

Master Data Management Plan

Submitted: February 27, 2017

Data description

This Master Data Management Plan (MDMP) defines the strategy for managing the digital final deliverables produced by the Collaborative Sciences Center for Road Safety (CSCRS), a National University Transportation Center funded by the U.S. Department of Transportation. Headquartered at the University of North Carolina Highway Safety Research Center, the CSCRS includes five consortium members: University of North Carolina (UNC), Duke University, Florida Atlantic University, University of California at Berkeley, and University of Tennessee at Knoxville. The aim and purpose of the MDMP is to present details of our plan to preserve and make available all final reports and datasets created by CSCRS projects.

This plan describes how the CSCRS will manage final documentation and the associated final datasets from projects for which CSCRS is a full or partial funder. For the purposes of the MDMP, CSCRS will use the following definitions:

- **Final Datasets:** The recorded factual material commonly accepted in the scientific community as necessary to validate research findings, i.e., the data needed to reproduce the final results.
- **Metadata:** The set of data that describes and gives information about and context for the dataset as recommended by the Data Documentation Initiative (DDI).
- **Dataset description document:** Describes all variables in the dataset, the measurement units used, variable labels, and label values. This document should specify the data position of each variable, describe the contents of each variable, and identify the range of possible codes and the meanings of the codes.
- **Code:** Any routines, scripts, queries, or other code needed or desired to reproduce final results. The code should contain comments or other documentation.
- **Final Documentation:** Final reports, papers, handbooks, guides, manuals, or presentations derived from research and produced as a final deliverable.
- **Final Deliverables:** Final documentation, final datasets with their associated metadata, data description documents and, if applicable, code.

CSCRS will only deposit final deliverables in the CSCRS archive. The final deliverables to be submitted to CSCRS by a Principal Investigator (PI) will be detailed in the project specific data management plan (DMP) submitted by the PI to CSCRS and accepted by CSCRS. CSCRS will keep a copy of the DMP for each CSCRS-funded project.

In contrast, individual CSCRS research project datasets may include a wide variety of information created and used in digital form including: experimental, observational, and simulation data; codes, software and algorithms; text; numeric information; images; video; audio; and associated metadata. These data will be stored according to the project DMP. CSCRS will not archive these raw datasets or interim products used to manage and analyze the data.

Plans for archiving and preservation

Final deliverables as defined above will be archived in the CSCRS Dataverse hosted by the H. W. Odum Institute for Research in Social Science located at the University of North Carolina at Chapel Hill. There, the data will be preserved and made publicly available online via the Odum Institute Dataverse Network virtual archive, <https://dataverse.unc.edu>. The Odum Institute virtual archive has been awarded the internationally recognized 2014-2015 Data Seal of Approval by meeting quality guidelines for trustworthy digital repositories. As a member of the Data Preservation Alliance for the Social Sciences (Data-PASS) and the Library of Congress National Digital Stewardship Alliance (NDSA), the Odum Institute is recognized as a trusted and well-established data archive in the social science research and archiving communities. As such, it provides a strong archival and data distribution resource to the CSCRS.

In addition, these final deliverables, including datasets and accompanying material, will also be made publicly available through links on the CSCRS website and be deposited in the National Transportation Library. The CSCRS Dataverse will store submitted CSCRS files in perpetuity. If UNC were to close their instance of the Dataverse, there is a survivorship agreement among the member institutions to ensure that all deposited files will continue to be hosted.

Data format and metadata standard

Datasets will be stored in platform-independent, non-proprietary digital formats. Data in proprietary formats such as Microsoft Word, Excel, PowerPoint, Access and Publisher may be stored temporarily, but CSCRS will archive final documents as portable document format (.pdf) files, and final data as comma-separated values (.csv) and/or text (.txt) files. The PI will submit final deliverables to CSCRS in both native format (Microsoft Word, Excel or PowerPoint commonly) and in .pdf, .csv or .txt, as described in the project DMP. These application-independent file types are less likely to lose their backward compatibility with new software upgrades. In the future, if other file types (i.e. audio, video, image, geospatial) need to be addressed, the optimal preservation format as defined by the Library of Congress will be accepted. For example, geospatial data currently require .shp, .shx, and .dbf files for preservation and future access, all of which would be archived.

When submitting the final datasets to our repository and any other depository required by the individual project, metadata, data descriptions and code will be stored with their associated final dataset files. The CSCRS-archived data files will be described with the DDI compliant metadata fields that have been widely adopted by the international data archives community. The fields populated during the creation of descriptive metadata – filename, data type, author, abstract, keywords, publisher, geographic coverage, temporal period of collection, response rate, funding, rights, etc. – will enhance search and discovery of project data.

Policies for access and sharing

To facilitate and encourage data sharing and re-use, de-identified public use versions of the final datasets will be made accessible online via the CSCRS Dataverse. Along with these data, the Dataverse will also preserve and provide access to supporting materials necessary to interpret and use the data appropriately. The CSCRS Dataverse will automatically assign a unique and persistent digital object identifier (DOI) for each dataset to ensure that the data will be accessible online even if their location should change. The DOI will also link the material to the CSCRS website and the National Transportation Library. The Dataverse system also generates standardized data citations to encourage attribution to the project team.

The CSCRS is not aware of any reasons that might prohibit the sharing and re-use of public-use versions of final datasets. An embargo period of one year may be placed on the data to allow the project team to retain first use rights of the data. It is important that investigators have the ability to produce and publish peer-reviewed manuscripts of their research before the data are available to external users. After the one-year embargo period, the data will be made publicly available via the CSCRS Dataverse. If survey data are included in the final datasets, responses will be voluntary, anonymous, confidential, and unidentifiable. Any other data with personally identifiable information (PII) will be anonymized prior to deposit. Results will be released only as aggregate statistics. PIs will be responsible for removing all

identifiers from datasets before release. In the event that a final dataset includes embargoed data, restricted access will be imposed until open access is granted by the PI or the one year embargo ends, or an exception to this timeframe is agreed to by CSCRS, whichever comes first.

During the active phase of the project period, data files will be stored and backed up on an access-restricted, password-protected server as determined by the PI and their affiliated university requirements. They will be responsible for ensuring that their chosen computing support follows optimal policies and procedures for safeguarding sensitive data. This includes provisions for secure data storage, transmission, access restrictions, and incident management. Only the PI and authorized project personnel will be allowed access to project files. In addition, the PI is responsible for addressing their respective Institutional Review Board (IRB) policies and ensuring compliance with any conditions placed on approval.

Any proprietary data acquired for a project will be handled in accordance with the data usage agreement as outlined in the Project DMP.

Policies for re-use, redistribution, derivatives

These materials are intended to be used for research. All content archived in .pdf format will be readable but will be secured against copying and editing.

All final datasets in .csv or .txt format will carry a Creative Commons, CC – “Public Domain Dedication.” The user may share by copying and redistributing the material in any medium or format and may adapt by remixing, transforming, and building upon the material. Dataverse Community Norms (<http://best-practices.dataverse.org/harvard-policies/community-norms.html>) as well as good scientific practices expect that proper credit is given via citation.

Data Management Timeline

Task	Timeframe
PI submits a project DMP with the work plan to CSCRS	Within 1 month of the project start date
CSCRS reviews the project DMP and either accepts it or provides revision comments to the PI	Within 1 month of DMP submission
PI submits revised project DMP, addressing comments received from CSCRS (if necessary)	Within 1 month of receiving CSCRS revision comments
Project ends	Date as stipulated in project work plan
PI submits final data sets and additional materials in original (not for archiving) and non-proprietary (.pdf, .csv, .txt) formats (for archiving) to the CSCRS (info@roadsafety.unc.edu)	Within 30 days of project end date

Task	Timeframe
<p>CSCRS:</p> <p>Secures all report and presentation (.pdf) files</p> <p>Verifies operability of all data (.csv and .txt) files</p> <p>Uploads all archival files to CSCRS Dataverse with appropriate access and restrictions on individual files for the embargo period</p> <p>Links final materials to project description on CSCRS website</p> <p>Sends the following information via one joint email to research.hub@dot.gov, NTLDigitalSubmissions@dot.gov, and TRIS-TRB@nas.edu:</p> <p>Final Report URL(s) or PDFs for any resulting publications</p> <p>URL(s) to, and associated descriptive metadata for, any final datasets from the research project</p> <p>The funding agreement number of the project</p> <p>The RH Display ID for the project</p> <p>ORCIDs (unique researcher IDs) for all project investigators, contributors, and publication author(s)</p> <p>Any documented project outputs or outcomes resulting from the research project (see Exhibit B for more information)</p>	<p>Within 60 days of project end date</p>
<p>CSCRS:</p> <p>Releases embargo (if applicable)</p> <p>Makes any previously restricted materials public (published) in Dataverse</p>	<p>No longer than one year after project end date unless an exception has been granted</p>

###