

## **DATA MANAGEMENT PLAN**

### **1. Types of data produced**

The project will produce survey-based questionnaire data. These data will be in the form of spreadsheets, word processing, and text files (TXT and CSV). In addition, the project will produce computer code, database schema documentation, and metadata documentation in text and machine-readable file formats (HTML, JSON, CSS).

### **2. Data and metadata standards**

Questionnaire data will be stored in CSV formats with an accompanying data dictionary in JSON, Markdown, or XML. Databrary will use open data schema (schema.org, Id-schema.org, datacite.org) for all metadata, and the formats will be shared publicly.

### **3. Policies for access and sharing**

We will seek permission to share from survey participants using language developed by Databrary. Where participants give permission, we will share de-identified responses via a public access dataset (project or volume) on Databrary and preserved on the Penn State institutional repository, ScholarSphere.

All computer code and schemas will be shared publicly on GitHub ([github.com/databrary](https://github.com/databrary)).

### **4. Policies for reuse and redistribution**

Data will be made available for educational and research purposes. Access will be provided using the web-based Databrary application whose software is free and open source. Materials generated under the project will be disseminated in accordance with the policies of NSF and participating institutions. Publication of data shared on Databrary shall occur during the project, if appropriate, or at the end of the project, consistent with normal scientific practices. Users are provided the tools to cite data sources hosted on Databrary using an automatically-generated persistent identifier. No data may be redistributed outside the principles of the Databrary Access Agreement.

### **5. Plans for archiving and & preservation**

Databrary assets will be preserved indefinitely in a secure data center facility at NYU, Databrary's host institution, and mirrored on a server in upstate New York. These facilities are administratively managed by the Information Technology Services (ITS) group at NYU, the university's central IT organization. Central IT staff at both sites handle storage, network, and backup systems. Should the current file format for Databrary access copies become obsolete, Databrary would seek guidance and support from the NYU Libraries and NYU ITS staff prior to converting formats.

Research assets generated by the project will be shared through ScholarSphere (<https://scholarsphere.psu.edu/>) repository service administered by the University Libraries at Penn State. ScholarSphere has been developed with the intent of helping

researchers comply with grant-funding agency requirements. It enables dissemination and long-term preservation and curation (management, use, and re-use) of data.

The project team will work with the University Libraries curatorial staff to generate appropriate metadata and ensure proper and complete documentation of the data to maximize the ability to understand and reuse the data. DOIs will be assigned to the assets as needed ensuring the ability to uniquely identify and cite the data.

ScholarSphere assets resides in two data center locations, a primary (production) data center, and a secondary location for disaster recovery/business continuity. Both storage clusters are configured to recover from two simultaneous drive or single node failures without sustaining any data loss. All data is backed up to tape for disaster recovery, and is stored off-site.