

[Home \(/\)](#) ▶ [Services \(/services\)](#) ▶ [Research data services \(/services/data\)](#) ▶ Examples of data management plans

## Examples of data management plans

These examples of data management plans (DMPs) were provided by University of Minnesota researchers. They feature different elements. One is concise and the other is detailed. One utilizes secondary data, while the other collects primary data. Both have explicit plans for how the data is handled through the life cycle of the project.

### School of Public Health

#### featuring data use agreements and secondary data analysis

All data to be used in the proposed study will be obtained from XXXXXX; only completely de-identified data will be obtained. No new data collection is planned. The pre-analysis data obtained from the XXX should be requested from the XXX directly. Below is the contact information provided with the funding opportunity announcement (PAR\_XXX).

**Types of data:** Appendix # contains the specific variable list that will be used in the proposed study. The data specification including the size, file format, number of files, data dictionary and codebook will be documented upon receipt of the data from the XXX. Any newly created variables from the process of data management and analyses will be updated to the data specification.

**Data use for others:** The post-analysis data may be useful for researchers who plan to conduct a study in WTC related injuries and personal economic status and quality of life change. The Injury Exposure Index that will be created from this project will also be useful for causal analysis between WTC exposure and injuries among WTC general responders.

**Data limitations for secondary use:** While the data involve human subjects, only completely de-identified data will be available and used in the proposed study. Secondary data use is not expected to be limited, given the permission obtained to use the data from the XXX, through the data use agreement (Appendix #).

**Data preparation for transformations, preservation and sharing:** The pre-analysis data will be delivered in Stata format. The post-analysis data will also be stored in Stata format. If requested, other data formats, including comma-separated-values (CSV), Excel, SAS, R, and SPSS can be transformed.

**Metadata documentation:** The Data Use Log will document all data-related activities. The proposed study investigators will have access to a highly secured network drive controlled by the University of Minnesota that requires logging of any data use. For specific data management activities, Stata "log" function will record all activities and store in relevant designated folders. Standard file naming convention will be used with a format: "WTCINJ\_[six letter of data indication]\_mmddyy\_[initial of personnel]".

**Data sharing agreement:** Data sharing will require two steps of permission. 1) data use agreement from the XXXXXX for pre-analysis data use, and 2) data use agreement from the Principal Investigator, Dr. XXX XXX (XXXXXX@umn.edu and 612-xxx-xxxx) for post-analysis data use.

**Data repository/sharing/archiving:** A long-term data sharing and preservation plan will be used to store and make publicly accessible the data beyond the life of the project. The data will be deposited into the Data Repository for the University of Minnesota (DRUM), <http://hdl.handle.net/11299/166578>. This University Libraries' hosted institutional data repository is an open access platform for dissemination and archiving of university research data. Data files in DRUM are written to an Isilon storage system with two copies, one local to each of the two geographically separated University of Minnesota Data Centers. The local Isilon cluster stores the data in such a way that the data can survive the loss of any two disks or any one node of the cluster. Within two hours of the initial write, data replication to the 2nd Isilon cluster commences. The 2nd cluster employs the same protections as the local cluster, and both verify with a checksum procedure that data has not altered on write. In addition, DRUM provides long-term preservation of digital data files for at least 10 years using services such as migration (limited format types), secure backup, bit-level checksums, and maintains a persistent DOIs for data sets, facilitating data citations. In accordance to DRUM policies, the de-identified data will be accompanied by the appropriate documentation, metadata, and code to facilitate reuse and provide the potential for interoperability with similar data sets.

**Expected timeline:** Preparation for data sharing will begin with completion of planned publications and anticipated data release date will be six months prior.

### College of Education and Human Development

#### featuring quantitative and qualitative data

##### Types of data to be collected and shared

The following quantitative and qualitative data (for which we have participant consent to share in de-identified form) will be collected as part of the project and will be available for sharing in raw or aggregate form. Specifically, any

individual level data will be de-identified before sharing. Demographic data may only be shared at an aggregated level as needed to maintain confidentiality.

Student-level data including

1. Student-level data including:
  1. Pre- and posttest data from proximal and distal writing measures
  2. Demographic data (age, sex, race/ethnicity, free or reduced price lunch status, home language, special education and English language learning services status)
2. Teacher-level data, including
  1. Pre/post knowledge and skills data (collected via secure survey tools such as Qualtrics)
  2. Teacher efficacy data (collected via secure survey tools such as Qualtrics)
  3. Fidelity data (teachers' accuracy of implementation of Data-Based Instruction; DBI)
  4. Teacher logs of time spent on DBI activities
  5. Demographic data (age, sex, race/ethnicity, degrees earned, teaching certification, years and nature of teaching experience)
3. Qualitative field notes from classroom observations and transcribed teacher responses to semi-structured follow-up interview questions.
4. Coded qualitative data
5. Audio and video files from teacher observations and interviews (participants will sign a release form indicating that they understand that sharing of these files may reveal their identity)

### **Procedures for managing and for maintaining the confidentiality of the data to be shared**

The following procedures will be used to maintain data confidentiality (for managing confidentiality of qualitative data, we will follow [additional guidelines \(https://qdr.syr.edu/guidance\)](https://qdr.syr.edu/guidance)).

1. When participants give consent and are enrolled in the study, each will be assigned a unique (random) study identification number. This ID number will be associated with all participant data that are collected, entered, and analyzed for the study.
2. The link between participants' names and study ID numbers will be kept in a separate electronic file on secure University servers at each site (using password protected computers that are kept in locked offices when not in use), ensuring that all data prepared for analysis are de-identified. Directly identifying information will never be maintained in the same files.
  1. All paper data will be stored in locked file cabinets in locked lab/storage space accessible only to research staff at the performance sites. Whenever possible, paper data will only be labeled with the participant's study ID. Any direct identifiers will be redacted from paper data as soon as it is processed for data entry.
  2. All electronic data will be stripped of participant names and other identifiable information such as addresses, and emails.
3. During the active project period (while data are being collected, coded, and analyzed), data from students and teachers will be entered remotely from the two performance sites into the University of Minnesota's secure BOX storage (box.umn.edu), which is a highly secure online file-sharing system. Participants' names and any other direct identifiers will not be entered into this system; rather, study ID numbers will be associated with the data entered into BOX.
4. Data will be downloaded from BOX for analysis onto password protected computers and saved only on secure University servers. A log (saved in BOX) will be maintained to track when, at which site, and by whom data are entered as well as downloaded for analysis (including what data are downloaded and for what specific purpose).

### **Roles and responsibilities of project or institutional staff in the management and retention of research data**

Key personnel on the project (PIs XXXXX and XXXXX; Co-Investigator XXXXX) will be the data stewards while the data are "active" (i.e., during data collection, coding, analysis, and publication phases of the project), and will be responsible for documenting and managing the data throughout this time. Additional project personnel (cost analyst, project coordinators, and graduate research assistants at each site) will receive human subjects and data

management training at their institutions, and will also be responsible for adhering to the data management plan described above.

Project PIs will develop study-specific protocols and will train all project staff who handle data to follow these protocols. Protocols will include guidelines for managing confidentiality of data (described above), as well as protocols for naming, organizing, and sharing files and entering and downloading data. For example, we will establish file naming conventions and hierarchies for file and folder organization, as well as conventions for versioning files. We will also develop a directory that lists all types of data and where they are stored and entered. As described above, we will create a log to track data entry and downloads for analysis. We will designate one project staff member (e.g., UMN project coordinator) to ensure that these protocols are followed and documentation is maintained. This person will work closely with Co-Investigator XXXXX, who will oversee primary data analysis activities.

At the end of the grant and publication processes, the data will be archived and shared (see Access below) and the University of Minnesota Libraries will serve as the steward of the de-identified, archived dataset from that point forward.

### **Expected schedule for data access**

The complete dataset is expected to be accessible after the study and all related publications are completed, and will remain accessible for at least 10 years after the data are made available publicly. The PIs and Co-Investigator acknowledge that each annual report must contain information about data accessibility, and that the timeframe of data accessibility will be reviewed as part of the annual progress reviews and revised as necessary for each publication.

### **Format of the final dataset**

The format of the final dataset to be available for public access is as follows: De-identified raw paper data (e.g., student pre/posttest data) will be scanned into pdf files. Raw data collected electronically (e.g., via survey tools, field notes) will be available in MS Excel spreadsheets or pdf files. Raw data from audio/video files will be in .wav format. Audio/video materials and field notes from observations/interviews will also be transcribed and coded onto paper forms and scanned into pdf files. The final database will be in a .csv file that can be exported into MS Excel, SAS, SPSS, or ASCII files.

### **Dataset documentation to be provided**

The final data file to be shared will include (a) raw item-level data (where applicable to recreate analyses) with appropriate variable and value labels, (b) all computed variables created during setup and scoring, and (c) all scale scores for the demographic, behavioral, and assessment data. These data will be the de-identified and individual- or aggregate-level data used for the final and published analyses.

Dataset documentation will consist of electronic codebooks documenting the following information: (a) a description of the research questions, methodology, and sample, (b) a description of each specific data source (e.g., measures, observation protocols), and (c) a description of the raw data and derived variables, including variable lists and definitions.

To aid in final dataset documentation, throughout the project, we will maintain a log of when, where, and how data were collected, decisions related to methods, coding, and analysis, statistical analyses, software and instruments used, where data and corresponding documentation are stored, and future research ideas and plans.

### **Method of data access**

Final peer-reviewed publications resulting from the study/grant will be accompanied by the dataset used at the time of publication, during and after the grant period. A long-term data sharing and preservation plan will be used to store and make publicly accessible the data beyond the life of the project. The data will be deposited into the Data Repository for the University of Minnesota (DRUM), <http://hdl.handle.net/11299/166578> (<http://hdl.handle.net/11299/166578>). This University Libraries' hosted institutional data repository is an open access platform for dissemination and archiving of university research data. Data files in DRUM are written to an Isilon storage system with two copies, one local to each of the two geographically separated University of Minnesota Data Centers. The local Isilon cluster stores the data in such a way that the data can survive the loss of any two disks or any one node of the cluster. Within two hours of the initial write, data replication to the 2nd Isilon cluster commences. The 2nd cluster employs the same protections as the local cluster, and both verify with a checksum procedure that data has not altered on write. In addition, DRUM provides long-term preservation of digital data files for at least 10 years using services such as migration (limited format types), secure backup, bit-level checksums, and maintains persistent DOIs for datasets, facilitating data citations. In accordance to DRUM policies, the de-identified data will be accompanied by the appropriate documentation, metadata, and code to facilitate reuse and provide the potential for interoperability with similar datasets.

The main benefit of DRUM is whatever is shared through this repository is public; however, a completely open system is not optimal if any of the data could be identifying (e.g., certain types of demographic data). We will work with the University of MN Library System to determine if DRUM is the best option. Another option available to the University of MN, ICPSR (<https://www.icpsr.umich.edu/icpsrweb/> (<https://www.icpsr.umich.edu/icpsrweb/>)), would allow us to share data at different levels. Through ICPSR, data are available to researchers at member institutions of ICPSR rather than publicly. ICPSR allows for various mediated forms of sharing, where people interested in getting less de-identified individual level would sign data use agreements before receiving the data, or would need to use special software to

access it directly from ICPSR rather than downloading it, for security proposes. ICPSR is a good option for sensitive or other kinds of data that are difficult to de-identify, but is not as open as DRUM. We expect that data for this project will be de-identifiable to a level that we can use DRUM, but will consider ICPSR as an option if needed.

**Data agreement**

No specific data sharing agreement will be needed if we use DRUM; however, DRUM does have a general end-user access policy ([conservancy.umn.edu/pages/drum/policies/#end-user-access-policy](http://conservancy.umn.edu/pages/drum/policies/#end-user-access-policy) (<http://conservancy.umn.edu/pages/drum/policies/#end-user-access-policy>)). If we go with a less open access system such as ICPSR, we will work with ICPSR and the Un-funded Research Agreements (UFRA) coordinator at the University of Minnesota to develop necessary data sharing agreements.

**Circumstances preventing data sharing**

The data for this study fall under multiple statutes for confidentiality including multiple IRB requirements for confidentiality and FERPA. If it is not possible to meet all of the requirements of these agencies, data will not be shared.

For example, at the two sites where data will be collected, both universities (University of Minnesota and University of Missouri) and school districts have specific requirements for data confidentiality that will be described in consent forms. Participants will be informed of procedures used to maintain data confidentiality and that only de-identified data will be shared publicly. Some demographic data may not be sharable at the individual level and thus would only be provided in aggregate form.

When we collect audio/video data, participants will sign a release form that provides options to have data shared with project personnel only and/or for sharing purposes. We will not share audio/video data from people who do not consent to share it, and we will not publicly share any data that could identify an individual (these parameters will be specified in our IRB-approved informed consent forms). De-identifying is also required for FERPA data. The level of de-identification needed to meet these requirements is extensive, so it may not be possible to share all raw data exactly as collected in order to protect privacy of participants and maintain confidentiality of data.