M | LIBRARY

Building a Data Lifecycle Management Toolkit to Support Diversity Scholarship

Rachel Woodbrook (woodbr@umich.edu)
March 16, 2022
RDAP Summit

Hi, my name is….

Before I start, I want to mention that I'm on land belonging to Coast Salish peoples, specifically the Duwamish, the host tribe of Seattle and King County. If anyone else lives in the area or feels connected to it and hasn't yet heard about the real rent campaign, it's a material way to acknowledge and support the tribe, who have been custodians of this land for thousands of years. You can also find out more about supporting their petition for state and federal recognition as a tribe, at https://www.duwamishtribe.org/

https://tinyurl.com/WoodbrookRDAP2022

Rachel
If you'd like to follow along, the slides are available at [read URL]. I will try to add this to the chat as well.
[Add URL to chat during live session] -
https://tinyurl.com/WoodbrookRDAP2022

Acknowledgements

LYRASIS®

ALA American Library Association

[Our Team](#)

Agenda

I also wanted to give a wholehearted thanks to the sponsors of our project, without whom most of this would not have been possible. Lyrasis's Catalyst Fund is an amazing opportunity that I encourage anyone with a wish list project idea to consider applying for; and the ALA Office for Diversity, Literacy and Outreach Services awarded us a Diversity Research Grant that allowed us to keep on one of our amazing student assistants, Emma De Vera, to help with the toolkit creation and revision. (You can see our entire team, past and present, on our website).

Project Overview

on quantitative research results

Very briefly, then–this project was based at the University of Michigan Library, and conducted in partnership with the National Center for Institutional Diversity, which is an organization supporting scholars doing diversity scholarship. Diversity scholarship is scholarship that is examines and seeks to affect social issues such as identity, representation, oppression, and inequality — at the individual, group, community, and institutional level.
(https://lsa.umich.edu/ncid/engagement-opportunities/diversity-scholars-network.html)

The first part of the project consisted of conducting qualitative and quantitative research with scholars who are part of a diversity scholars network created by NCID. I've linked slides sharing some of these results. The second part, which I'll focus on today, was about building the resource.

The basis of this project (as several other scholars have mentioned in lightning talks and presentations already) was a recognition that there is a disconnect between conversations about open data and data sharing requirements, and support for scholars to feel confident in sharing their

data, knowing that they have been able to take all the necessary steps to make sure they are doing so appropriately. At the time the project started (fall 2018), there were almost no resources explicitly tying data ethics to the data lifecycle for management purposes, which was a framework we thought would be useful.

The overarching question shaping our project was, "What implications do Diversity, Equity, Inclusion and Accessibility considerations have for best practices in each step of the data lifecycle?"

Research Findings Overview

- 131/140 respondents "somewhat likely" or "very likely" to use a toolkit if available.

- Biggest **anticipated barrier**: time and resources (83/140)

- What toolkit resources would diversity researchers find **most useful**?

  - 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)

- Important data lifecycle stages: most **uncomfortable**

  - Data sharing (32)
  - Data archiving/preservation (33)

One of our biggest findings (no surprise) was that respondents all wanted more support, *and* worried that even if a resource were available, they would not have the time or resources to invest the effort to use it. This research helped further clarify the need for **institutional, structural, and cultural** changes to support best practices around data ethics. Many researchers already feel overwhelmed, and timelines and funding models are not supportive of truly ethical data practices, which require making space for **slowness, maintenance, and investment in relationships.** These are not problems which can be solved on an individual level, though there are things we can do.

Although we didn't see a correlation between demographics and desire for support, we posited that such a resource could be especially useful for scholars earlier in their careers, with fewer resources at their disposal (or not associated with an institution), and/or with minoritized identities that might leave them open to greater scrutiny or consequences.

Our goal, then, was to identify the most useful resources, and find a way to collate and make these available that was not overwhelming or difficult to wade through, so that scholars could quickly and easily

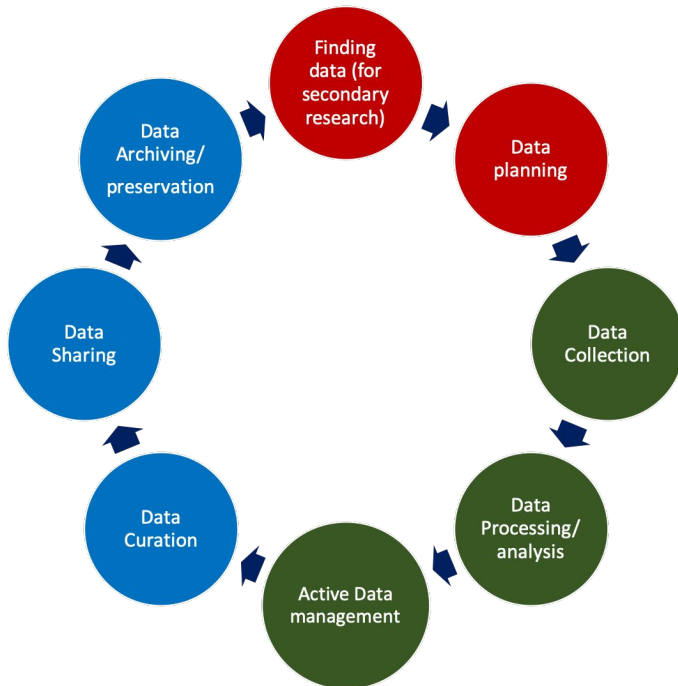identify the resource(s) that might be helpful for them.

Toolkit creation: Parameters

The toolkit must:

- Have a **highly curated** set of useful resources
- Include **easily understandable and widely applicable** resources
- Be **easy to navigate**
- Not overwhelm the researcher seeking guidance

We came up with parameters for the toolkit based on on interviews with researchers, and did an environmental scan of available resources online while we were doing our research. We ended up with about 60 resources to review from about Fall 2018-Spring 2020 (we've continued to add to this, and I would like to do a more comprehensive update).

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)" then "Data planning." The next three stages take place during the project: "Data collection," "Data processing/analysis," then "Active data management." The final three stages are undertaken after the project is completed: "Data curation," "Data sharing," then "Data archiving/preservation." This final stage may lead back into finding data.

—---------------

We used the research data lifecycle as a structure for our project, and coded resources to various stages of the lifecycle–likely familiar to most of you.
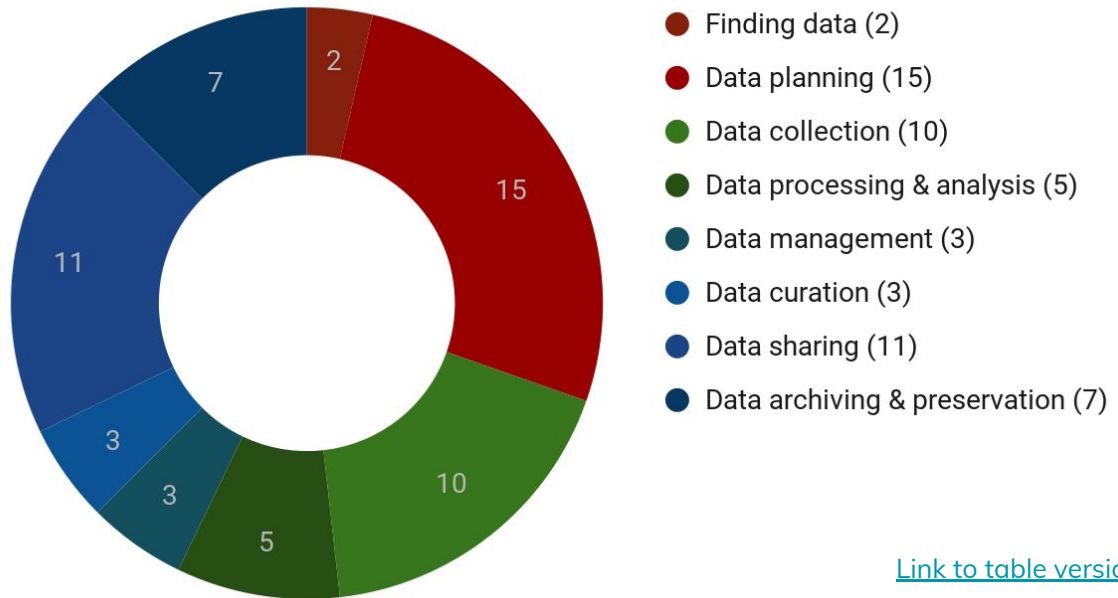
Toolkit content: Inclusion criteria

- Applicability

- Accessibility*

- Usefulness

- Scope/breadth

- Timeliness/Currency

- Addresses Data Lifecycle Stages

- Authorship

To evaluate the resources we found, we looked at these factors:

- **Applicability:** was the resource created to address particular needs of researchers working with diversity scholarship, or created to address ethical and/or social justice issues raised when dealing with research data more generally?
- **Accessibility**: is the resource readily accessible to most researchers? Does it sit behind a paywall? Are there accessibility concerns about its format or platform?* * Our process for assessing this needs to be refined
- **Usefulness**: is the resource directly applicable in its current form? Does it require specialty skills or knowledge to apply? Is it succinct and concrete? Does it have enough guidance and substance to be widely useful to researchers?
- **Scope/breadth**: is the resource applicable to multiple disciplines or research methods?
- **Timeliness/Currency:** Has the resource been recently updated

- or created? If it is older, is it still relevant to today's needs?
- **Data Lifecycle Stages:** Does the resource provide advice that specifically applies to one or more of the data lifecycle stages for research data?
- **Authorship:** Is the resource produced by an author or collective that is experienced and/or knowledgeable about DEIA issues and/or diversity research?

What's in the toolkit?

- Finding data (2)
- Data planning (15)
- Data collection (10)
- Data processing & analysis (5)
- Data management (3)
- Data curation (3)
- Data sharing (11)
- Data archiving & preservation (7)

Link to table version

In the end, the first version of the toolkit has 17 resources total. (Our initial goal was 3-5 for each stage, or 25-35 total). Both data sharing and data archiving and preservation, the areas researchers indicated least comfort with, are fairly well-represented. However, you can see that there are only 2 resources on finding data; these were the only 2 we were able to find so far that were specifically oriented towards those wanting to work with data that can be reused. Similarly, we found very little on accessibility of data as it relates to assistive technology, etc.

For researchers (or others) who wish to dig deeper into the range of resources available, we do provide access to the full Google spreadsheet with all the resources identified, and have a mechanism for suggesting additional resources as well (https://docs.google.com/forms/d/e/1FAIpQLSdqsNkz5mj4_OCKAQwptx A2lvd7mFH2Pd4Q__ssseTMoTnt-A/viewform?usp=sf_link).

Toolkit 1.0

Open Data DEIA Toolkit 1.0

There are many considerations touching on Diversity, Equity, Inclusion, and Accessibility that inform researchers' perspectives, practices, and decisions regarding their data.* Although some fields have extensive research ethics histories that include detailed data practices, many do not. Many researchers end up learning best practices either through trial and error, or from whatever personal and professional networks they happen to have.

Our goal with this project is to collate and make visible some of the resources relevant to different stages of work with data throughout a research project; that is, we sought to include resources with practical applications at various decision points in the research process. When this project was started (3 years ago now), there were many fewer resources available. As these conversations have become more frequent, additional resources are being created. This list is not exhaustive; if you know of a resource you think should be included, you can contact us at **um-deia-data-toolkit [at] umich.edu**. For more information about the data stage definitions or our inclusion and exclusion criteria, see the full spreadsheet linked below (which also includes additional information for each resource).

A good short introduction to the types of intersections between DEIA considerations and data that might arise during a research project can be found in the Principles for Advancing Equitable Data Practice report.

| Last updated ▼ | Format ▼ | Data stage(s) ▼ |

Linked Title ▲	Creator(s)	Description	Why we chose it
Anti-Racism Data Reuse Guide	ICPSR	The purpose of this document is to advance racially equitable data reuse practices for researchers to ensure that analysis does not harm the population(s) being studied or contribute to further marginalization of racialized communities.	This resource focuses on secondary data reuse including evaluating data and analysis methods, and is recently-created.
Data Ethics Canvas	The Open Data Institute	The Data Ethics Canvas is a tool for anyone	This tool can be used at

https://tinyurl.com/deia-data-toolkit-v1

This is what the initial version of the toolkit looks like currently [go to toolkit]. It's pretty basic; we have a few pieces of metadata visible, and more behind the scenes. There are filters for data stage, format, and date the resource was last updated. We ended up using Google Data Studio for the platform; although this had some advantages (especially since we didn't have a budget for web design or execution) and passed muster with our accessibility folks, it's not the most intuitive tool to navigate, and is more oriented toward the visual presentation and analysis of data, which is slightly different than what we are doing.

Linked Title ▲	Creator(s)	Description	Why we chose it
Data Ethics Canvas	The Open Data Institute	The Data Ethics Canvas is a tool for anyone who collects, shares or uses data. It helps identify and manage ethical issues – at the start of a project that uses data, and throughout. It encourages you to ask important questions about projects that use data, and reflect on their responses.	This tool can be used at multiple points throughout the research without feeling redundant, and provides a clear outline through which to approach various parts of the research process.
Equitable Open Data Report	Detroit Digital Justice Coalition and Detroit Community Technology Project	This report has on-the-ground advice for researchers working with data from at-risk/marginalized communities, based on interviews with Detroit residents considering the potential benefits and harms of various data sets currently available on the City's Open Data Portal. The report assesses what actions can be taken by the City to maximize benefits and minimize harms, and investigates open data policies and tools in place in other cities that model their vision for data justice.	Easy to follow advice and general issues to think about-- this resource is focused on open government data but includes direct perspectives from community members thinking about their data being used and shared.
Feminist Data Manifest-No	Marika Cifor, Patricia Garcia. In addition to their efforts, the first complete draft is the collective labor of TL Cowan; Jasmine Rault; Tonia Sutherland; Anita Say Chan; Jennifer Rode; Anna Lauren Hoffmann; Niloufar Salehi;	The Manifest-No is a declaration of refusal and commitment. It refuses harmful data regimes and commits to new data futures.	The authors address data production as contextual and frame it within "matrices of oppression"; this resource primarily focuses on big data, but can also be useful as a framework for thinking about potential data harms

Here are a few examples of the first resources listed in the "Data sharing" category.

For example, the first resource listed here, the Data Ethics Canvas…

open
data
institute

This document includes:
➔ The Data Ethics Canvas
➔ User guide
➔ Facilitator guide

Data Ethics Canvas

Title	Description	Comments #1	Comments #2	Actions and decisions	Responsible	Date
1. Data sources	Name/describe key your project's data sources, whether you're collecting data yourself or accessing via third parties. Is any personal data involved, or data that is otherwise sensitive?					
2. Rights around data sources	Where did you get the data from? Is it produced by an organisation or collected directly from individuals? Was the data collected for this project or for					

…walks users through a series of questions in table format to interrogate intention, potential uses, and the consent of research participants to consider when preparing data for sharing. This corresponds with the data sharing lifecycle and could be used by any researcher who intends or may want to share their data outside of their project.

Linked Title ▲	Creator(s)	Description	Why we chose it
Data Ethics Canvas	The Open Data Institute	The Data Ethics Canvas is a tool for anyone who collects, shares or uses data. It helps identify and manage ethical issues – at the start of a project that uses data, and throughout. It encourages you to ask important questions about projects that use data, and reflect on their responses.	This tool can be used at multiple points throughout the research without feeling redundant, and provides a clear outline through which to approach various parts of the research process.
Equitable Open Data Report	Detroit Digital Justice Coalition and Detroit Community Technology Project	This report has on-the-ground advice for researchers working with data from at-risk/marginalized communities, based on interviews with Detroit residents considering the potential benefits and harms of various data sets currently available on the City's Open Data Portal. The report assesses what actions can be taken by the City to maximize benefits and minimize harms, and investigates open data policies and tools in place in other cities that model their vision for data justice.	Easy to follow advice and general issues to think about--this resource is focused on open government data but includes direct perspectives from community members thinking about their data being used and shared.
Feminist Data Manifest-No	Marika Cifor, Patricia Garcia. In addition to their efforts, the first complete draft is the collective labor of TL Cowan; Jasmine Rault; Tonia Sutherland; Anita Say Chan; Jennifer Rode; Anna Lauren Hoffmann; Niloufar Salehi;	The Manifest-No is a declaration of refusal and commitment. It refuses harmful data regimes and commits to new data futures.	The authors address data production as contextual and frame it within "matrices of oppression"; this resource primarily focuses on big data, but can also be useful as a framework for thinking about potential data harms

…You can see from the Equitable Open Data Report, the second resource, that in some cases we had to go outside academia to find good resources. Many resources come from a particular frame of reference–big data, indigenous data, government open data–but include perspectives and questions that provide useful starting points in other situations as well.

Curating these resources to form a finished product was challenging, as it is ultimately a patchwork. Our multidisciplinary and introductory-level approach meant that many high-quality resources we did find that were very relevant to the issues we were concerned with (e.g., Traditional Knowledge labels for metadata) were too specific to include in a general toolkit. And some challenges researchers undoubtedly need support with, such as true data de-identification, are also just too complex to really address in this type of a resource.

However, I do feel that we were able to come up with a number of useful resources for those looking for an entry point, or somewhere to direct others when questions arise.

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)" then "Data planning." The next three stages take place during the project: "Data collection," "Data processing/analysis," then "Active data management." The final three stages are undertaken after the project is completed: "Data curation," "Data sharing," then "Data archiving/preservation." This final stage may lead back into finding data.

—---------

I was going to do some more process talk about deciding on a platform and building the toolkit, but suffice it to say–I am not a web designer nor a data visualization expert, and a lot more time and effort went into building what we currently have than what it probably looks like! But I do believe that we are at a point now where the toolkit is usable and feedback would be helpful, so if this is something that interests you I encourage you to take a look and feel free to use it if it can be incorporated into your work, or to send us any comments or suggestions.

I will continue to work on improving navigation and visual formatting. Based on feedback we've received so far, we may revisit using the data lifecycle visual as an entry into the toolkit; users could then select a

lifecycle stage, such as Data sharing…

[Data Ethics Canvas](#)

Description: The Data Ethics Canvas is a tool for anyone who collects, shares or uses data. It helps identify and manage ethical issues – at the start of a project that uses data, and throughout. It encourages you to ask important questions about projects that use data, and reflect on their responses.

Creator(s): [The Open Data Institute](#)

Format: Document - table with questions and space for answers

Last updated: 2019

Applicable to Discipline(s): Any

Applicable to Methodologies: Any

Why we chose it: This tool can be used at multiple points throughout the research without feeling redundant, and provides a clear outline through which to approach various parts of the research process.

Data stage(s): Data planning, Data collection, Data sharing

…and be directed to something more like a webpage, with all of the metadata we assigned to each resource for that lifecycle stage displayed.

Next steps...and beyond!

- Finish interviews and initial toolkit revision
- Survey Diversity Scholars Network
- Plan for sustainability
- Expansion? Crowdsourcing, broader directory?
- Application and use - by researchers and institutions

Once we complete collecting feedback and implementing initial revisions, we will be sending out a short feedback survey to NCID's Diversity Scholars Network to obtain additional impressions and suggestions from researchers.

In terms of planning for the future, the dissemination and incorporation of the toolkit for use is a new phase and will be partly determined by what we hear back, and/or how others want to use the toolkit. We will continue to talk with NCID about effective ways to introduce the toolkit to the Network, and I'm excited to keep working on this and find out what shape it might take next!

For example, In addition to the highly-curated list I think something like a broader directory could be manageable and useful. Of course, one of the biggest questions for whatever we do next is sustainability; at this point, our student researchers have cycled off the team so we need to figure out what is manageable, and potentially find new partners.

Thinking big, I would love to find a way to crowdsource experiences with particular tools, or to facilitate rating tools, and to make the tool more

interactive to further engage the research community in academia and beyond. We could also expand to incorporate more research participant perspective and go beyond the academic setting to data application as well. The toolkit is not meant to be a static, finished resource but instead a catalyst for conversations and interaction, and a tool where helpful for workshops, classes, etc.

I could even see either the toolkit or some of the resources within it becoming incorporated more institutionally into approaches to data ethics and research education modules beyond the library's influence. Of course, that would take a lot of additional work and buy in, but when I think about what we've been able to do so far and conversations happening now across our field, I think this shift is a matter of "when" rather than "if," and our work could be part of this progress.

Thank you!