4-8-2020

UK Libraries: Your Partner in Data Management, Data Sharing, and Data Preservation

Lauren E. Robinson
University of Kentucky, laurene.robinson@uky.edu

Adrian K. Ho
University of Kentucky, hoadriank@gmail.com

UK Libraries: Your Partner in Data Management, Data Sharing, and Data Preservation

Lauren Robinson, Medicine Liaison Librarian, *laurene.robinson@uky.edu*
Adrian Ho, Director of Digital Scholarship, *adrian.ho@uky.edu*

University of Kentucky

Overview

- Define Data
- Discuss why research data management (RDM) is important
- Discuss the Data Lifecycle and RDM Tools
- Discuss how UK Libraries can help with RDM

WHAT IS DATA?

National Institutes of Health (NIH) defines data as “recorded factual material commonly accepted in the scientific community as necessary to validate research findings.”

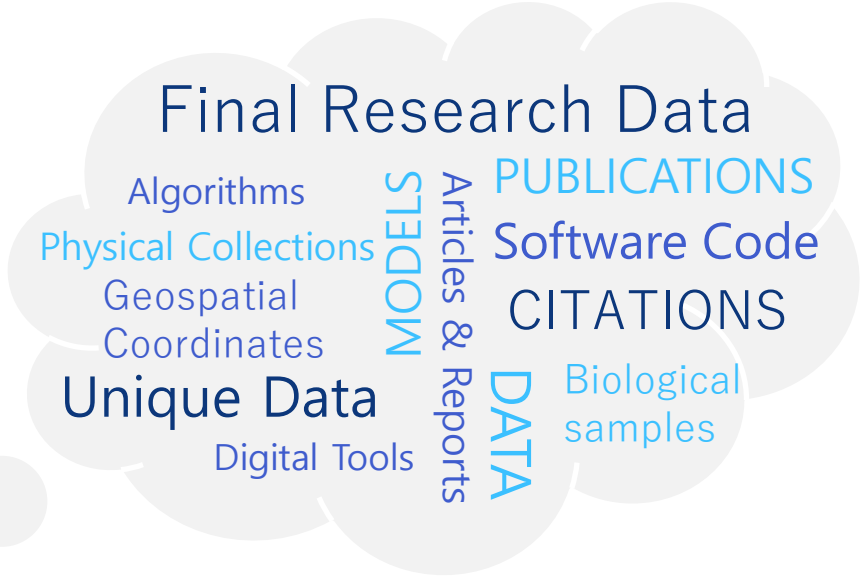
WHAT IS DATA?

The National Science Foundation (NSF) considers data to be something determined by the community of interest through the process of peer review and project management. NSF's definition points out the data are essential to the research community and not just to individual researchers or research teams.

WHAT IS DATA?

The National Endowment for the Humanities (NEH) includes a wider array of data types than we see from NSF or NIH. Any of these data types NEH defines as humanities data, or "materials generated or collected during the course of conducting research."

WHAT IS DATA?



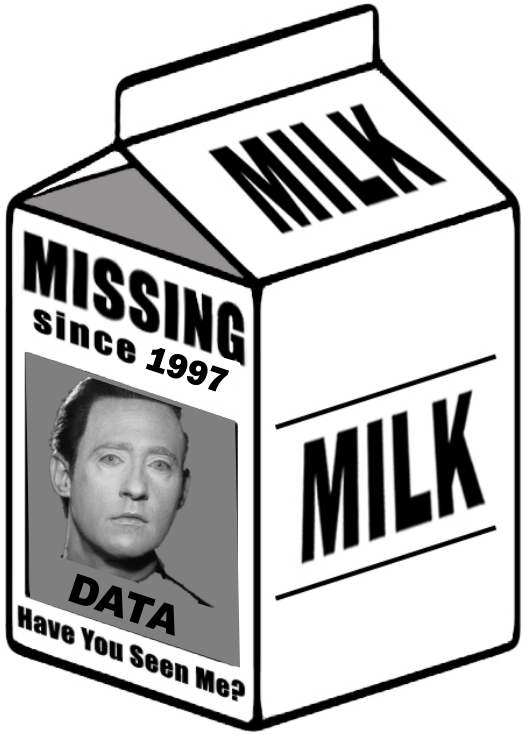
THERE IS NO CLEAR DEFINITION OF DATA. DATA MEANS DIFFERENT THINGS TO DIFFERENT GROUPS.

BUT IF WE HAD TO DEFINE DATA…

> *“Data should be valid, shared, and are heterogeneous, and contextualized within research communities.”*

WHY MANAGE DATA?

"A 2013 study found that poor data management practices have led to a concerning loss of scientific research data, with 80 percent of data unavailable after twenty years."



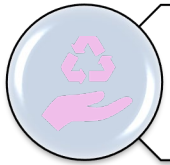
Elizabeth Howe and Richard Van Noorden, "Scientist Losing Data at a Rapid Rate," Nature News, December 19, 2013

WHY MANAGE DATA?

Be a Responsible Researcher

To Make Collaboration Easier

Be a Good Steward



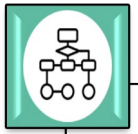
Reusability and Reproducibility

DATA LIFECYCLE

Developed using the DataOne Lifecycle
https://www.dataone.org/data-life-cycle

How can UK Libraries help?

UK Libraries can help…

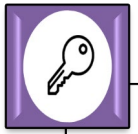


PLAN

Planning entails describing the data that will be compiled, and how that data will be managed and made accessible throughout its lifetime.

DESCRIBE

Describe entails accurately and thoroughly describing your data using the appropriate metadata standards



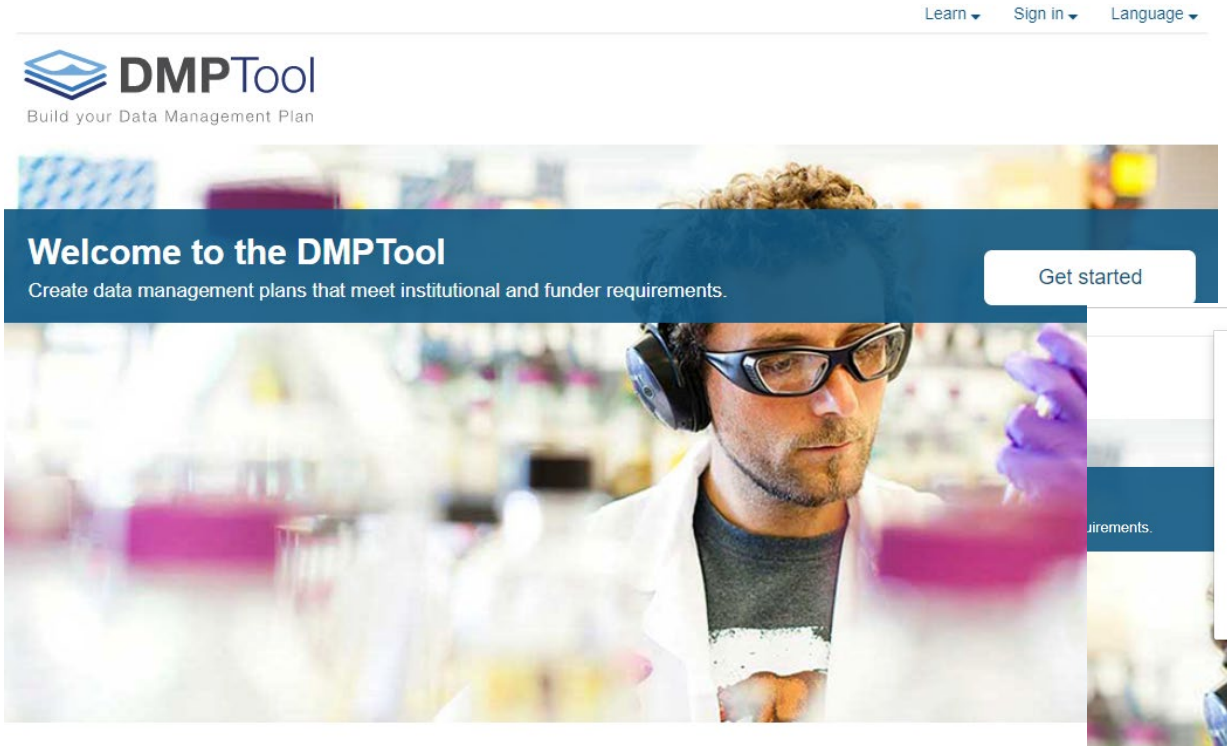
PRESERVE

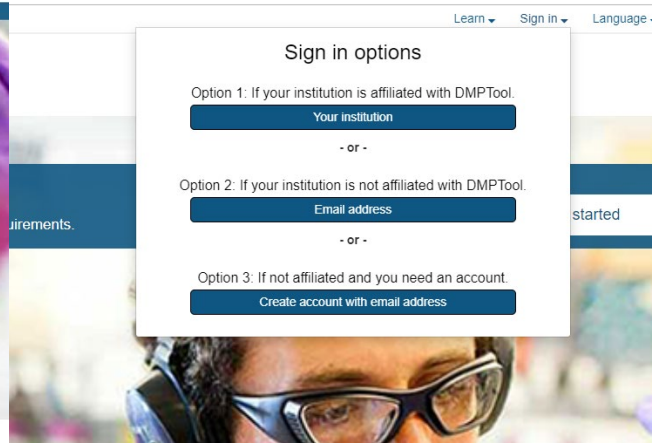
Preserve entails ensuring that your data are submitted to an appropriate long-term archive (i.e. data center)

DISCOVER

Discover entails making potentially useful data accessible, along with the relevant information about the data (metadata)

DMP Tool





DMPTool by the Numbers

Top Templates

Digital Curation Centre
NIH-GEN: Generic
NSF SBE: Social, Behavioral, Economic Sciences

Data Management Plans for the United States Migraine Study Protocol

A Data Management Plan created using DMPTool

Creator: Catherine Zappia

Affiliation: University of New Mexico (UNM)

Template: National Institutes of Health (NIH)

Last modified: 09-29-2015

Data Management Plans for the United States Migraine Study Protocol

Data sharing plan

Project Information

This Data Management Plan (DMP) entails the data which will be collected from a study conducted on behalf of the United States Migraine Study Protocol Group. The study is being conducted to assess rizatriptan's effectiveness in treating acute migraine at its onset versus in a later phase of the attack.

The study will ask those who experience migranes to treat their headaches twice; once with rizatriptan at the early onset of a migraine, and again with rizatriptan once the migraine becomes more severe. The patients will then be asked to report how they felt after using the medication each time. The study will assess: time to onset of headache relief, headache severity 2 hours postdose and whether or not patients became largely symptoms-free (including associated symptoms), and when they were able to resume normal activities.

General Data Management Plan Information

This study is being conducted to draw conclusions about the impact timing has on the medical benefits of the medication. The results will be used to better inform rizatriptan users on when they should be taking their medication to achieve maximum migraine relief in the quickest time possible. The results will also be shared accross the United States Migraine Study Protocol Group. This Data Management Plan is intended for review by the USMAP Group before disemination of the results.

Additional data sharing requirements

Policies for Access and Sharing

There are no requirements for making the results of the study private. All participants have signed waivers to be able to participate in the study, and to have their results shared, with the exception of any personal information. The USMAP group wishes to make their results available to the public, for anyone who is interested in taking rizatriptan or has questions about the drug, and for anyone within the USMAP group who was not directly part of the study.

Legal Guidelines and Requirements

The results of this study will not contain any personal information of the participants involved, nor will it contain any unique personal identifiers that could cause assumptions to be made about those involved. In order to assure that the results can be shared publicly, precautions have been taken to ensure that the participant's identities are kept anonymous.

Types of Data Produced

The data being produced from this study will include statistical values written down and made into charts. There are a number of variables being observed (time drug was taken, headache severity 2 hours postdose, etc.), and each of the results will be calculated and formed into a visual aid. Along with the charts, a report will be written up including the statistics and the conclusions that can be drawn from the results of the study.

Plans for Archiving and Preservation

There are different plans for archiving and preserving the data for the short term and the long term. However, for both plans security is a concern. The results of the study will be made public, but there is always a concern that someone may attempt to steal the data and conclusions and claim the study for their own. To protect against this, we are going to store and archive the data on a secure, password protected system that conducts a back up daily.

Short Term

The short term plans for archiving and preserving the data includes saving the results on the USMAP Group website. The results of the study will be dated and saved so that future researchers have access to the information.

Long Term

The long term plans for archiving and preserving the information involve saving the results on the USMAP group database once the results of the study become outdated. Although the results may be old and possibly irrelevant, we still want furture USMAP group members to have access to the study that will be conducted, either to base their studies off of or to encourage further conclusions to be drawn.

Metadata and File Formats

METADATA

- UK Libraries can help with finding and using the appropriate metadata schema that ensures that you capture the most important information about the research data.

- We can also help you identify a suitable vocabulary for the data description to ensure accuracy and consistency.

FILE ORGANIZATION

UK Libraries can provide assistance in managing data files in these areas:

- File naming convention to facilitate organization and versioning
- File formats most suitable for long-term preservation

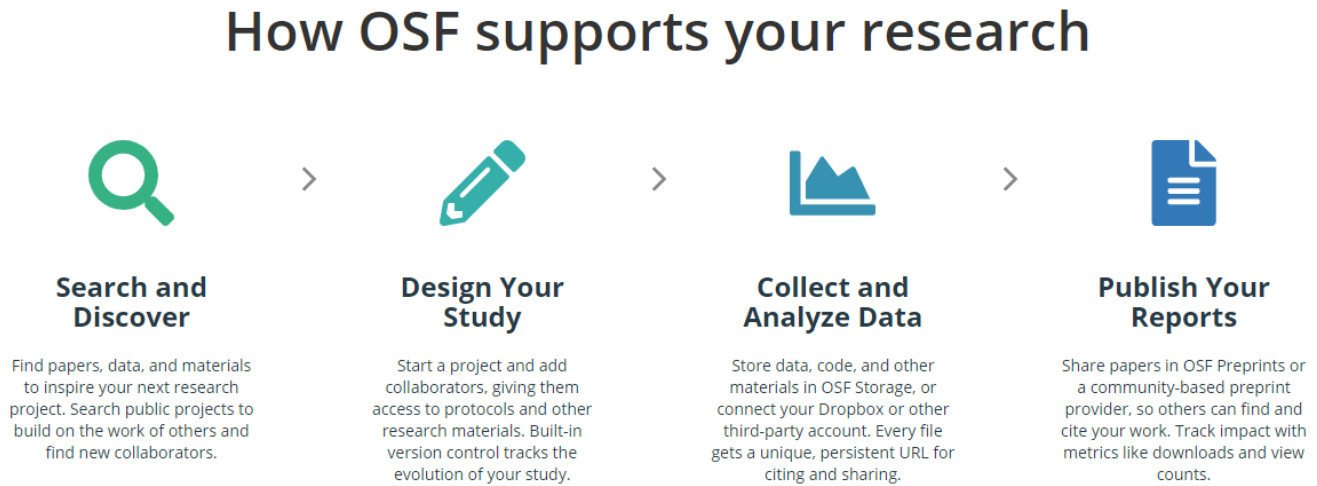
Sharing and Preservation

UKnowledge

UK's institutional repository that archives and provides access to UK community's research outputs

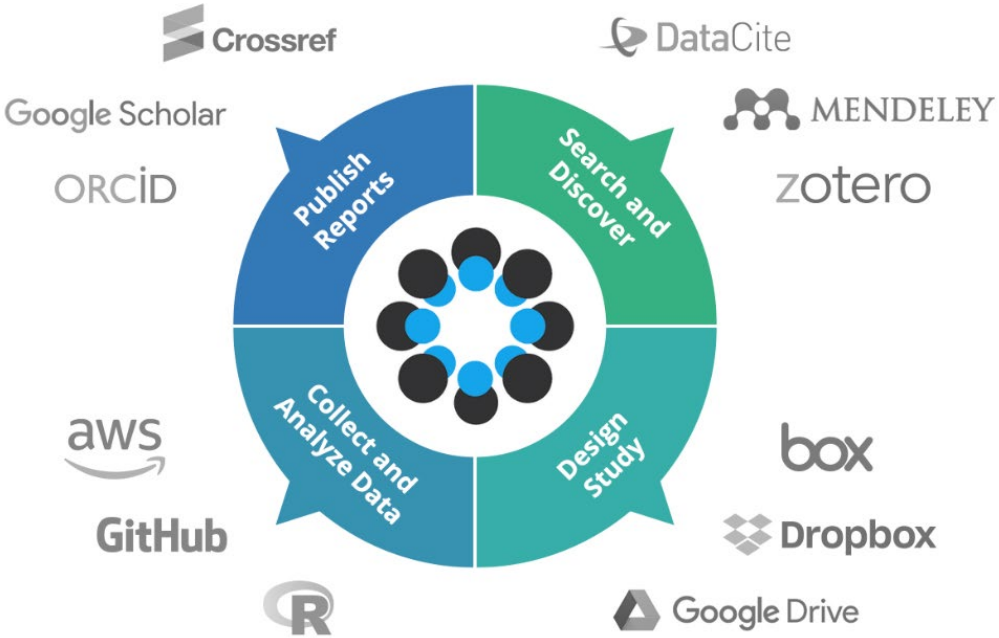
https://uknowledge.uky.edu/

OPEN SCIENCE FRAMEWORK (OSF)



Open Science Framework. Our Projects. OSF. https://cos.io/our-products/osf/

OPEN SCIENCE FRAMEWORK



Open Science Framework. Our Projects. OSF. https://cos.io/our-products/osf/

DATA & COPYRIGHT

> *Raw data generally falls into the category of factual information and is not afforded copyright protection.*

> *Expressions of data (e.g., charts and graphs) may be copyrightable*

Caitlyn Bakker, University of Minnesota Libraries

Finding Data Repositories

SUBJECT SPECIFIC REPOSITORIES

- There are established research data repositories for certain disciplines, e.g., GenBank for research data about DNA and a repository at Syracuse University especially for qualitative data.
- There is an online directory for finding appropriate research data repositories for you: Registry of Research Data Repositories (re3data.org)

Recommended Practices

FAIR Guiding Principles

To increase the user-friendliness of research data, a group of researchers have proposed the FAIR Guiding Principles in 2016: https://doi.org/10.1038/sdata.2016.18

F= Findable
A= Accessible
I= Interoperable
R= Reusable

FAIR Guiding Principles

F (Findable), e.g., a DOI (digital object identifier) assigned to the dataset to prevent a broken link in the future

A (Accessible), e.g., dataset available from an online repository that records descriptions about the data and holds the data files

I (Interoperable), e.g., dataset available in different file formats, accompanied by rich metadata about how the dataset can potentially be integrated with other datasets

R (Reusable), e.g., availability of terms, license, and descriptions regarding how the dataset can be reused

Useful Links

UK Libraries Resources

- DataOne Data Life Cycle
  - https://www.dataone.org/data-life-cycle
- Data Sources
  - https://libguides.uky.edu/datasources/home
- DMPTool
  - https://dmptool.org/
- FAIR Guiding Principles
  - https://doi.org/10.1038/sdata.2016.18

UK Libraries Resources

- Mapping and GIS Services
  - https://libguides.uky.edu/gis_mapping
- Open Science Framework
  - https://osf.io/
- Readme file template
  - https://bit.ly/2NPeu4z
- Research Data Services
  - https://libguides.uky.edu/research_data

UK Libraries Resources

- Registry of Research Data Repositories (re3data)
  - http://re3data.org/
- UK Data Retention & Ownership Policy
  - https://www.research.uky.edu/office-research-integrity/university-kentucky-data-retention-ownership-policy
- UKnowledge
  - https://uknowledge.uky.edu/

IMAGE CREDITS

Flaticons:

- Eucalyp; https://www.flaticon.com/authors/eucalyp
- Linector; https://www.flaticon.com/authors/linector
- Mynamepong; https://www.flaticon.com/authors/mynamepong
- Freepik; https://www.flaticon.com/authors/freepik
- Smashicons; https://www.flaticon.com/authors/smashicons
- Pause08; https://www.flaticon.com/authors/pause08
- Smartline; https://www.flaticon.com/authors/smartline