

Data Sharing and Data Management Plans

Syracuse University OSP
Sebastian Karcher, QDR
April 13, 2022



DRYAD

Dissemination and Sharing of Research Results -
NSF Data Management Plan Requirements

NSF DATA SHARING POLICY



Why Care About Data Sharing & Research Transparency?

Final NIH Policy for Data Management and Sharing

Notice Number:

NOT-OD-21-013

Key Dates

Release Date:

October 29, 2020

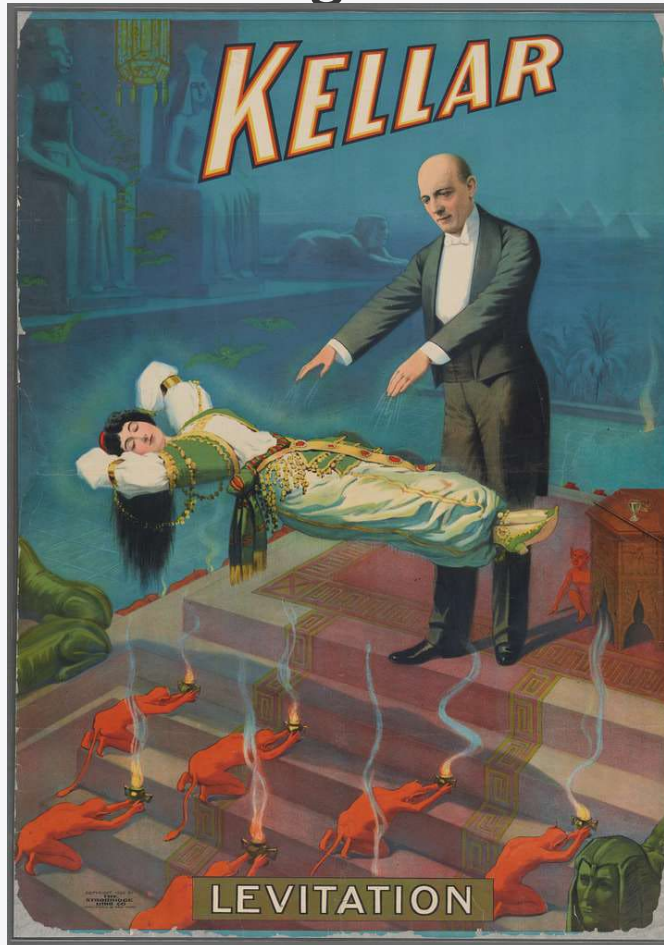
Effective Date:

January 25, 2023

QDR
The Qualitative
Data Repository

Being a good researcher

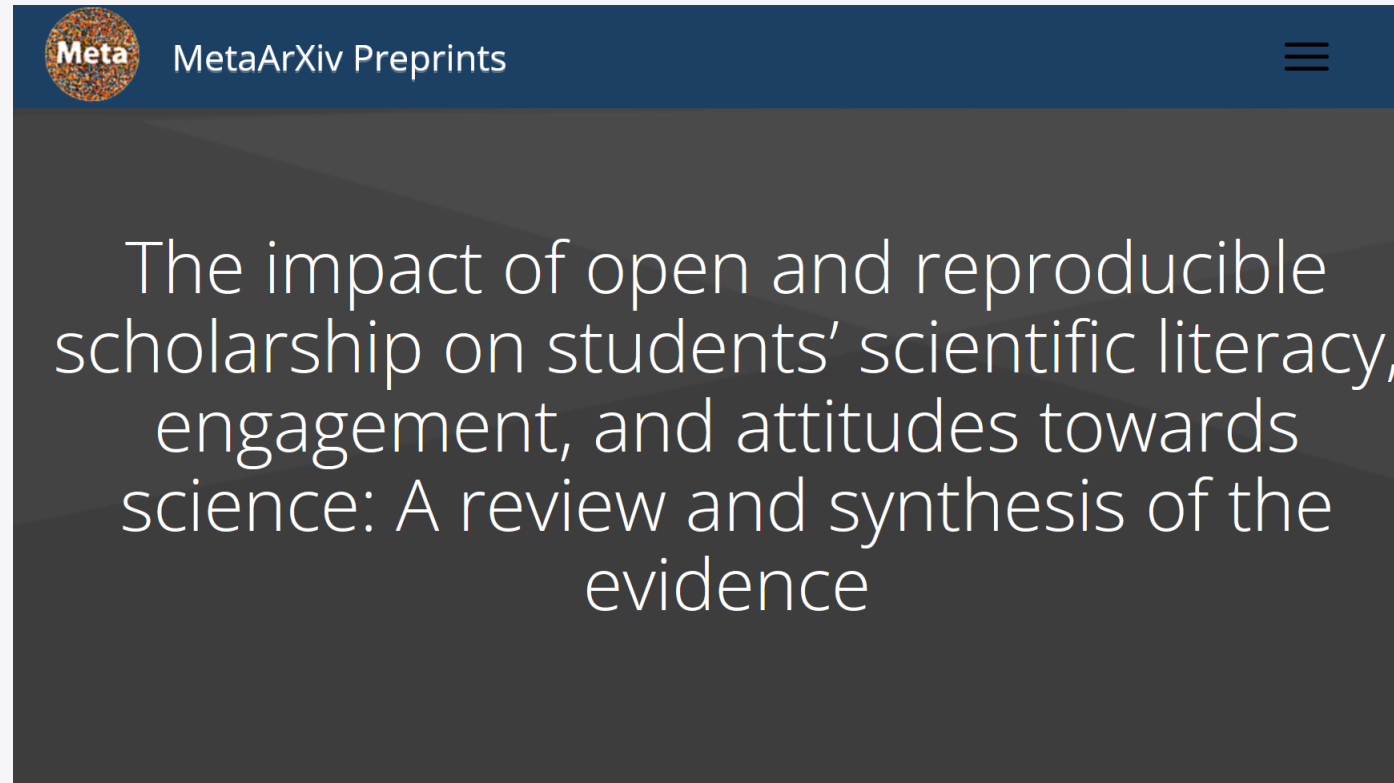
How we come to gain knowledge is foundational to science



This is cool! But not science

Being a good teacher

Shared data are foundational for effective teaching



“Our literature review has indicated that explicitly embedding open and reproducible scholarship into Higher Education can clearly enhance student engagement, scientific literacy, and attitudes.” (p. 46)

Source <https://doi.org/10.31222/osf.io/9e526>

Being a good colleague

Allow others to scrutinize...



Joscha Legewie
@jlegewie

Update on "Police Violence and the Health of Black Infants": After publication, a reader discovered classification errors in the openly shared data. After learning about errors, I conducted a thorough investigation focusing on a larger sample of cases that revealed: 1/4

10:58 PM · Dec 11, 2019 · Twitter Web App



Joscha Legewie
@jlegewie

Therefore, I requested a retraction of the article and the editors agreed. I apologize that errors were not discovered before publication. I am grateful that someone found the classification errors allowing me to investigate the issue and correct it quickly. 4/4

10:58 PM · Dec 11, 2019 · Twitter Web App

...and build on your work

Publication, Publication

Gary King, *Harvard University*

<https://gking.harvard.edu/files/abs/paperspub-abs.shtml>

“I show herein how to write a publishable paper by beginning with the replication of a published article. (...)

Some students ask: ‘Why begin an original research paper by replicating some old work?’ A paper that is publishable is one that by definition advances knowledge. If you start by replicating an existing work, then you are right at the cutting edge of the field. If you can then improve any one aspect of the research that makes a substantive difference and is defensible, you have a publishable paper.”

Be a good citizen

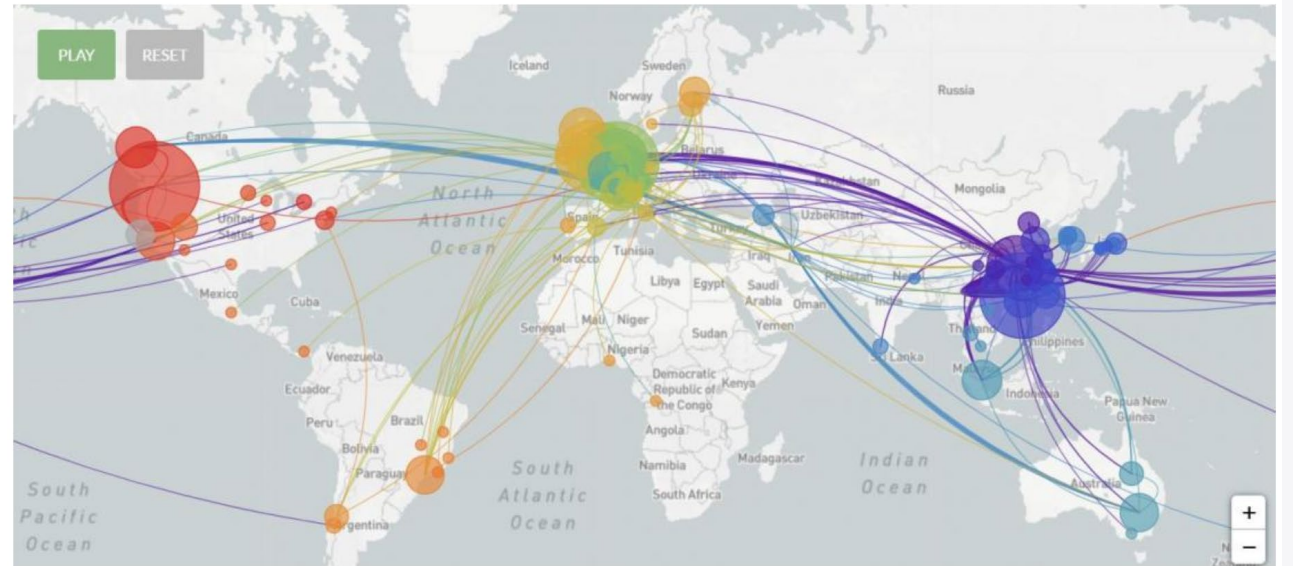
Open science is more impactful

Covid-19: How unprecedented data sharing has led to faster-than-ever outbreak research

23 March 2020

by *Ian Le Guillou*

Republish ↗



Source: <https://ec.europa.eu/research-and-innovation/en/horizon-magazine/covid19-how-unprecedented-data-sharing-has-led-faster-ever-outbreak-research>

Get funded and published


Data Availability

The following policy applies to all PLOS journals, unless otherwise noted.

Introduction

PLOS journals require authors to make all data necessary to replicate their study's findings publicly available without restriction at the time of publication. When specific legal or ethical restrictions prohibit public sharing of a data set, authors must indicate how others may obtain access to the data.

Source: <https://journals.plos.org/plosone/s/dataavailability>



The TOP Guidelines were created by journals, funders, and societies to align scientific ideals with practices.

TOP provides a suite of tools to guide implementation of better, more transparent research.

The TOP Guidelines are a recognized standard in publishing and funding.

With over 5,000 signatories, the TOP Guidelines are a widely used tool for implementing open science practices.

Data Management Plans

Paul H. Bern, Ph.D.

April 13, 2022



Research Data Services



Syracuse University Libraries

[Libraries](#) / [Research Guides](#) / [Research Data Services](#) / Home

Research Data Services: Home

[Home](#) [Data Management Planning](#) [Data Collection](#) [Analysis & Visualization](#) [Research at S.U.](#) [Research Tutorials](#) [Data Purchase Program](#)

Qualtrics Workshops

Research Data Services will be presenting a series of [workshops on using Qualtrics](#) online survey software!!

Related Sites

- [Numeric Data Resources](#)
- [GIS - Geographic Information Systems](#)
- [Resources for Doctoral Dissertation / Thesis Writers](#)
- [SURFACE](#)
- [Electronic Theses and Dissertations](#)
- [Data Science](#)
- [Arts and Humanities Funding](#)

Contact Us

SU Libraries Research Data Services

SU Libraries offers a broad range of research data services related to the identification, collection, management, analysis, and curation of quantitative and qualitative research data. To contact the Research Data Services group, please send an email to datasvcs@syr.edu.

Data Management Planning

Many funding agencies, such as the National Science Foundation (NSF) and the National Institutes of Health (NIH), have requirements for data sharing and data management plans. Research Data Services can help you to put together such a plan to comply with the requirements.

Data Collection and Data Discovery

Research Data Services can provide consulting in research methods, study design, and questionnaire and interview design. We also provide assistance in locating and using freely available as well as proprietary quantitative, qualitative, and GIS data.

Data Analysis

Research Data Services can assist you with quantitative and qualitative data analysis, use of software, especially SAS, Stata, SPSS, Qualtrics, and ArcGIS. Services include research methodology, instrument design, and data analysis.

Data Visualization

Research Data Services can help you identify data visualization and [GIS](#) tools and resources.

What Is a Data Management Plan?

- A DMP describes how you will collect, organize, store, secure, back up, preserve and share your data
 - the types of data, samples, physical collections, software, curriculum materials, and other materials to be produced in the course of the project;
 - the standards to be used for data and metadata format and content (where existing standards are absent or deemed inadequate, this should be documented along with any proposed solutions or remedies);
 - policies for access and sharing including provisions for appropriate protection of privacy, confidentiality, security, intellectual property, or other rights or requirements;
 - policies and provisions for re-use, re-distribution, and the production of derivatives; and
 - plans for archiving data, samples, and other research products, and for preservation of access to them.

http://www.nsf.gov/pubs/policydocs/pappguide/nsf11001/gpg_2.jsp#dmp

Why Do I Need A DMP?

- Required by NSF, NIH and other funding agencies as well as some journals
 - Even if there is no requirement from the funder, journal, creating a DMP is still a good idea
- Sharing your data is required
 - Fully justify why you cannot share the data
 - Having human subjects is not, in and of itself, a reason to not share

Types of Data

- The types of data, samples, physical collections, software, curriculum materials, and other materials to be produced during the project;
 - What type of data you will collect and how, including secondary data
 - Data format and size over the course of the project
 - A plan to back up the data
 - The tools or software needed to create, process, visualize the data

Metadata

- The standards to be used for data and metadata format and content (where existing standards are absent or deemed inadequate, this should be documented along with any proposed solutions or remedies)
 - Different disciplines have different standards, as do different archives
 - Metadata aids in discoverability
 - How can someone else use this data without my help?

Sharing and Access Policies

- Policies for access and sharing including provisions for appropriate protection of privacy, confidentiality, security, intellectual property, or other rights or requirements
 - How and where will you make the data available when the time comes?
 - What will be the process for accessing the data?
 - How will you protect privacy?
 - What, if any, complications might there be vis-à-vis IRB requirements?
 - Data sharing must be included in Informed Consent

Re-use of the Data

- Policies and provisions for re-use, re-distribution, and the production of derivatives; and
 - Who needs to access the data during the project and how they will do so
 - Designate a data manager
 - Who will want to use the data after the project is over?

Archiving and Preservation

- Plans for archiving data, samples, and other research products, and for preservation of access to them.
 - What archive/repository?
 - Retention period
 - Will the data be curated?

Active vs Archival vs Shared Data

- You will need different procedures and standards for the working or active data as opposed to the archival
- Active data may need frequent updating and access by several people
 - Version control!
- Archival data is relatively fixed, only need infrequent access
 - Version control!
- Shared data is what the rest of the world gets

Points to Ponder

- Data format and size over the course of the project
- Retention period
- Privacy and security requirements – IRB!
- A plan to back up the data
- Who needs to access the data during the project and how they will do so
- Designate a data manager
- The tools or software needed to create, process, visualize the data

DMPTool

- Online tool to assist in completing a DMP
- Has templates for all NSF and NIH directorates as well as several other funding agencies
- Has samples, suggestions
- <https://dmptool.org/>

Thank you!

- phbern@syr.edu
- <http://researchguides.library.syr.edu/dataservices>

Thinking about Data Management Plans

Data Management Plans: A Very Short History

- Originated in 1960s and 1970s for complex engineering projects, especially aerospace
- In early 2000s, renewed interest to accommodate growth of digital data
- Mandated for all funding bodies in UK since mid-2000s
- In US, mandated by NSF since 2011; by NIH starting 2023; focus on data sharing;
- Mandated for EU funding as part of Horizon 2020

DMP vs. IRB

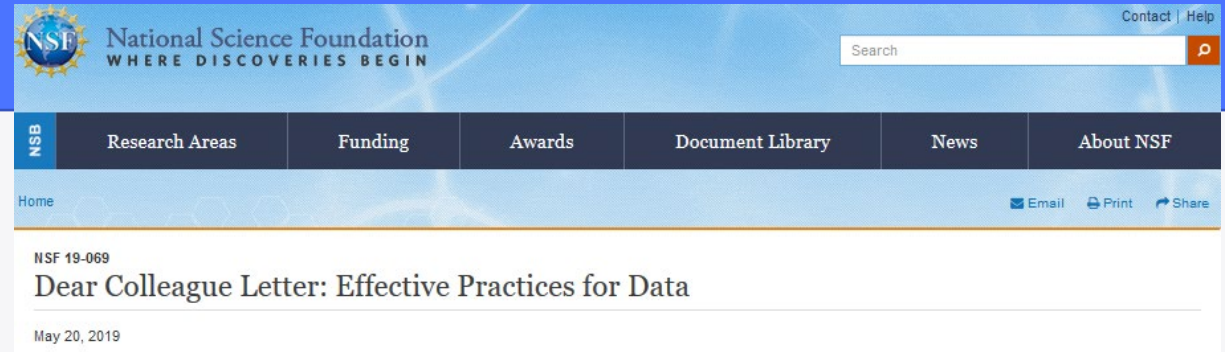
- IRB: Required based on federal regulations
- DMP: Required by many funders
- *Both require you to document data collection and sharing procedure*
- Critical to ensure DMP and IRB application align

DMP: Your Audiences

- Your funder
 - Main interest: funding impact, data sharing
- Grant reviewers
 - Main interest: your proposal, does the DMP support your proposal
- Your team
 - Co-authors, student workers, translators, transcribers, committee members
 - Main interest: data collection, does DMP help convey principles to team members
- Your future self
 - Main interest: does DMP help remind you of/prepare you for key decisions

What Funders Want

- Maximize impact
- Data sharing
- Use of suitable infrastructure
- DMP Tool
- Data repository
- NSF: Explicitly allows for DMP budget



Grant reviewers

NSF Example

- DMP is part of review package
- Panel *has to* comment on (i.e. at least skim) DMP

DMP = Additional 2 pages to showcase your competence

Your Team

- DMP as reference document for team members
 - “How did we say we were going to store/organize...”
- Be specific! DMP can & should contain specifics about file/folder organizations, formats, etc.
- DMP as “living document”—keep your DMP updated with changes in plans

Future Self

- DMP: Opportunity to plan key decisions ahead of time
- Check with others involved
 - IRB
 - IT
 - Data Repository
 - Local partner organizations

Find examples of great DMPs

DMP Competition Winners: DMPs so good they go to 11



<https://qdr.syr.edu/qdr-blog/dmp-competition-winners-dmps-so-good-they-go-11>



<https://www.faseb.org/resources/data-science-and-informatics/dataworks-dmp-challenge>

Stay in Touch

- qdr@syr.edu
- Twitter: @qdrepository
- Blog: <https://qdr.syr.edu/qdr-blog>
- Available for curation consultation for DMPs involving qualitative data