Data Management Plan example from the University of Minnesota School of Public Health

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. Date 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.