Using Institutional Data to Identify and Address Educational Inequities in STEM

Data Management Plan

Recognizing that data management is a complex process (Borghi & Van Gulick, 2021), we have drafted this plan in accordance with current guidance from the NSF EHR Directorate about data management (https://www.nsf.gov/bfa/dias/policy/dmpdocs/ehr.pdf). All project team members will be trained to manage project data through their university's standard trainings on human subjects data and data management. While all project team members will share responsibility for complying with this plan, the PI will assume primary responsibility for ensuring that the project activities are conducted with appropriate data management strategies.

Borghi, J. A., & Van Gulick, A. E. (2021). Data management and sharing: Practices and perceptions of psychology researchers. PloS one, 16(5), e0252047.

Types of data, metadata, samples, physical collections, software, curriculum materials, documentation, publications and other materials that may be produced during the project:

- Institutional Review Board applications and associated materials.
- University documents and artifacts related to the departmental equity reports (e.g., internal memos).
- Interview and focus group protocols and resulting audio recordings as well as associated consent and incentive forms, verbatim transcripts, and field notes, as applicable.
- Survey instruments and responses to surveys.
- Demographic information as well as students' grades in target, prerequisite, and subsequent courses, all obtainable from the university Registrar's Offices and associated databases administered by our Information and Technology Services units.
- Files that support quantitative data analysis, such as spreadsheets and databases, and files that support qualitative data analysis, such as coding documents created in qualitative software.
- Publication drafts and associated documents including those for conference presentations and journals.

Standards to be used for data and metadata format and content:

- The research team will use common digital data formats (such as the .docx, .txt, .pdf, .xlsx, and .csv formats normally associated with text documents and spreadsheets) and metadata standards specific to each aspect of the project. We will work to provide data in those formats for centralized collection, management, and storage. Only research team members will be provided secure access to these stored data.

Policies for access and sharing, including provisions for appropriate protection of privacy, confidentiality, security, intellectual property, and other rights and requirements:

Access and sharing:

- All data collected in an electronic format will be kept on secure, password-protected computers and servers (provisioned by our Information and Technology Services units) that are protected by a firewall and backed up daily. No one other than the research team members will have access to these data. Remote data access will require a virtual private network connection.
- All non-electronic data (e.g., written consent forms) will be scanned for electronic storage and then physically stored in a locked file cabinet in a secure space inside the locked office space of one participating member of the research team. These materials will be accessible only to the research team.
- Only secure platforms, such as Dropbox through University of Michigan which is covered by a Business Associate Agreement, will be used to transfer electronic files containing sensitive

information between members of the research team (we consider even deidentified data to be sensitive information).

Protection of privacy:

- Data will only be collected directly from participants if they consent to provide us the data for our use. Such data will be kept completely confidential.
- Intellectual property issues are not applicable.

Policies and provisions for reuse, redistribution, and the production of derivatives:

- Assuming reasonable timeframes and no more than incremental costs, to support maximizing the impact of our collected data, the project team will make deidentified data and codebooks available in electronic format to researchers outside the project team upon request after the project team has analyzed the data, with the only stipulation that credit is given to the original project team and to NSF. These activities will occur insofar as they are compliant with our Institutional Review Board protocols.
- Production of derivatives is not applicable.

Plans for archiving data, samples, and other research products, and for preservation of access to them:

Consistent with NSF's Public Access Plan (NSF 15-52), we commit to 1) submitting the data that support research results at the smallest "grain size" possible to the appropriate repository (usually as a supplemental file alongside a journal article) based on the type of publication, and including the NSF award number and attribution with the documentation of such data; 2) storing all deidentified data throughout the duration of the project and for at least three years after the final project report with any associated costs being covered by either our Information and Technology Services units or the home unit of one of the PIs; and 3) archiving appropriate and deidentified data in University of Michigan's "Deep Blue Data" archive (https://deepblue.lib.umich.edu/data/), where it will be available to anyone, anywhere in the world for free. We will leave the data in the Deep Blue Data archive for five years following the date of the last publication related to the project, as is custom for social science research.

Dissemination and sharing of research results:

Consistent with NSF's Grant General Conditions (GC-1) Article 51 entitled "Sharing of Findings, Data, and Other Research Products" and NSF's Companion Guidelines on Replication and Reproducibility in Education Research (2018), we commit to: 1) publishing significant results from the project in a timely manner with author lists that reflect the contributions of participating researchers, persistent identifiers as appropriate (such as digital object identifiers (DOIs) and ORCIDs), and proper attribution to NSF as a funding source including the NSF award number; 2) in the interest of promoting reproducibility, including in such publications all data cleaning steps such as criteria for data inclusion and exclusion, analyses (including any original code and codebooks), and comparisons to previous similar studies; and 3) sharing all primary and supporting materials to the fullest extent possible while maintaining participant confidentiality and privacy as outlined in our Institutional Review Board protocol.