2020-12-09

Reflections on Research: Toward an Open Data Toolkit Centered on Diversity, Equity, Inclusion, and Accessibility Principles

Woodbrook, Rachel

https://dx.doi.org/10.7302/10
https://hdl.handle.net/2027.42/166087

Reflections on Research: Toward an Open Data Toolkit Centered on Diversity, Equity, Inclusion, and Accessibility Principles

This project was made possible in part by a 2019 award from the Catalyst Fund at LYRASIS.

Table of Contents

Project Goals

This project grew out of an existing relationship between the University of Michigan Library and the National Center for Institutional Diversity (NCID), and a shared realization in late 2018 that there were not enough data management resources available that fully incorporated diversity, equity, inclusion, and accessibility[1] (or DEIA) considerations into their advice and best practices. (NCID is an organization that funds, produces, and supports diversity research and scholarship across the nation and the world, and seeks to build intergenerational communities of scholars and leaders.[2] It defines diversity scholarship as scholarly work that "advanc[es] understandings of historical and contemporary social issues related to identity, difference, culture, representation, power, oppression, and inequality—as they occur and affect individuals, groups, communities, and institutions."[3])

Through work on NCID's own data, it became apparent that many of the questions the larger organization was encountering were also areas where the scholars they work with would benefit from more support, and the idea of creating a toolkit emerged. The Library partnered with NCID to create this resource; in order to identify scholars' needs and existing resources, we broke our process into several smaller goals:

1. Identify gaps in support for diversity scholars regarding their data, and identify current data practices that could be more widely disseminated.

Although DEIA considerations apply to data management broadly, for scoping purposes the intended audience for this project was NCID's Diversity Scholars Network (DSN), roughly 850 scholars from various disciplines who self-identify as working in diversity scholarship. When we began to explore data management and sharing through the lens of DEIA in 2018, we searched for existing work on these principles using the organizing framework of the research data lifecycle (the processes data moves through as part of a research project--see Appendix A for the list and description of stages we used). We came up with surprisingly few existing resources for several of the stages. We specifically wanted to work with the DSN to better understand their needs and what they would consider most useful in meeting those needs. We also suspected that some diversity scholars might already be using data practices or tools that would be useful to include in the toolkit and share more widely.

[1] See Glossary for our reference definitions of these concepts, which are in line with U-M Library definitions.
[2] "About | U-M LSA National Center for Institutional Diversity." Accessed August 3, 2020. https://lsa.umich.edu/ncid/about.html.
[3] "What Is Diversity Scholarship? | U-M LSA National Center for Institutional Diversity." Accessed August 2, 2020. https://lsa.umich.edu/ncid/about/what-is-diversity-research-.html.

2. Gauge scholar response to proposed toolkit resources.

We did have some initial ideas of potential gaps in support based on our literature review and previous experience. Data management can be an abstract and contextual topic even among those who work with it--different disciplines and communities may treat and understand data differently, and have different priorities. We took a very broad approach to what we considered data (see Glossary), and wanted to provide a concrete list of potential resources for scholars to respond to and prioritize. This would help determine if there was consensus around areas of greatest need, and also identify potential easy wins. We used our environment scan and literature review as well as our own experience to come up with an initial list of resources intended to improve satisfaction and confidence in making data decisions; see Appendix B for the list presented in the survey. (In both our qualitative and quantitative research, we presented these resources after first asking about perceived gaps and areas where more support was needed).

3. Draft a toolkit to support diversity scholars in making decisions about their data based on explicit DEIA considerations.

Diversity scholars work in many different disciplines; our hope was to create a resource useful across fields but also one that could be further developed and potentially made relevant to other types of cultural heritage institutions engaged in research, including museums and public libraries. While there is work already being done in these sectors as well as academia around DEIA and data (particularly data sharing), we did not find an existing resource pulling together tools from different disciplines for use across the research data lifecycle, and we anticipated that our research would indicate the usefulness of such a resource. This goal was the final step in our process, and is still in progress.

An additional, internal goal of our project was to model the best practices we discovered as we progressed, including effective partnering and transparency around labor, processes, and data.

Process

This study used a mixed-methods exploratory approach, consisting of a literature review, interviews, and a survey. The primary research question driving our work was:

> *"What implications do Diversity, Equity, Inclusion, and Accessibility (DEIA) principles and values have for each stage of the research data lifecycle?"*

Literature review

To address this question, we conducted a literature review and environmental scan of existing work. Our topic encompassed a wide array of fields and the terms used were correspondingly diverse. Thus, rather than conducting a comprehensive review with standardized search strings, we identified key areas of scholarship discussing topics related to our area of inquiry. For each area, we sourced publications and borrowed terms and references to expand our search within that field.

We found bodies of relevant research mainly in publications on open access/open data, research data management, critical data studies, data justice, indigenous data governance, community-accountable research, public scholarship, and public health. Because emerging concepts take time to enter the published literature, we also conducted an environmental scan of other scholars, academic and community projects, and organizations doing relevant work. Again, although we saw indications of work in the areas we were concerned with, nothing emerged that specifically addressed the need we saw for more centralized access to resources, especially when considering the entire data lifecycle.

Based on this review, we identified several themes that shaped the second stage of our research:

- The importance of connectivity, context, and relationships, and how much overall communication and research practices shape data practices.
- A diversity of language and metaphors emphasizing process over product (e.g., “data friction,” “data journeys,”[4] “data assemblages,”[5] and “data ecosystems,”[6]).
- The importance of research participants’ agency and choice around their own visibility (oftentimes, vulnerable communities are both over-researched and under-represented in available data).
- Issues of trust, ownership, and control, in the context of data stewardship as well as community vs. institutional capacity for data management.

[4] Bates, J., Lin, Y.-W., & Goodale, P. (2016). Data journeys: Capturing the socio-material constitution of data objects and flows. Big Data & Society. https://doi.org/10.1177/2053951716654502
[5] Kitchin R. The data revolution: Big data, open data, data infrastructures and their consequences. London: Sage, 2014.
[6] Yoon, A., Copeland, A., & McNally, P. (2018). Empowering communities with data: Role of data intermediaries for communities' data utilization. Proceedings of Association for Information Science and Technology (ASIS&T) Annual Meeting. 583-592.

Qualitative research: Interviews

Our team collaboratively designed a 10-question semi-structured interview protocol to explore how diversity scholars understand their research in relation to data decisions and the data lifecycle, and to elicit current data practices or needs for support in deeply considering DEIA principles around data as they move through the research process. After submitting our study design for review we obtained an institutional IRB exemption, although we made conscientious efforts to be explicit with plans for data use, storage, and sharing with participants, and used consent forms when collecting data. We revised our protocol, consulting with the U-M library's Assessment Specialist, and piloted the interview process with several graduate students before conducting about 10 hour-long interviews with faculty in southeastern Michigan, as well as one community data organization.

Through these interviews, we identified areas for further investigation as well as potential limitations in discussing the topics we were investigating across disciplines. One example of the way these interviews informed our next phase of research--survey design--was the identification of terminology which did not have clear meaning (or had differing connotations) depending on the discipline or type of work a researcher conducted. This terminology, which we decided to minimize in our next stage of research, included the phrases "open access"/"open data," "data lifecycle," and even "DEIA." For example, multiple scholars, in conversation about their data, focused on their research topic and goals more generally, rather than on their data specifically. The idea of data having importance and value separate from the publications that come out of its analysis is still emerging in many disciplines, and so it may not be surprising that participants did not respond or relate as expected to some of the terminology we used. This is a testament to the complex and nuanced contexts in which researchers work, and the intersections of the topics we are trying to address.

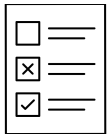
Emergent themes from this stage of research included:

- The impact of researchers' own identity(ies) on their practices and concerns.
- The perception that data sharing introduces vulnerabilities for both the researcher and participants.
- A reiteration of the importance of building trust and relationships and acknowledging power dynamics (on research teams and between researchers as well as between researchers and subjects).
- The tension between scholars' strong desires to act with full consideration for the impact of their choices, and internal and external pressures around time and resources.

- The challenge of prioritizing venues for research dissemination to increase impact. (In particular, preparing and sharing data was sometimes seen as requiring a large investment of time and effort for less direct benefit to research participant communities).

- Interest in the idea of a toolkit and the potential resources we proposed, as well as reservations about finding time to use the tools or even navigate through the resource.

Quantitative research: Survey



The last stage of our research consisted of designing and administering a survey. We used Qualtrics to create and distribute the survey to the roughly 850 members of the Diversity Scholars Network through the NCID's mailing list (the network is international, but most members are located in the United States). Out of 209 people who started the survey, 168 finished (for a completion rate of about 80%), and 140 of the completed responses were usable, for a response rate of roughly 20%. The response rate was likely affected by both the level of involvement required by the survey (at around 50 questions and a 15-20 minute average completion time) as well as its distribution during the COVID-19 pandemic shutdown in April 2020. However, we also took the number of scholars willing to invest their time in completing the survey as evidence of the importance of the topics we sought to address.

The survey was collaboratively designed, reviewed, and revised by our team in conjunction with the U-M library's Assessment Specialist and Accessibility Specialist. This utilized institutional expertise in DEIA considerations/diversity scholarship, survey design, statistics, and research data management (see Acknowledgements for more information on team members' specializations). The aim of the survey was to determine correlating factors for diversity scholars' perceived likelihood of utilizing a DEIA-specific data toolkit, as well as to better understand in what areas diversity scholars feel more need for support in managing their data according to DEIA concerns, and what types of toolkit resources they might find most useful.

Survey results did not show a correlation between researcher demographics and likelihood of using a toolkit, with only 8 out of 140 respondents indicating they would be unlikely to use one. The most anticipated barrier to using a toolkit was lack of resources (time, funding, staffing), with 83 respondents selecting this option. Overall, scholars indicated that they do feel a need for more support to be comfortable with all stages of the data lifecycle. When asked what stage was the best example of a process both important to their project and comfortable for them, scholars most often indicated data collection, processing/analysis, and finding existing data. Interestingly, the top two responses when asked what stage was important but most *uncomfortable*--data archiving/preservation and data sharing--are areas libraries and other centralized campus resources are often intended to support.

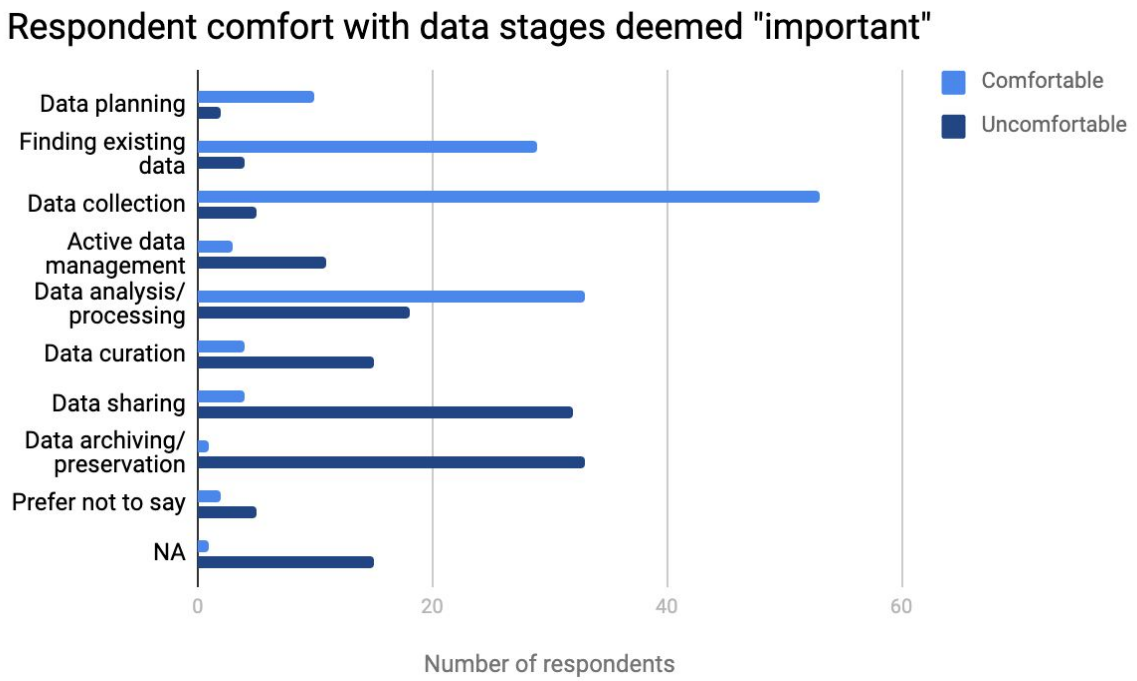


Figure 1: Data lifecycle stages marked as important/comfortable and important/uncomfortable (See Appendix C for table underlying chart)

Most toolkit items were rated as useful by a fair number of respondents (who could choose multiple items). The items most often selected were “successful examples of engaging communities in research design and data governance”; ”a checklist of questions for making data decisions”; and “templates for one-page data applications or data use agreements.” (See Table 1 below, and Appendix B for full descriptions).

Potential resource	**No. Respondents**
Examples of community engagement	100
Checklist of questions for making data decisions	97
Templates for one-page data applications or use agreements	95
Consent form language for data sharing	90
Resources on hidden metadata	84
Resources on data accessibility	75
Case studies of data shared back to participants	75
Resources on participant rights	70
Bibliography/Reading material	69
Resources on de-identification	66

Table 1: Perceived usefulness of potential toolkit resources

The research so far thus indicates general consensus from respondents that they would like more support in areas libraries are equipped to address; and that a toolkit containing at least some of the items we proposed would be useful.

Modifications to original proposal: Timeline

The main change to the initial proposal is likely a common one: scoping and timeline had to be adjusted as the project progressed. Although we were diligent in preparing a plan, we found that over the course of conducting the project there were multiple instances where we needed to slow down our decision-making process in order to fully embody the ethical best practices we were researching. We also hired more student team members at fewer hours each than anticipated, and it was the PI's first time supervising a research project, so many processes could not be taken for granted and required teamwide discussion to establish. These factors, combined with the usual fluctuations in bandwidth and team composition, as well as the completely unexpected COVID-19 outbreak that began affecting life in the U.S on a large scale as our survey got underway, meant that our timeline stretched in order to accommodate the quality of work we were committed to.

The size of our team and the nature of our research meant that we could adjust fairly easily, but one area we would spend more time on up front if conducting a similar project in the future is anticipating presentation venues ahead of time, and planning what types of presentations to propose before results were complete. Although these challenges have meant that we will not be delivering the final product as soon we anticipated, this was a unique opportunity to build a team and research process integrating DEIA data considerations from the ground up, and a growth opportunity for both staff and students to gain research experience while creating something new and beneficial to the field.

For other proposals considering following this model (conducting multi-level research, building a product based on analysis of research results, and disseminating the results), we would recommend doing so only if as many of the variables for the project, team, and target venues for distribution as possible are established ahead of time, and team leaders have experience gauging the length of time needed for such a project. We also recommend budgeting extra person-time to accommodate for inevitable unanticipated delays.

Accomplishments

Although we still have work to do in completing the toolkit (Goal 3), we feel this project has been successful. We identified areas of the data lifecycle where diversity scholars feel they need more support (Goal 1), and we will be able to prioritize including and potentially developing toolkit items based on respondent feedback to our proposed resources (Goal 2).

Process

We built a team with members who had quantitative as well as qualitative expertise and backgrounds ranging from information science and DEIA to health and scientific computing. This helped extend our team conversations around research approaches and presentations beyond a single disciplinary approach. We also worked together successfully to distribute the workload and move forward on multiple tasks at once. In addition, we were able to balance best practices around documentation, consent, and other responsible research practices (such as verifying the accessibility of our survey) with maintaining forward momentum.

Another accomplishment of this project was the successful continuation and strengthening of the Library's relationship with NCID. As a project partner, we will continue to consult with them on the platform for the toolkit and publications, further opportunities or directions to pursue in our efforts to make the resulting toolkit most useful to their constituents, and planning write ups and presentations of the overall project results.

Results

In line with our project goals, we gathered a rich dataset based on unique access to the Diversity Scholars Network. The scholars we interviewed and surveyed were generous with their time and trust, and we are confident in the quality of the data we generated based on our ability to iteratively review the research instruments we created. The increasing visibility of DEIA as applied to research data over the past year in publications, conferences, and presentations on research data clearly evidences it as an important area of study, and we are hopeful our work will contribute to these larger conversations. Initial responses from interviews and survey comments show that the types of resources we are considering for the toolkit are indeed valuable to scholars in their work.

Products

Because of the ways in which our timeline shifted, some of our anticipated products are still in process. At the 2019 LYRASIS summit, we talked about the outcomes of our project including "not only an easily-accessible version of the finalized toolkit as well as presentations on the material, but also contextual information from the process, including our own documentation and data where releasing such information is possible and appropriate." We are currently in the process of planning out multiple write ups based on our work, focused not just on our results but also maintaining transparency around our process and making our data and methodological details available.

In the meantime, over the course of the project we have shared out information on our progress in various venues, and established an online presence to facilitate engagement and dissemination. The main avenue to find out information about our project while it is still

ongoing is our project website: https://um-deia-data-toolkit.github.io/home. Also linked on the site are outputs of the project so far, including:

- May 2019: Poster presentation at RDAP Summit
- June 2019: Poster, lightning talk presented at U-M Library ShareFest
- July 2019: Lightning talk at Data Curation Network All-Hands Meeting
- September 2019: Project write up on Michigan Publishing Website
- October 2019: Presentation at LYRASIS Leaders' Summit
- March 2020: Lightning talk at Research Data Access and Preservation (RDAP) Summit

More project outputs, including links to data and instruments and additional write ups on our results and methodology, will be added to the site as they become available. Data and instruments will eventually be deposited in Deep Blue Data or another repository for public access.

Lessons Learned

Based on our experience, we reshaped our approach to several aspects of the project. At first, we framed our work to encourage open data sharing for diversity scholars, but through conversations with our participants--and especially with the community data organization we interviewed--we found ourselves shifting our focus to finding the right audience for data, and working toward making the consideration of opening data up an explicit question while being careful not to assume that openness is appropriate in every case. Put more succinctly, we pivoted toward a framing of making data "as open as possible, as closed as necessary."[7] This allowed us to more effectively address the interdisciplinary nature of diversity scholarship and broaden the impact of our work.

Over the course of the project, we also learned first-hand how complex a balancing act it is to put the best practices we were discovering into use. The flexibility of LYRASIS and support of our liaison in adjusting the project timeline as needed was crucial in this regard. Together, these factors illustrate one of the principles that came up in our research: in order to produce research products that are truly consistent with ethical research practices and thoroughly informed data decisions, individual researchers' processes must be supported at the institutional, funder, and publisher level as well. This is often at odds with the traditional research ecosystem, which values urgency and productivity over developing and nurturing relationships with research partners, or shoring up existing resources.

This was also a valuable exercise in reflecting on our own process before making recommendations to other researchers. As the project progressed we learned about a number

[7] "Open Access - H2020 Online Manual." European Commission, 2020. https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/open-access_en.htm.

of existing resources (and more are being created even as we write). We also learned that researchers' concerns about time constraints and navigating and applying new tools are of great concern. Thus, we have shifted our toolkit focus to providing effective curation and navigation.

Finally, this project has given us an opportunity to reflect on the challenges of engaging the entire ecosystem of those affected by research data. More specifically, we would have liked to partner with existing projects working with research participants and community data organizations, in addition to researchers and those who collect data. This would have given us access to a greater diversity of perspectives in building our toolkit; however, it would also have expanded our scope even farther, so at this point we consider it a known limitation of the project and an area for further exploration. We do hope to encourage more community engagement as we draft and hopefully develop the toolkit draft for a wider audience.

Next Steps

Analysis and write ups

We are approaching the end of the research phase, and are conducting our analyses of the quantitative research data, and cleaning and quality assurance for our qualitative data. As we complete these processes, we are also planning and beginning to write up our results and identify possible publication venues for writing on the project's literature search, data and methodology practices, qualitative and quantitative findings, and overall reflection on the process and partnership with NCID, including the application and assessment of the toolkit once it is drafted.

Toolkit creation

Simultaneously, we are moving into working on the toolkit itself. The first steps on this front are defining requirements, identifying candidate platforms, strategizing navigability, and comparing our list of existing resources against the priorities expressed by our research participants.

Dissemination

Finally, we are working to identify channels beyond publication that hold promise for distributing our toolkit to scholars, librarians, and others for whom it may be useful. This will likely include working with partners internal to U-M as well as externally. In addition to our upcoming LYRASIS webinar we have discussed other webinars, workshops, and/or online education modules as potential methods of dissemination for this work.

Rachel Woodbrook, University of Michigan, woodbr@umich.edu

Acknowledgements

In addition to the generosity of the LYRASIS Catalyst Fund, this project would not have been possible without the dedicated work of our project team and other collaborators such as Craig Smith, U-M Library Assessment Specialist, and Stephanie Rosen, U-M Library Accessibility Specialist. Our team members represent a diversity of social identities, disciplinary backgrounds, academic training and roles within academia. It is also important to acknowledge limitations of the project based on aspects of identity not represented on our team; two perspectives we know we do not directly represent are research participants, and assistive technology users. Future work on these topics should look for ways to incorporate perspectives up and down the research chain, as well as additional identities. For more information about individuals' contributions to the project, please see our website.

Staff team members

Rachel Woodbrook	*Data Curation Librarian, U-M Library; Principal Investigator*
Karen Downing	*Education Librarian, U-M Library; Co-Investigator. Ph.D. Education. DEI researcher, qualitative research methods specialty.*
Megan Segoshi	*Program Lead for Scholar and Community Engagement, National Center for Institutional Diversity (NCID); Advisor*
Jake Carlson	*Director, Deep Blue Repositories and Research Data Services, U-M Library; Advisor*
Laura Sanchez-Parkinson	*(former) Assistant Director of Programs & Development / Program Manager for Research, National Center for Institutional Diversity (NCID); (former) Advisor*

Student team members (in order of hire)

Emma De Vera	*MSI, U-M School of Information; Student Assistant*
Elyse Thulin	*M.Sc, PhD candidate in Health Behavior/Education and Scientific Computing, U-M; Graduate Research Fellow*
Emily Oxford	*MSI Candidate, U-M School of Information; Student Assistant*
Chanese Forté	*PhD Candidate in Environmental Health Science and Scientific Computing, U-M; (former) Graduate Research Fellow*
Tierra Shannon	*BA, English, U-M; (former) Student Assistant*

Glossary

Accessibility: Accessibility means supporting meaningful access to resources by all people, with their diverse range of needs, abilities, bodies, minds, and backgrounds. Promoting accessibility require identifying and removing existing barriers, incorporating best practices of accessible design, and learning from the expertise of individuals who regularly face barriers to access. (Adapted from Stephanie Rosen)

Data lifecycle: The processes data moves through as part of a research project--see Appendix A for the list and description of stages used in our survey.

Diversity: Diversity is expressed in myriad forms, including race and ethnicity, gender and gender identity, sexual orientation, socioeconomic status, language, culture, national origin, religious commitments, age, (dis)ability status and political perspective. (Adapted from U-M Library)

Diversity scholarship: research that "advanc[es] understandings of historical and contemporary social issues related to identity, difference, culture, representation, power, oppression, and inequality—as they occur and affect individuals, groups, communities, and institutions." (National Center for Institutional Diversity)

Equity: Equity requires working actively to challenge and respond to bias, harassment, and discrimination. It requires commitment to equal opportunity for all persons and does not discriminate on the basis of race, color, national origin, age, marital status, sex, sexual orientation, gender identity, gender expression, disability, religion, height, weight, or veteran status. (Adapted from U-M Library)

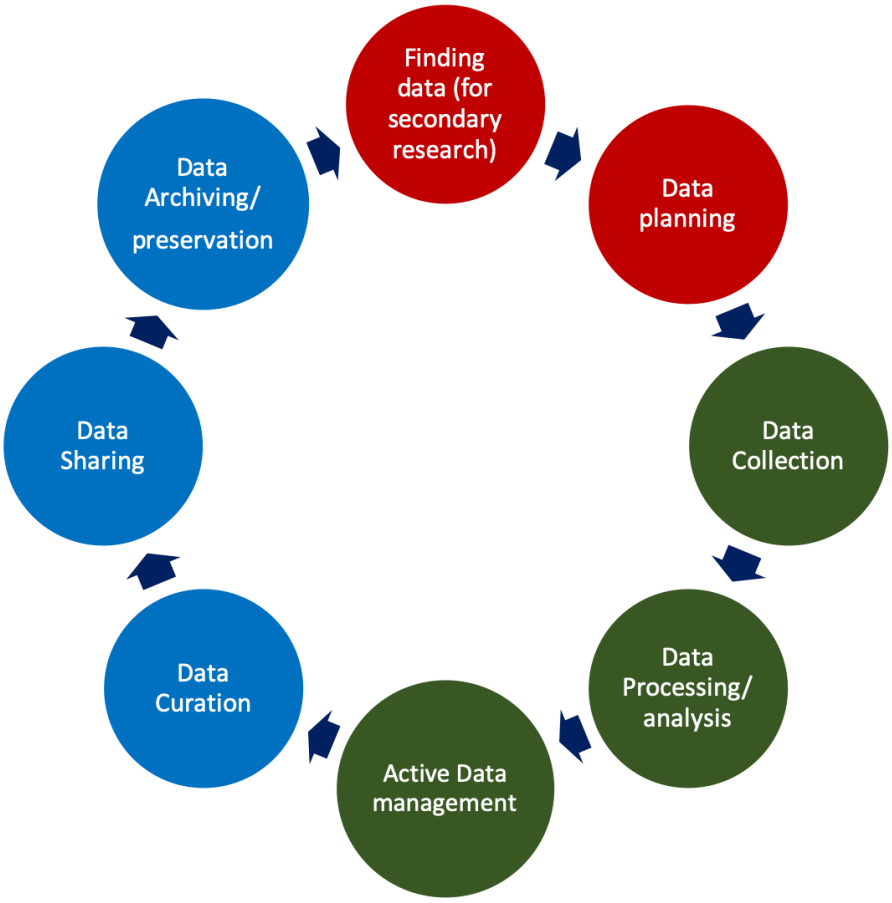
Inclusion: Inclusion is demonstrated when differences are welcomed, different perspectives are respectfully heard, and every individual feels a sense of belonging. By building a critical mass of diverse viewpoints and creating a climate of inclusiveness, we can more effectively advance our collective capabilities. (Adapted from U-M Library)

Research data: Any data produced during the process of research in any discipline. This includes materials collected, observed or generated for analysis, that serve as a basis for original research or scholarship in any discipline.

Icon references

- articles by I Putu Kharismayadi from the Noun Project
- interview by DailyPM from the Noun Project
- Question by Graphic Tigers from the Noun Project
- Seo Report by I Putu Kharismayadi from the Noun Project
- application by Richa from the Noun Project
- tools by arif fauzi hakim from the Noun Project
- multicultural by john melven from the Noun Project

Appendix A: Data lifecycle (from survey)



Text description of image: The data lifecycle is depicted as eight stages in a circle leading into each other. The first two stages, which take place before starting a research project, are "Finding data (for secondary research)" and "Data planning." The next three stages take place during the project: "Data collection," "Data processing/analysis," and "Active data management." The final three stages are undertaken after the project is completed: "Data curation," "Data sharing," and "Data archiving/preservation." This final stage may lead back into finding data.

1. **Finding existing data:** searching, locating, and accessing data for secondary analysis or reuse.

2. **Data planning**: writing data management plans (DMPs) or setting up processes for any stage of working with data. For example,
    - Data security protocol for IRB proposals
    - Study design/population choices
    - Informed consent design

    - Designating roles and responsibilities

3. **Data collection**: gathering or measuring information on topics or variables of interest. For example,
    - Recruiting, consenting and/or interviewing participants
    - Web scraping
    - Archival research
    - Collecting and recording metadata

4. **Data processing and analysis**: the process(es) by which understanding or conclusions are drawn from data. What questions are being asked of the data? How are data cleaned and quality checked?

5. **Active data management**: steps taken while data is collected and analyzed to facilitate access and keep track of the data. For example,
    - Adding metadata and documentation to preserve context
    - Storing data in an appropriate location using a file and folder structure
    - Versioning and workflow
    - Implementing security protocols

6. **Data curation**: preparing data for consumption or sharing outside the research team. For example,
    - Deciding what documentation is needed to understand the data in context for a given audience
    - Designating what will be shared (consent documents, raw or processed data, protocols, codebooks, guidance on appropriate uses of the data)
    - Formatting, cleaning, aggregating, and anonymizing or de-identifying data

7. **Data sharing**: deciding and implementing protocols about who should have access to underlying data, who makes these decisions, and how access is implemented. This could include sharing data back to participants, managing data access applications, or licensing data.

8. **Data archiving and preservation**: Making and implementing decisions about where to store data long-term, who will be responsible, how to maintain access and for how long, and meeting funder or journal requirements for data accessibility.

Appendix B: Potential toolkit resources (from survey)

1. **Consent form language/templates** for sharing (raw data, de-identified data, sharing with other researchers, publicly, etc.) See for example the Qualitative Data Repository's templates).

2. Successful examples of **engaging communities** in research design and data governance. See for example "Good Data Practices for Indigenous Data Sovereignty" in *Good Data*.

3. Resources for researchers and research participants on **participant rights** around data. See for example the Chicago Beyond guidebook, "Why Am I Always Being Researched?"

4. A bibliography or other **reading material on ethical considerations** for data decisions. See for example the Responsible Data Handbook.

5. Resources on potential **hidden identifying information** or embedded metadata. (For example, underlying geotagging in the "Healthy Minds" study was used to identify survey respondent locations and shut down a dorm that typically housed minority/art students.)

6. Resources on **de-identification** of human subjects data.

7. Resources on formatting or hosting data for **accessibility** to those using assistive technologies.

8. Templates for one-page **data applications** or **data use agreements** when sharing data.

9. **Case studies** of underlying data shared back to participants in useful ways.

10. A **checklist of questions** for making data decisions, including how far to open data, and to whom. See for example the Data Ethics Canvas.

Appendix C: Table for Figure 1 - Data lifecycle stages marked as important/comfortable and important/uncomfortable

Stage	No. Comfortable	No. Uncomfortable
Data planning	10	2
Finding existing data	29	4
Data collection	53	5
Active data management	3	11
Data analysis/ processing	33	18
Data curation	4	15
Data sharing	4	32
Data archiving/ preservation	1	33
Prefer not to say	2	5
NA	1	15