broadened.

Example Anonymization Log

	Original	Changed to
page 1	Jay	Carl (pseudonym)
page 1	age 16	teenager
page 1	African American	minority group
page 2	Fritz High School	a high school
page 3	Mrs. Jones	teacher
page 3	Mr. Franklin	parent
page 4	Austin	Texas city

Why Share in a Repository?

- You'll receive stable URLs for your deposits
- Your data becomes a citable object, so you can track citations and the impact it's having on other's work
- Your work becomes more discoverable
- They'll often help with curation and ingest
- Allow for various level of restricted access
- They'll do the work of preservation and maintenance



<https://www.flickr.com/photos/johnjack/3449280414>

Data Publication & Preservation: Repositories

re3data.org : Find your disciplinary repository.

Texas Data Repository: UT's tool for preserving and sharing research data. <http://texasdatarepository.org/>

QDR (Qualitative Data Repository) <https://qdr.syr.edu/>

ICPSR: <https://www.icpsr.umich.edu/>

What to Publish or Share?

In addition to, or in lieu of your raw or transformed data, you might share ...

- Readme files and user guides
- Consent forms
- Survey instruments, interview questions
- Anonymization protocols, scripts, etc.
- Data narratives, dictionaries and documentation
- Codebooks / coding schema / coding methods
- Data analysis protocols, documents and notes
- Word frequency queries and other QDCAS project files
- Terms of Use or licensing details



<https://www.flickr.com/photos/ryanr/142455033>

Examples of Published Qualitative Data Collections

Texas Data Repository

- [Points of Comparison: What Indicating Gestures tell us About the Origins of Signs in San Juan Quiahije Chatino](#)
- [Iron Fellows: Commitment and Activism in a Poor People's Movement](#)
- [FB Ads from 2015-2017](#)

ICPSR

- [Gender & Multicultural Leadership Project](#)
- [Polyvictimization Among Girls in the Juvenile Justice System](#)
- [Teacher Quality Grants, Texas](#)

Examples of Published Qualitative Data Collections, cont.

QDR (Qualitative Data Repository)

- [A Qualitative Study on the Experiences and Perspectives of Public Sector Patients in Cape Town..](#)
- [Brokers, Voters and Clientelism](#)
- [High Performance School-aged Athletes at Australian Schools](#)

Data Ownership at UT

Sec. 9 Use of Research Data.

“Research data or results created by an employee are **owned by the Board of Regents** and except to the extent that rights to such research data are contractually assigned or licensed to another by the Board of Regents, the **creator shall have a nonexclusive license to use** such data for patient care, teaching, scholarly, and other academically related purposes and nonprofit research, provided such activities are within the scope of the employee's employment.”

<https://www.utsystem.edu/board-of-regents/rules/90101-intellectual-property>

Ownership, Copyright & Data Use Agreements

Basic understanding of copyright law is YOUR responsibility. Only share what you have rights to share!

Quantitative data are often considered facts and not copyrightable, but qualitative data may be (e.g., video? images? field notes?).

Even when you can't copyright your data, take the time to create licenses and think about your terms of use.

Ownership, Copyright & Data Use Agreements, cont.

Don't automatically use the default settings in your repository of choice. Many repositories allow for varying levels of access across your deposited files.

Access settings can usually be changed over time. You may deposit restricted data now, and you can choose to make it public in the future.

Support

[UT Libraries Research Data Services](#)

[UT Libraries Subject Librarians](#)

We can help with:

- Research data management planning
- Metadata and documentation advice
- Tools and repository recommendations
- Copyright and fair use questions
- Texas Data Repository deposits

For best results, contact us early in the process!

Sources & Recommended Reading

Antes, A. L., Walsh, H. A., Strait, M., Hudson-Vitale, C. R., & DuBois, J. M. (2018). [Examining data repository guidelines for qualitative data sharing](#). *Journal of Empirical Research on Human Research Ethics*, 13(1), 61–73.

[De-identifying human subject data for sharing your research](#). (2015). *Johns Hopkins Data Management Services*.

DuBois, J. M., Strait, M., & Walsh, H. (2017). [Is It Time to Share Qualitative Research Data?](#) *Qualitative Psychology*.

Elman, C., & Kapiszewski, D. (2013). [A guide to sharing qualitative data](#). *Center for Qualitative and Multi Method Inquiry (CQMI), Syracuse University*.

Sources & Recommended Reading, cont.

[Guide to social science data preparation and archiving](#). (2012).
Inter-University Consortium for Political and Social Research.
University of Michigan.

Mannheimer, S., Pienta, A., Kirilova, D., Elman, C., & Wutich, A. (2018). [Qualitative data sharing: Data repositories and academic libraries as key partners in addressing challenges](#). *American Behavioral Scientist*.

Tsai, A. C., Kohrt, B. A., Matthews, L. T., Betancourt, T. S., Lee, J. K., Papachristos, A. V., ... & Dworkin, S. L. (2016). [Promises and pitfalls of data sharing in qualitative research](#). *Social Science & Medicine*, 169, 191-198.