CAREful FAIRness in the Data Lifecycle - Intro to Ethical Research Data

CAREful FAIRness in the Data Lifecycle - Intro to Ethical Research Data (link to web version of syllabus)

Session 0 – Background

Description

This section provides a background to the main themes of this syllabus: the research data lifecycle and how to navigate the different stages with both the principles of FAIR – Findable, Accessible, Interoperable, and Reusable and of CARE – Collective Benefit, Authority to Control, Responsibility, and Ethics in mind.

Questions

- What is the Research Data Lifecycle?
- What is FAIR?
- What is CARE?
- How do they differ and how do they support each other?
- Why are FAIR and CARE both are important, especially when it comes to data about people/communities?

Figure 1 Harvard Biomedical Research Data Lifecycle

Resources

- CARE Principles for Indigenous Data Governance (September 2019)
- Data Curation Network (DCN) CARE Primer (2023)
- FAIR Principles (2016)
- Harvard Biomedical Research Data Lifecycle
- Bossaller, J. and Million, A.J., The research data life cycle, legacy data, and dilemmas in research data management in *Journal of the Association for Information Science and Technology*, 74(6), 701–706 (March 2022)

Session 1 – Plan and Design

The CARE Principles speak to how scientific data are used in ways that are purposeful and oriented towards enhancing wellbeing of people. The CARE Principles are likely to find expression across the data lifecycle from collection to curation, from access to application, with implications and responsibilities for a broad range of entities from funders to data users.[i]

Description

In this section, we'll enter the research data lifecycle at the plan and design stage, where you create your plan for your research data – the planning, storage, and sharing, among other things. Instead of the basics of creating a budget, creating a data management plan (DMP), and defining roles and responsibilities, we must consider how to weave or make actionable ("operationalize") the CARE and FAIR principles. When your research data concerns people and/or communities, particularly those who are marginalized/minoritized, "research with people instead of about people" enhances the collective benefit and make you, the researcher/research group, responsible to those people.

Questions

- When crafting your DMP, how do/how can you plan to incorporate both CARE (the principles) and care (in terms of ethics)?
- What world will you make with your research data?[ii]

Resources

- Carroll, S.R., Herczog, E., Hudson, M. *et al.,* Operationalizing the CARE and FAIR Principles for Indigenous data futures in *Sci Data* **8**, 108 (April 2021)
- Belarde-Lewis, M., Littletree, S., Braine, I.S., Srader, K., Guerrero, N. and Palmer, C.L., Centering Relationality and CARE for Stewardship of Indigenous Research Data in *Data Science Journal*, 23: 32, pp. 1–16 (May 2024)
- de la Bellacasa, M. P., Matters of care in technoscience: Assembling neglected things in *Social Studies of Science*, *41*(1), 85-106 (December 2010)
- A primer on an intersectional approach to data (2018)

- Boenig-Liptsin, M., Tanweer, A., & Edmundson, A., Data Science Ethos Lifecycle: Interplay of Ethical Thinking and Data Science Practice in *Journal of Statistics and Data Science Education*, 30(3), 228–240 (July 2022)
- Barrett, T., Okolo, C. T., Biira, B., Sherif, E., Zhang, A., & Battle, L., African Data Ethics: A Discursive Framework for Black Decolonial AI in *Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency*, 334–349 (June 2025)
- A Toolkit for Centering Racial Equity Within Data Integration (2020)

Section 2 – Collect and Create

Differences in the etymological roots of the terms data and capta make the distinction between constructivist and realist approaches clear. Capta *is "taken" actively while* data *is assumed to be a "given" able to be recorded and observed. From this distinction, a world of differences arises. Humanistic inquiry acknowledges the situated, partial, and constitutive character of knowledge production, the recognition that knowledge is constructed, taken, not simply given as a natural representation of pre-existing fact.*[iii]

Description

The collection and creation stage of the lifecycle is the first moment when both the CARE and FAIR principles would be implemented. While research data created or collected is done so to answer a question or provide new/more information, this work isn't done in a vacuum and the implications for the "subject(s)", the data creator/collector(s), and the data user/re-user must be evaluated. Additionally, documentation is an important component in this stage of the lifecycle. Without proper and full documentation for the data collected/created, discovery is limited, it is difficult for other researchers to reuse or reproduce the data, and importantly, the context is lost.

Questions

- Have you considered the best method(s) of data collection that allows for the gathering of relevant data and not unnecessary data or data that may be useful for a future research project but not within the scope of the current research and scope of consent?
- If gathering consent from a potential participant, have you thoroughly explained what kind of data will be collected, how it will be collected, and how it will be potentially used?
- Is there an option to opt-out of certain activities or moments of participation?
- When organizing collected data, is there documentation explaining and/or supporting the reasoning or method for organizing?
- If conventional/traditional methods of documentation fail to adequately describe the content and context of the data, how will you go about find the best way to document or explain the gap/issue with the available documentation method(s)?
    - Are there specific protocols or vocabularies that could be used?

Resources

- Leon, Sharon M., The Peril and Promise of Historians as Data Creators: Perspective, Structure, and the Problem of Representation in *Journal of Slavery and Data Preservation* 4, no. 5, 1-17 (December 2023)
- Brown, N. M., Mendenhall, R., Black, M., Moer, M. V. *et al.*, In Search of Zora/When Metadata Isn't Enough: Rescuing the Experiences of Black Women Through Statistical Modeling in *Journal of Library Metadata*, *19*(3–4), 141–162 (August 2019)
- Henrietta Lacks, the Tuskegee Experiment, and Ethical Data Collection: Crash Course Statistics #12 [video] (April 2019)
- Gebru, T., Morgenstern, J., Vecchione, B., Vaughan, J. W. *et al.*, Datasheets for Datasets (March 2018)
- Pushkarna, M., Zaldivar, A., & Kjartansson, O., Data Cards: Purposeful and Transparent Dataset Documentation for Responsible AI (April 2022)

Section 3 – Analyze and Collaborate

Interoperable: The data usually need to be integrated with other data. In addition, the data need to interoperate with applications or workflows for analysis, storage, and processing.[iv]

Description

Once the data creation/collection stage has been completed (at least, initially), the data must be analyzed to answer the research question(s) and/or provide additional information for research – whether as a report, visualization, or in some tabular format, to name a few examples. In this stage, we should also consider the best method(s) of documenting the analysis and the steps taken before the research data is ready for analysis. This documentation is not only important for future use and reproducibility but also for sharing out with others in your research group, if working collaboratively.

Questions

- From "7 Key Ethical Considerations":
    - Am I approaching this analysis with objectivity, and have I considered all potential sources of bias?
    - Can other researchers reproduce my results based on the information I've provided?
- From Rawson and Munoz (2016) "Against Cleaning":
    - With those communities [impacted by the data] in mind and even in dialogue, we would ask, what are the concepts that structure this data?
    - And how can this data, structured in this way, point to other people's data?
- What distinguishes your "messy" data from "clean" data? Has that difference been documented?
- Can the "messy" data you've encountered be seen in a new way?

Resources

- Çay, D., Rodighiero, D., & Zhang, W., Visualizing as a Form of Collective Care in *Nightingale* (June 2025)
- Barter, R., A quick guide to developing a reproducible and consistent data science workflow (March 2019)
- Danchev, V., Reproducible Data Science with Python (2021)
- Reproducibility 4 Everyone
- Framework for Open and Reproducible Research Training (FORRT) List of Open Science Syllabi
- D'Ignazio, C., & Klein, L., Chapter Two: On Rational, Scientific, Objective Viewpoints from Mythical, Imaginary, Impossible Standpoints in *Data Feminism* (November 2018)
- Hicks, C., When We Miss Missingness (July 2022)
- The Turing Way (2024)
- Rawson, K. and Muñoz, T., Against Cleaning (July 2016)
- Nowviskie, B., capacity through care (February 2016)
- 7 Key Ethical Considerations in Research Data Analysis and Interpretation (September 2024)

Section 4 – Evaluate and Archive

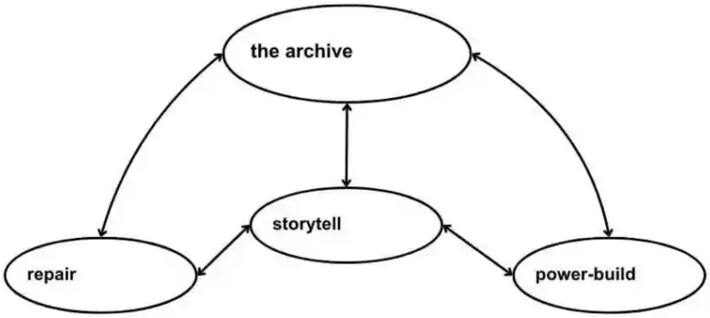

Figure 2 Nasir Montalvo's the diagram (Montalvo, 2023)

We further imagine a world where their narratives, whole or fragmented, are accessible to those who prioritize protection.[v]

Description

As your research project concludes, you need to evaluate your data to ensure they are complete, accurate, and well-documented. Additionally, you will need to select where and how the data will be archived. While this is another opportunity to focus on the completeness of your data and related documentation, a focus should be placed to how you complete your data and documentation – looking to more inclusive description.

Questions

- From Montalvo (2023) "bad tats, jesus christ, lemons; everything is archival"
- **NOTE:** all "questions to consider" for each section are valuable in considering what and how you archive your research data, particularly data about people/communities
    - what are your identities? how do you show up in this work?
    - who is your audience? what are their access needs?
    - what harms has the community you're archiving faced?

    - what through lines from repair can be examined and subverted to power build in your context?
- How can you provide enough metadata for your data that not only provides technical insight but also non-technical (historical, social, etc.) insight?

Resources

- King, S., Build Your Archive
- Montalvo, N. A., bad tats, jesus christ, lemons; everything is archival
- Webb, S., "Digital Archives in Communities – Practice and Preservation" : A Summary (or at least an attempt) (August 2018)
- Uncovering hidden narratives in Georgia Tech's history
- CURATE(D) Steps and Checklists (2018)
- Shankar K., Jeng W., Thomer A., Weber N., and Yoon A., Data curation as collective action during COVID-19 in *J Assoc Inf Sci Technol*.; 72: 280–284 (September 2020)
- Rogers, J. and Rae, R., Ethical Data Considerations for Engaging in Reparative Archival Practice (November 2023) in *Works of the FIU Libraries*. 141

Section 5 – Share and Disseminate

Too often in science, we put the onus for ethical behavior on individuals, but an individual cannot ultimately behave ethically within an unethical framework[vi]

Description

Once research is completed, many researchers look to share out their findings – many funding agencies and journals require it – and have to determine how much access to allow to others, certain agreements that satisfy institutional/organizational requirements, as well as proper stewardship of the research data. With the prevalence of AI and data concerning vulnerable populations, CARE and community-driven oversight of data helps narrow the "moral distance"[vii] that occurs when data is shared. While there are many benefits to data being Accessible, oceanographer and Native Hawaiian Rosie Alegado remarks, "'Accessible' means somebody can access it but you are still allowed to regulate it" (cited in Ravindran, 2024).

Questions

- How can a participant(s) and their data be protected from ways of sharing/reuse that are new or may not exist at the time of giving consent?
- Is it possible to reduce harm for individuals/communities while contributing to the greater benefit that may result from the research data?
- What does "openness" and "transparency" mean to you when considering sharing research data? Would the definitions change as the context of the research and the data changes?

Resources

- Evans, E.A., Delorme, E., Cyr, K. *et al.*, A qualitative study of big data and the opioid epidemic: recommendations for data governance in *BMC Med Ethics* **21**, 101 (October 2020)
- Suchikova, Y., Nazarovets, S., Extending the CARE Principles: managing data for vulnerable communities in wartime and humanitarian crises in *Sci Data* **12**, 420 (March 2025)
- Abebe, R., Aruleba, K., Birhane, A., Kingsley, S. *et al.*, Narratives and Counternarratives on Data Sharing in Africa in *Conference on Fairness, Accountability, and Transparency (FAccT '21)* (March 2021)
- Ravindran, S., Open with Care in *Science* (October 2024)
- Mitra, A., Open with care: transparency and data sharing in civically engaged research in *Politics, Groups, and Identities*, 1–26 (December 2023)
- Huvila, I., & Sinnamon, L. S., When data sharing is an answer and when (often) it is not: Acknowledging data-driven, non-data, and data-decentered cultures in *Journal of the Association for Information Science and Technology*, 75(13), 1515–1530 (October 2024)
- Exploring the Ethical Imperative for Data Sharing (September 2015)
- Villegas-Galaviz, C., Martin, K., Moral distance, AI, and the ethics of care (March 2023)

Session 6 – Publish and Reuse

Description

In fulfilling requirements for publishing and reuse, researchers much choose where to publish their findings and where their data and related documentation will live for others to reuse. Ensuring that your persistent identifiers (ORCID, DOI, etc.) – for yourself and your work – are included to make yourself and your data Findable and if your data belongs in a domain-specific, generalist, or institutional repository. Additional considerations include *if* it is feasible to share your data, that your data includes enough context to explain the origins of your data, and that you've done the work to prevent/mitigate misuse of your data. It is always important to remember that just because you have the ability to/capacity to share, it doesn't mean you should.

Questions

- Returning to the CARE Principles:
    - Is there a **Collective benefit** for sharing parts or all of your data?
    - Is there an agreement with the participants/communities/gatekeepers on the **Authority to control** where and how the data is shared? Are there any permissions or access restrictions? If so, what are they and how can they be accommodated?
    - How are you as the principal investigator/data collector/data analyst/curator **Responsible** once the data has been published and/or is shared/will be shared?
    - Is what (and how) you're sharing **Ethical**? Are "best practices" and similar considerations enough?

Resources

- Yoon, A., Data reusers' trust development in *Journal of the Association for Information Science and Technology*, 68: 946-956 (June 2016)
- Khodyakov, D., Mikesell, L., Schraiber, R., Booth, M., and Bromley, E., On using ethical principles of community-engaged research in translational science in *Translational Research*, *171* (May 2016)
- Responsible Datasets in Context
- Pasquetto, I. V., Cullen, Z., Thomer, A., and Wofford, M., What is research data "misuse"? And how can it be prevented or mitigated? in *Journal of the Association for Information Science and Technology*, 75(12), 1413–1429 (July 2024)
- Post45 Data Collective
- Huvila, I., & Sinnamon, L. S., When data sharing is an answer and when (often) it is not: Acknowledging data-driven, non-data, and data-decentered cultures in *Journal of the Association for Information Science and Technology*, 75(13), 1515–1530 (October 2024)
- O'Brien, M., Duerr, R., Taitingfong, R., Martinez, A., *et al.*, Earth Science Data Repositories: Implementing the CARE Principles in *Data Science Journal*, 23: 37, pp. 1–29 (July 2024)

Sesson 7 – Store and Manage

Description

While we're at the end of the cycle, the storage and management of research data are constant process throughout the research project. Storage should be considered for both data/documentation in the active stage of the project and once the project is completed. The storage solution(s) may be different at each point. Same with how the data/documentation is managed. In this stage, we must think long term and regularly consider if the availability of the data is still of benefit years later and if adjustments must be made.

Questions

- Who will manage the data at the conclusion of the project? Will participants or those immediately affected by the data have a say on how the data is managed after a certain point?
- Looking back to consent during the Create/Collect stage, will participants be made aware of the project's retention policy?
- If not at the beginning of the research data management planning process, have the appropriate stakeholders been contacted to discuss and determine the procedures, rights and responsibilities at this stage of the lifecycle?

Resources

- Chapter 5 Data Ethics
- Murillo, Luis Felipe R., How to avoid the "infrastructural blues"? Studying-while-caring for data stewardship in *Annals of Anthropological Practice* 48: 36–51 (2024)

- Guideline on retention of study information and data (February 2023)
    - Check with your institution/organization/funder for specific guidelines and policies for retention and disposal of research data

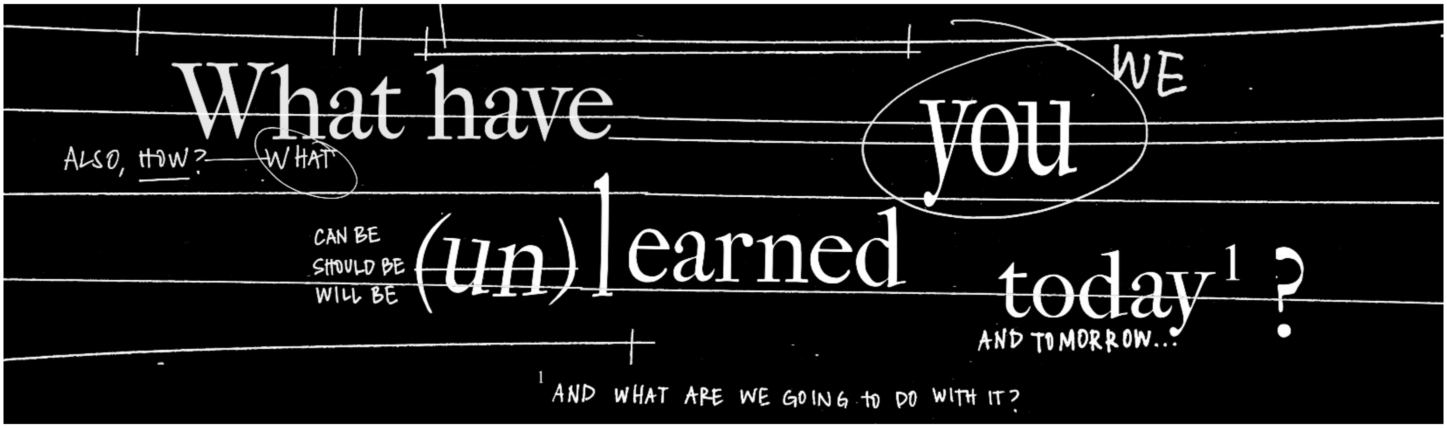

Figure 3 Work by Kameelah Janan Rasheed called "when j reads my work, he only ever reads aloud. never silently. often, his mouth will utter the words that are not on the page: revision by way of misreading. this is how we arrived at this piece" (2020)

---

[i] Carroll, S.R., Herczog, E., Hudson, M. *et al.* Operationalizing the CARE and FAIR Principles for Indigenous data futures. *Sci Data* **8**, 108 (2021). https://doi.org/10.1038/s41597-021-00892-0

[ii] "If we accept that the very object of data science—data—are made through the web of interested relationships among researchers, communities, instruments and institutions, we can see how the practice of producing insights and tools based on these relational objects ripples through and reshapes these relationships" (Boenig-Liptsin, Tanweer, and Edmundson, 2022)

[iii] Drucker, J., "Humanities Approaches to Graphical Display," *Digital Humanities Quarterly* 5, no. 1 (March 10, 2011): paragraph 3, http://www.digitalhumanities.org/dhq/vol/5/1/000091/000091.html

[iv] Wilkinson, M. D. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* 3:160018 doi: 10.1038/sdata.2016.18 (2016).

[v] King, S.. Build Your Archive. https://www.buildyourarchive.com/

[vi] Ravindran, S. (2024). Open with care. *Science*, *386*(6720), 372–375. https://doi.org/10.1126/science.adu0429

[vii] Committee on Population; Division of Behavioral and Social Sciences and Education; The National Academies of Sciences, Engineering, and Medicine. Sharing Research Data to Improve Public Health in Africa: A Workshop Summary. Washington (DC): National Academies Press (US); 2015 Sep 18. 4, Exploring the Ethical Imperative for Data Sharing. Available from: https://www.ncbi.nlm.nih.gov/books/NBK321546/