DEIA Open Data Toolkit Interview Protocol for Graduate Students

Share the informed consent form and offer to either read it for the participant or have them read it on their own. Ask if the participant is comfortable with recording the interview, and mention that we are planning to share de-identified data through deposit in Deep Blue Data or another online data repository.

Press record on the recorder and [State interviewer, interviewee names and the date for recording]

Scripted Introduction:

Good *morning/afternoon*, and thank you for agreeing to share your experience and time with us. The goal of this interview is to learn how scholars whose research addresses issues of diversity, equity, access, or inclusion make decisions around their data. This includes acquiring, managing, sharing, and preserving data. We are particularly interested in how these decisions potentially impact the people or populations you study, or who are affected by your work.

This project has passed through U-M's IRB and was exempted. This means that your participation involves no more than minimal risk and that our data collection does not involve participants who are in vulnerable populations. You were invited to participate because you were identified as a graduate student within the diversity scholar network. We have adopted the National Center for Institutional Diversity's definition of a diversity scholar as someone whose scholarship "furthers our understanding 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." We may also use the phrase "DEIA" in this interview as a shorthand--when we say this, we mean all of these same considerations related to diversity scholarship.

There may be quotes or stories from your interview that we think would be valuable to share more widely. You can let us know now if you would like to remain anonymous or would like attribution. If you don't want to decide now, we will contact you with relevant sections of the interview for your approval, and ask at that time if you would prefer to remain anonymous, or would like attribution.

We will be recording this interview and transcribing the interviews, then coding them to find emergent themes and stories. We plan to use interview themes to build a survey on the same topics for broader distribution. After the survey, we will use the data we've collected to guide creation of a toolkit to support diversity scholars in making decisions about their data. Building this toolkit may include gathering existing resources, and/or creating new resources.

Do you have any questions about the consent form, or the interview, before we proceed?

Questions:

1. Please describe your area of research interest.

    - Prompt:
        - Who or what do you study, or want to study?
        - Why?
        - What questions do you hope your research will answer?
        - What effects do you hope your research will have in the world?

The next few questions are focused on data collection, storage, and dissemination. Please focus on how you interact with your data for these questions.

2. Please describe the type or types of data you typically work with in your research. If you are not yet conducting independent research, what type(s) of data does your mentor work with?

Follow up: Regardless of whether you are working with your own data at this time, what types of data do you see yourself likely using or collecting in the future?

- Please collect at least the following:
    - Quantitative vs. qualitative
    - Primary data collection (interacting with participants, asking them questions via a survey or interview guide) vs. secondary/existing data collection (historical data, big data, census or other publically available data)
        - If collecting primary data from humans, do you have direct contact with communities?
        - If so, how do you (plan to) make connections?
        - *If so*, how do you (plan to) maintain your connections?

3. Do you think of your research as happening with people--that is, with individuals or within populations, communities, or other types of groups?

[If "yes"]: Can you describe the ways you identify your research population and samples of interest? Please describe where you learned how to define your research populations.

- In your current or future research, do you…

- Plan to have direct connections with these communities?
    - *If so*, how do you (plan to) make connections?
    - Who (if anyone) has modeled research participant community connections for you?
    - How did they make connections with communities? How did they maintain connections with communities?
    - What were the strengths and limitations you noted in their approach?
- If not, how have you seen these connections made and maintained by your mentors, advisors, or faculty in your department?

[If "no"]: Can you describe your process for gathering or collecting data?

Prompts if needed:

- How do you think of your work in relation to communities or individuals who might be the topic of the data, or affected by your research?
- How/do you work with multiple stakeholders to gather data?
- How do you decide how and when to collect data?
- How do you decide what data to use?

4. We use the phrase "data life cycle" to mean all of the actions and decisions regarding data that take place over the course of research. Are you familiar with the data life cycle (**show and review the hard copy graphic**)?

- Are you familiar with the data life cycle?
- *[If yes]* Where or from whom did you learn about the data life cycle?
- *[If yes]*: Have you thought about the data life cycle in relation to your own or your mentor's work?
    - What factors do you think about when deciding how to manage or plan for your data?
    - How do you take into account the population or topic that you are studying?
    - How do you take into account the historical and current context in which you are planning to collect this data?
    - What considerations do you have for data security?
    - Are there ways you process, clean, or quality check your data that are impacted by the population you study, or the context you work in?
    - How do you envision people will engage with your data and your findings?
    - How do you decide (or plan to decide) where you publish/present your results?

5. Have you had mentors/prior experiences that have circled back to the participants (and/or communities) they worked with about the findings from their project/research?

- Have you had any opportunities to circle back to communities/participants in your own work?
    - If so, how did you do this?
    - What led you to do this? (i.e., modeling from a mentor, something you read, etc.)
    - What did you share with the population? (i.e., report, presentation, data)
        - [Qualitative only] Do you ask participants to confirm your data is accurate and complete, either before or after you interpret it?

6. What do you think of when you hear the statement that data should be made “as open as possible, as closed as necessary[1]”?

Prompts if needed:

- What do you think of when you hear the phrase “open data”?
- Who do you think should make decisions about the availability of data?

7. Have you been involved on projects/used data shared with you by other researchers (i.e. mentors, other research teams, etc.)? We are specifically asking about datasets that are not publically available for download.

- Prompts:
    - Who shared this data with you?
    - How was the data shared? (e.g., through a data repository or other means?)
        - Why do you think the researcher(s) shared the data with you in this way?
    - What considerations do you think the researcher(s) thought about when choosing whether or not to share their data? How to share it?
    - In your past work, have ***you*** ever shared data you collected/curated?
        - How did you share it?
        - What considerations did you use when deciding whether and how to share it?
        - Were there any steps you took to ensure security/anonymity? What about useability (e.g., codebooks)?

8. In your present and future work, do you think you will share your data with others?

If/when you share data, do you anticipate communicating restrictions or guidelines for its use? (E.g., through documents, or by depositing in a restricted repository).

Prompts if needed:

- How do you or would you derive those recommendations? Who should determine them?
- How would you communicate them?
- Do you have any data use materials you could share with us as examples? (e.g. sharing agreements, protocols, etc.)

9. Our project's end goal is to create a toolkit to support diversity scholars in making confident decisions that allow them to share their data appropriately. Here is a list of some example resources the toolkit could include: (*show list*)

- As a student, are there any barriers you see in using these tools?
- Would you be likely to use any of these resources? Why or why not?
  - *If so*, which three look most interesting to you, and why?
  - Are there any other resources (existing or theoretical) you can think of that would be useful but are missing from this list?
  - Any other comments?

10. Is there anything else we should have asked you about the intersection of your data practices with your diversity scholarship?

- Do you know of other graduate student diversity scholars we should talk with?
- Do you know of any other resources of which we should be aware?

Many thanks for sharing your thoughts on these issues!

(Give timeline for project's completion--mid-next year. Ask if they would like to be notified when we launch the toolkit.)