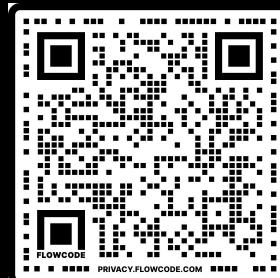


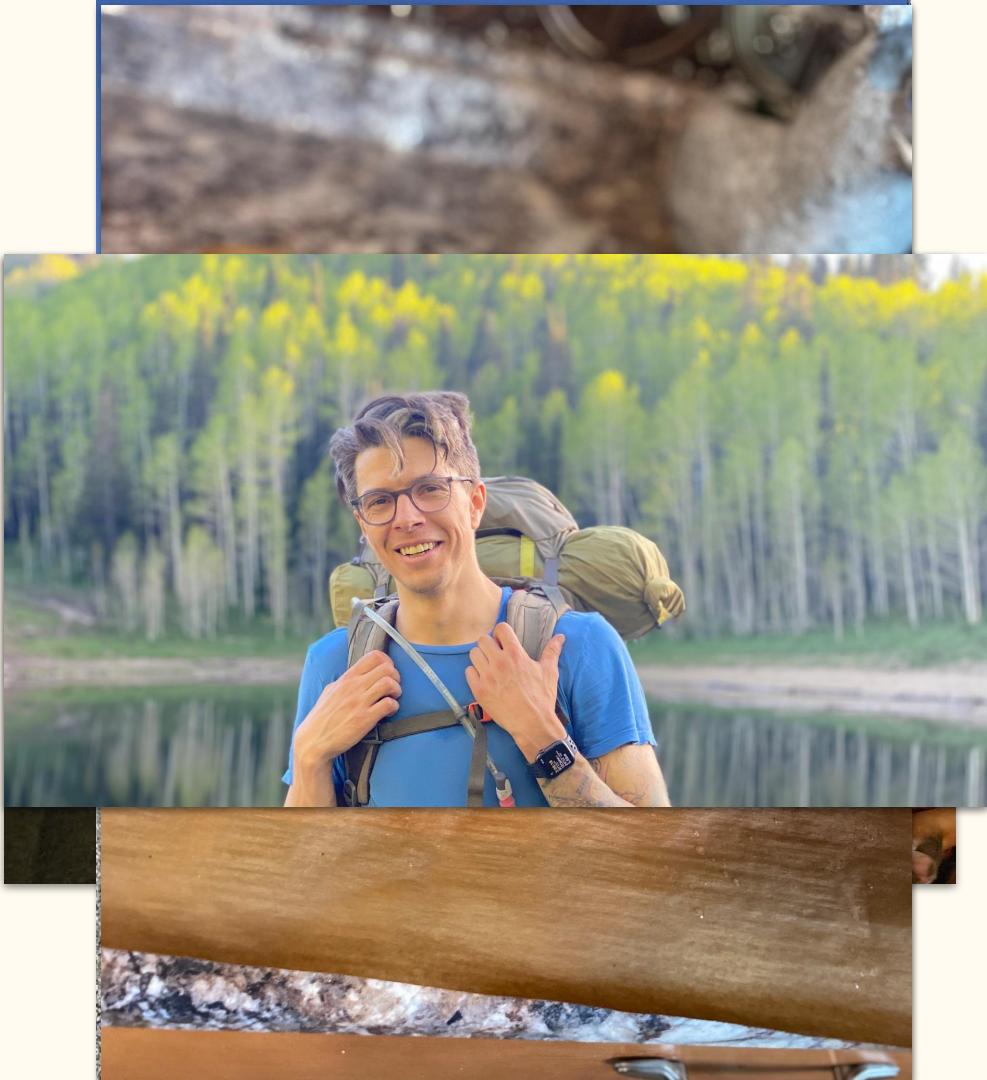
Maximizing Research Impact: the Role of Research Data Librarians

By: Keith Araneo-Yowell (He/Him)



Introduction

- Long difficult name
- University of Utah Alumnus
- Dogs
- Bikes
- Food
- Utah



Objective

Supporting the University of Utah's Research Community

- Facilitate the management of research data throughout its lifecycle
 - Broaden the reach of the UofU's scholarly contribution
-

Essential Attributes of Research Data

- “Recorded, factual material commonly accepted in the scientific community as necessary to validate research findings” (NIH 2023).
- Something essential to the research community and not just individual researchers (NSF 2023)
- “Materials generated or collected during the course of conducting research”. (NEH 2023)



The Importance of Data Management

- Long-term preservation
- Other researchers
- Funding agencies require it

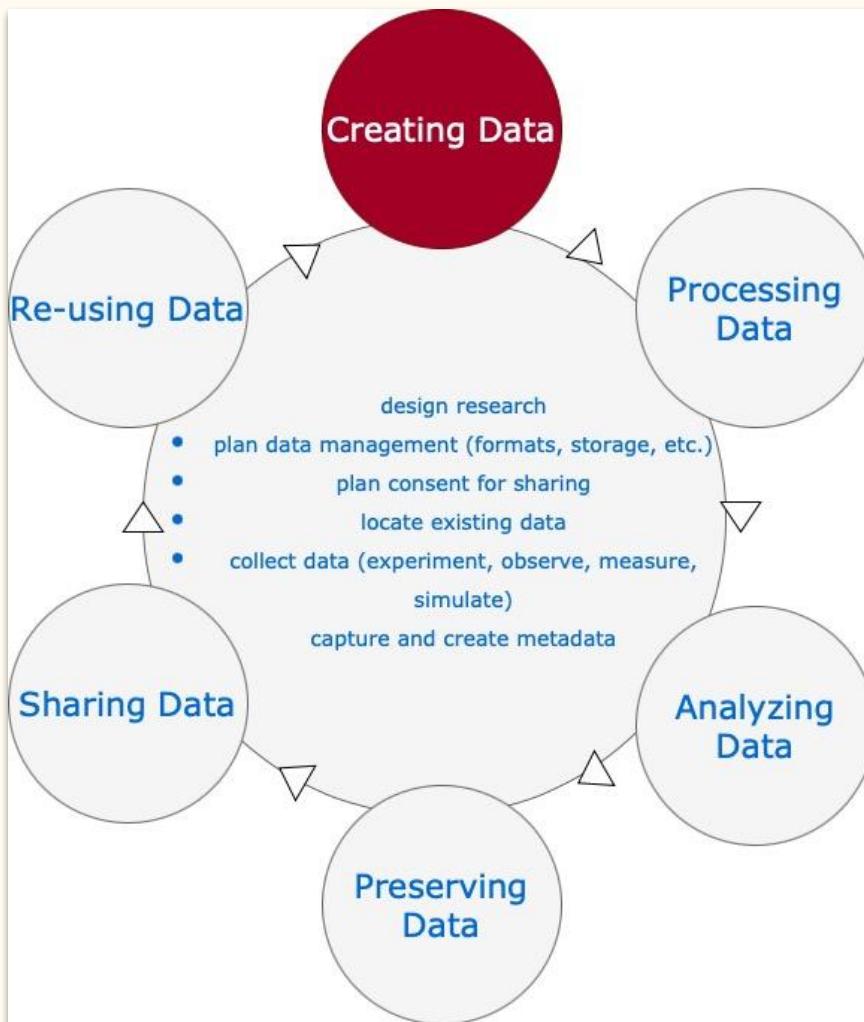


Data Management Plans (DMPs)

What are DMPs?

“A month in the laboratory can often save you an hour in the library”
-Frank Westheimer

- Formal documents that describe the data produced during a research project
- Detailed description of all aspects of the data management that will take place during the research data lifecycle

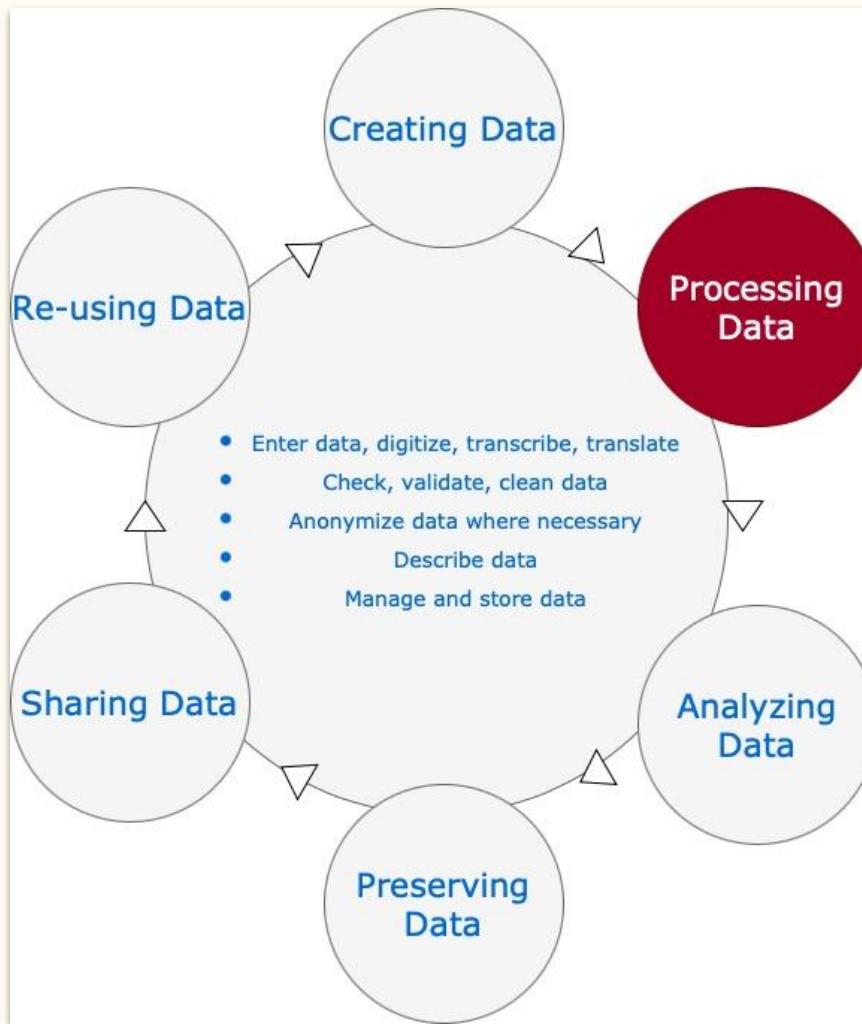


Creating Data

- Not every stakeholder will be considering the entirety of the data lifecycle

“Communication is the universal solvent”

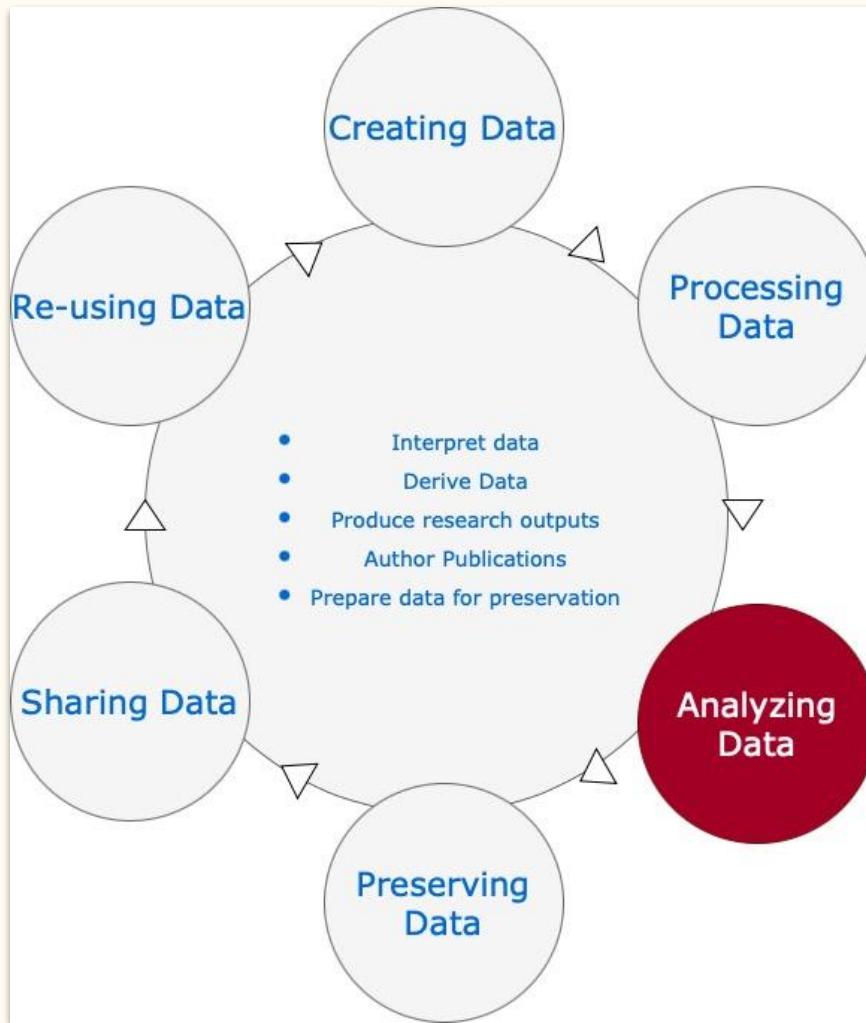
12	Glassware washer records are attached with BMR	<input type="checkbox"/>	CLICK TO SIGN
13	Autoclave records attached to BMR	<input type="checkbox"/>	CLICK TO SIGN
14	DPO records attached to BMR	<input type="checkbox"/>	CLICK TO SIGN
15	Depryg. Tunnel Record attached to BMR	<input type="checkbox"/>	CLICK TO SIGN
16	OPC data attached	<input type="checkbox"/>	CLICK TO SIGN
17	Interventions log reviewed	<input type="checkbox"/>	CLICK TO SIGN
18	filter COA and integrity Report attached with BMR	<input type="checkbox"/>	CLICK TO SIGN
19	stage-wise yield and reconciliation is written	<input type="checkbox"/>	CLICK TO SIGN
20	all details of visual inspection are mentioned correctly	<input type="checkbox"/>	CLICK TO SIGN



Librarian's Role During Data Processing

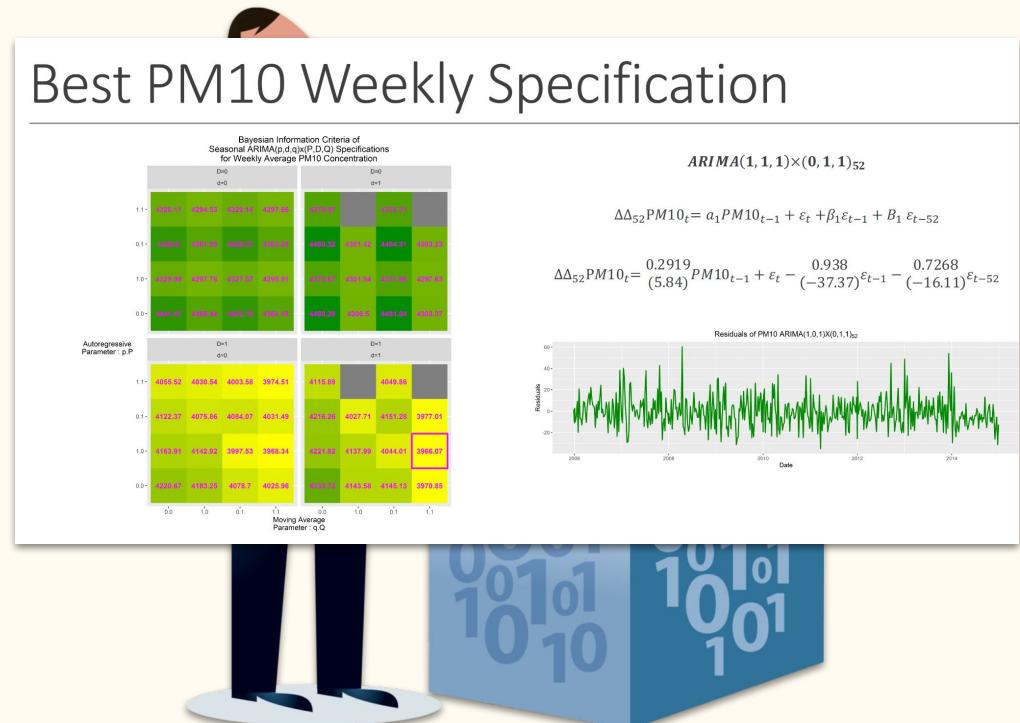
- Educating about data management best practices
- Selecting software and methods
- Data-cleaning
- Ensuring documentation for reproducibility of data processing

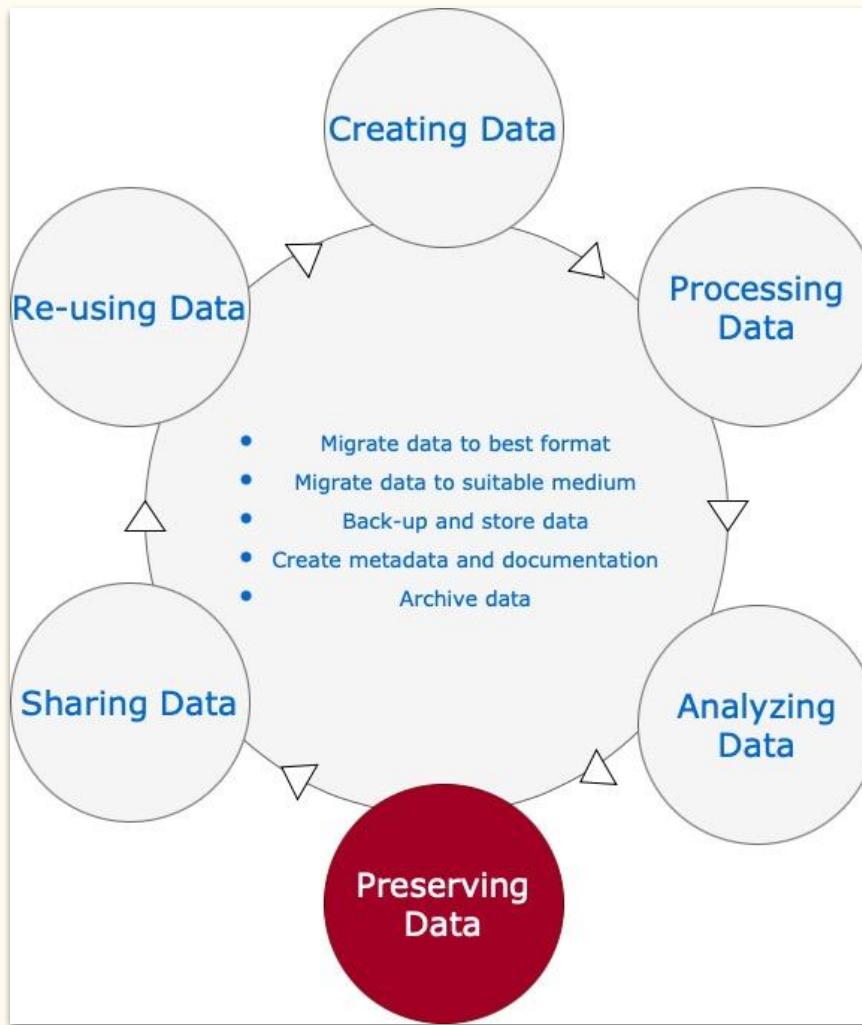




Librarian's Role During Data Analysis

- Guidance on data analysis tools and software
- Support on statistical methods and data visualization
- Provide training on data analysis best practices
- Lay groundwork for long-term preservation





Preserving Data

- **Authenticity**

- “The quality of being genuine, not a counterfeit, and free from tampering, and is typically inferred from internal and external evidence, including its physical characteristics, structure, content, and context.”

- **Integrity**

- The extent to which a digital object has been free from corruption over time or in transit
- The similarity of the dataset in-hand to the dataset at the time it was created.



Preserving Data

- Data are at risk if not properly preserved
 - Neglect
 - Bit Rot
 - Obsolescence



Role of the Librarian in Data Preservation

- Assist with Documentation and Versioning
- Finding trustworthy repositories
- Help create data sharing policies
- Ensure data is stored in proper format(s)
- Backup/Security/Encryption
- Metadata



Metadata: Data about Data

- Fashionable
- Necessary in a lot of cases
- The more you have the better
- Easier to find
- Live to see another day
- Unique and persistent identifier
- Universally understood and utilized
- It can get messy if you don't have it

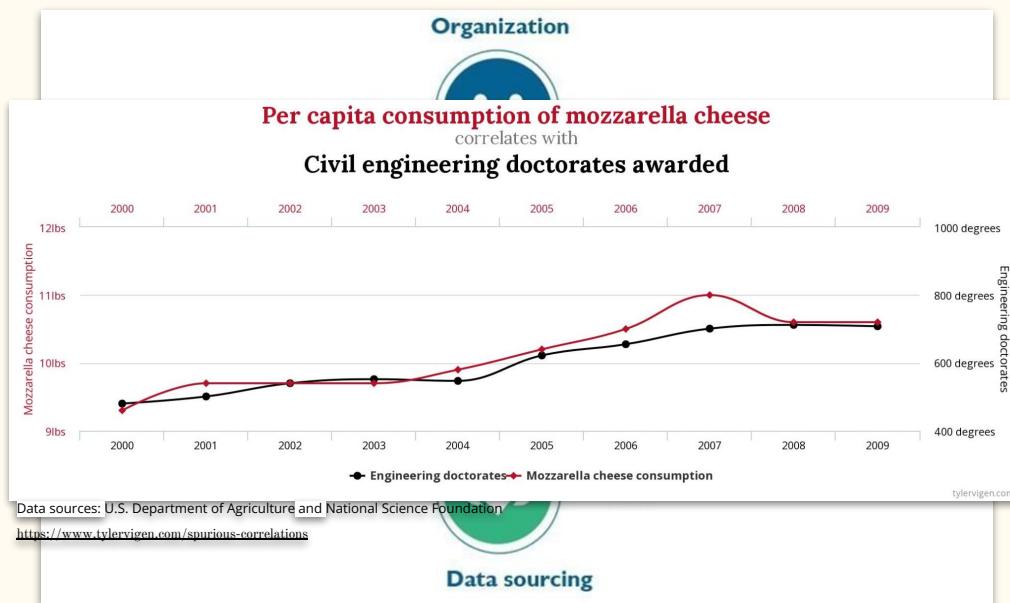
Name:	A token appended to the URI of a DCMI namespace to create the URI of the term.
Label:	The human-readable label assigned to the term.
URI:	The Uniform Resource Identifier used to uniquely identify a term.
Definition:	A statement that represents the concept and essential nature of the term.
Type of Term:	The type of term: property, class, datatype, or vocabulary encoding scheme.

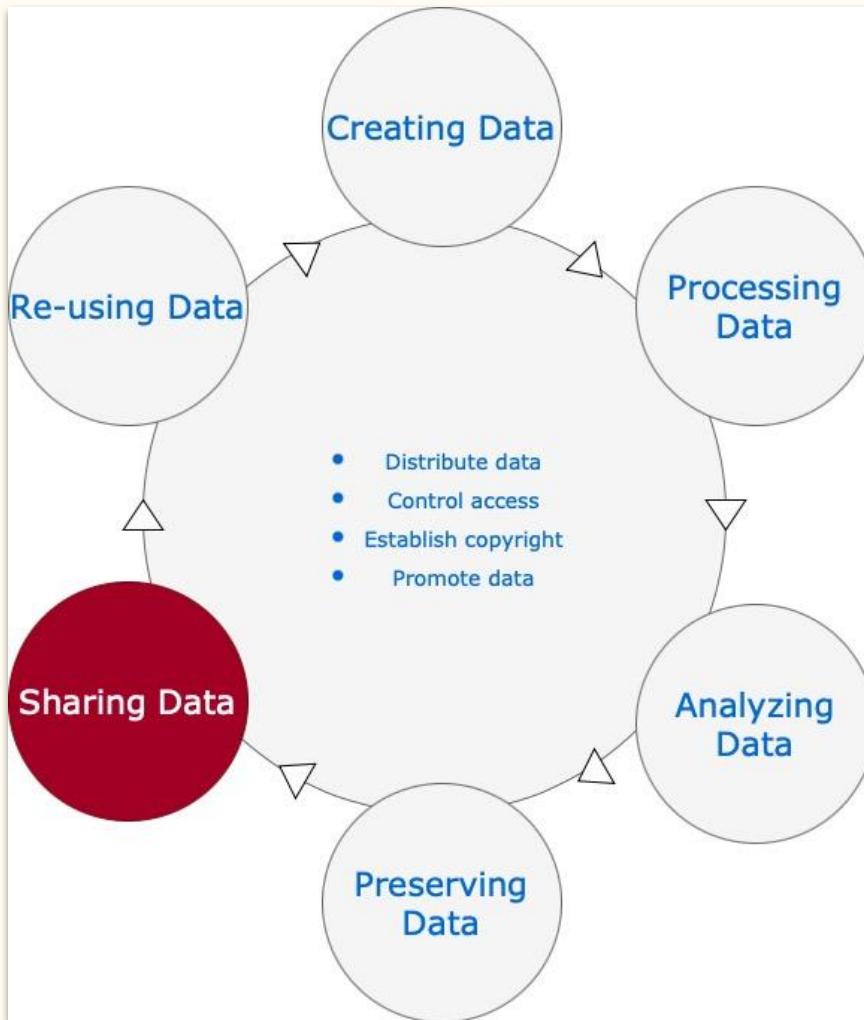
(DCMI 2021)



My Experience with Data Preservation

- Good Data
 - easily-validated findings
 - Reproducible results
- Bad Data
 - More cumbersome to make useable
 - Untrustworthy





Benefits of Sharing Data

- Encourages open scientific discourse
- Supports verification and replication of results
- Invites new questions
- Enables collaboration
- Puts forward valuable teaching resources
- Reduces Cost
- Safeguards against fraud
- Improves visibility
- Preserves for future use
- Contributes to broader community

“If I have seen further [than others], it is by standing on the shoulders of giants.”

- Isaac Newton

The Role of the Librarian in Sharing Data

1.12 - TEN YEAR ACCIDENT / INCIDENT OVERVIEW BY CALENDAR YEAR (January-December) You Chose Months January Through December Run Date: Thu, Apr 20, 2023 Reporting Level:... ALL - - - Railroad Group:... ALL RAILROADS UTAH, WEBER COUNTY ALL RAILROADS SELECTED													
Category	CY 2013	CY 2014	CY 2015	CY 2016	CY 2017	CY 2018	CY 2019	CY 2020	CY 2021	CY 2022	% Change From CY 2021 to CY 2022	% Change From CY 2013 to CY 2022	Total For CY 2013 to CY 2022
Number of railroads included	1	1	1	1	1	1	1	1	1	1	.	.	.
TOTAL ACCIDENTS/INCIDENTS 1/	8	2	5	5	12	16	20	16	9	10	11.111	25.000	103
--- Total fatalities	.	.	1	1	2
--- Total nonfatal conditions	3	3	1	4	3	4	4	9	1	5	400.000	66.667	37
--- Employee on duty deaths
--- Nonfatal EOD injuries	3	2	1	4	1	2	.	4	1	3	200.000	0.000	21
--- Nonfatal EOD illnesses
--- Total employee on duty cases	3	2	1	4	1	2	.	4	1	3	200.000	0.000	21
--- Cases with days absent from work	3	.	1	2	1	.	.	1	1	2	100.000	-33.333	11
--- Trespasser deaths, not at HRC	1	1
--- Trespasser injuries, not at HRC	1	3	2	.	2	.	.	8
--- Trespasser incidents, not at HRC	1	3	1	.	2	.	.	7
--- Passengers kld in train accs or crossing incidents
--- Passengers inj in train accs or crossing incidents
--- Passengers kld in other incidents
--- Passengers inj in other incidents
TRAIN ACCIDENTS (Not at Grade-Crossings)	3	1	2	1	9	11	16	6	7	5	-28.571	66.667	61

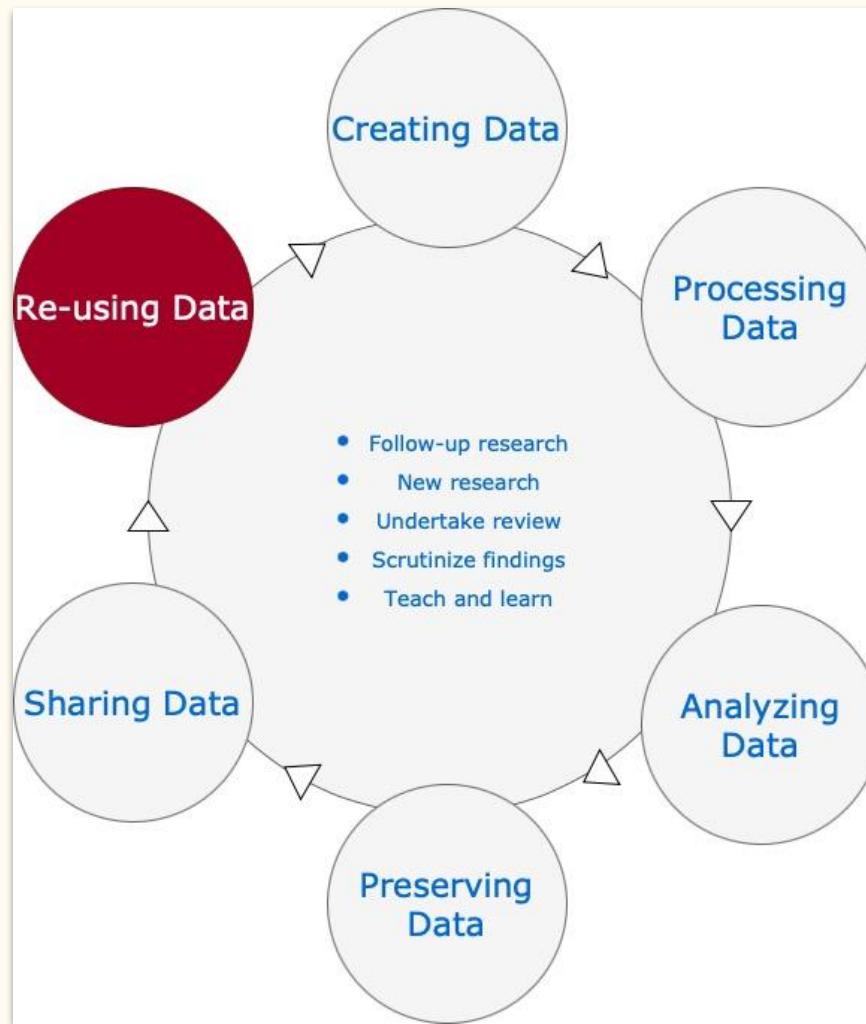
Federal Railroad Administration. (n.d.). Safety Data Portal. Retrieved April 20, 2023, from <https://safetydata.fra.dot.gov/OfficeofSafety/publicsite/Query/TenYearAccidentIncidentOverview.aspx>

- Have appropriate measures to protect the **privacy and confidentiality** of research participants
- **Data ownership:** Licensing and attribution requirements
- **Quality and Integrity:** have the data been evaluated for validity and longevity?
- **Format and Documentation:** are the data saved in multiple legacy formats in different locations? Have the data elements been properly documented for use in a data dictionary?
- **Re-use:** do the data represent value to the research community?
- What are the mechanisms for making data accessible?

F.A.I.R

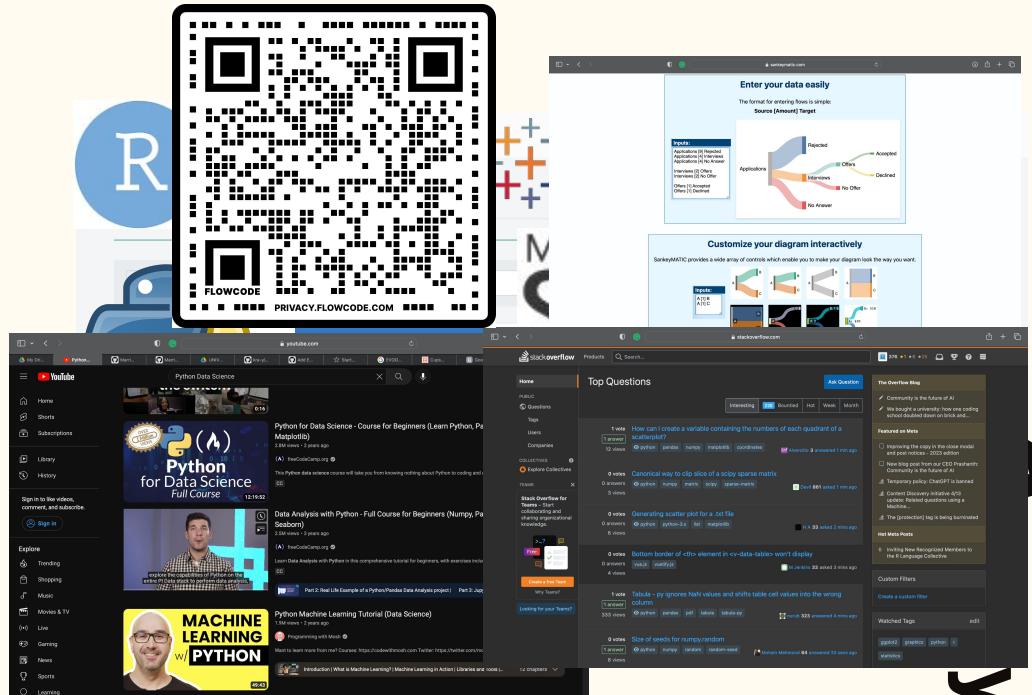
- Findable
- Accessible
- Interoperable
- Reusable





Key Qualifications

- Strong knowledge of research data management principles and practices
- Experience using data management tools and software, such as data repositories, data visualization tools, and AI language models
- Proficiency in programming languages commonly used in data analysis and manipulation, such as R and Python



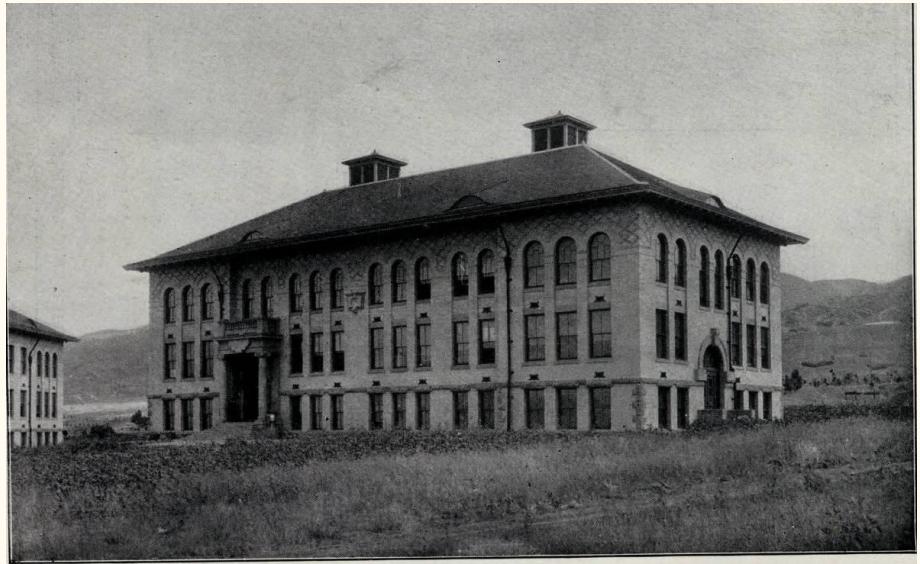
Key Qualifications (Ctd.)

- Strong Communication and interpersonal skills to collaborate effectively with researchers and other stakeholders to meet their data needs
- Knowledge of relevant laws and regulations related to data management, such as intellectual property laws, privacy laws, and funding agency requirements

The screenshot shows the DMPTool web application. At the top, there is a navigation bar with links: My Dashboard, Create Plan, Funder Requirements, Public DMPs, and Help. On the far right, it shows the user's name 'Keith Araneo-yowell' and a language dropdown. Below the navigation, the title 'Geospatial Analysis of U.S. Rail System Health' is displayed. The main content area has tabs: Project Details (selected), Collaborators, Write Plan, Research outputs, Finalize, and Download. Under 'Project title', the text 'Geospatial Analysis of U.S. Rail System Health' is entered. A checkbox labeled 'mock project for testing, practice, or educational purposes' is checked. The 'Project abstract' section contains two paragraphs of text. To the right, a sidebar titled 'Select Guidance' offers help in writing the plan, mentioning organizations like DMPTool, and provides a link to see the full list of organizations. A 'Save' button is located at the bottom right of the sidebar.

Why the Marriott Library/University of Utah?

- I want to give back to an institution that's always been part of my life
- In addition to helping others with their research, I hope to conduct studies of my own
- I want to continue learning
- I want to teach
- I want to interact with people who spend time at libraries
- Legacy
- Tenure

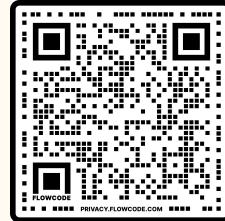


By University of Utah - Utah Yearbook, PD-US, <https://en.wikipedia.org/w/index.php?curid=23989223>

In Conclusion...

- Thank you so much for your consideration
- Questions?

The End?



References

- Johnson, K., Tibbo, Helen. *Research Data Management and Sharing*[MOOC]. Coursera. <https://www.coursera.org/learn/data-management>
- Dublin Core Metadata Initiative. (2012). DCMI metadata terms. Retrieved from <http://dublincore.org/documents/dcmi-terms/#elements-contributor>
- Crampon, Jean E. 1988. Murphy, Parkinson, and Peter: Laws for librarians. Library Journal 113. no. 17 (October 15), p. 41.
- National Institutes of Health (NIH). (2023) *NIH Data Management and Sharing policy*. Bethesda, MD: National Institutes of Health. Retrieved from <https://sharing.nih.gov/data-management-and-sharing-policy/about-data-management-and-sharing-policies/data-management-and-sharing-policy-overview#fin>
- National Endowment for the Humanities (NEH). (2015). *Data management plans for NEH office of digital humanities proposals and awards*. Washington, DC: National Endowment for the Humanities. Retrieved from <https://www.neh.gov/divisions/odh/grant-news/announcing-new-grant-program-digital-humanities-implementation-grants>
- ChatGPT. (2023, April 21). ChatGPT: A large language model. <https://openai.com/>. Accessed on April 21, 2023.
- DMPTool. (n.d.). About the DMPTool. Retrieved from <https://dmptool.org/about>
- J. Willard Marriott Library. (2022, November 29). In *Wikipedia*. https://en.wikipedia.org/wiki/J._Willard_Marriott_Library
- Center for Research Libraries. (n.d.). TRAC: Trustworthy Repositories Audit & Certification. Retrieved from <https://www.crl.edu/archiving-preservation/digital-archives/metrics-assessing-and-certifying/trac>
- Government of Canada. (n.d.). Digital Preservation Toolkit. Heritage Information Network. Retrieved from <https://www.canada.ca/en/heritage-information-network/services/digital-preservation/toolkit.html>
- Federal Railroad Administration. (n.d.). Safety Data Portal. Retrieved April 20, 2023, from <https://safetydata.fra.dot.gov/OfficeofSafety/publicsite/Query/TenYearAccidentIncidentOverview.aspx>
- Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., ... & Mons, B. (2016). The FAIR guiding principles for scientific data management and stewardship. *Scientific Data*, 3(1), 1-9. <https://doi.org/10.1038/sdata.2016.18>
- Lynch, C. A. (2003). Institutional Repositories: Essential Infrastructure for Scholarship in the Digital Age. Council on Library and Information Resources. Retrieved April 20, 2023, from <https://www.clir.org/pubs/reports/pub92/lynch/>
- Computero. (n.d.). '90s movie "Office Space" goes reality [Digital image]. Retrieved April 21, 2023, from <https://www.computero.com/90s-movie-office-space-goes-reality/>
- Flowcode. (n.d.). Home page. Retrieved April 23, 2023, from <https://www.flowcode.com>
- Google LLC. (n.d.). Google Maps. Retrieved April 21, 2023, from <https://www.google.com/maps/>
- Vigen, T. (n.d.). Spurious Correlations. Retrieved April 21, 2023, from <https://www.tylervigen.com/spurious-correlations>
- Araneo-Yowell, Keith. Econometric Analysis and Forecasting of Particulate Matter Concentration, Temperature, and Precipitation: Comparative Forecasting Methods. Master's Thesis. University of Utah, 2016