

Sponsoring Office: Office of the Assistant Secretary for Research and Technology

Grant Number: 69A3552348305

Project Title: Center for Connected and Automated Transportation (CCAT)

Center Director: Henry Liu
CCAT Director
Professor of Civil and Environmental Engineering
Phone: (734) 647-4796
Mobile: (651) 260-5876
Email: henryliu@umich.edu

Submitted By: Debby Bezzina
CCAT Managing Director
Phone: (734) 763-2498
Mobile: (734) 751-1778
Email: dbezzina@umich.edu

Submission Date: October 5, 2023

Version: 1.1

DUNS No.: 073133571

TIN No.: 38-6006309

Awarded To: Regents of the University of Michigan
3003 S. State Street
Ann Arbor, MI 48109

UM Account No.: F069665

Period of Performance: June 1, 2023 – May 31, 2029

Signature:



The CCAT Data Management Plan provides a framework for individual research projects that are awarded through the center. For each research project that is awarded, a project data management plan must be submitted that provides sufficient information to address the sections below. This DMP is a living document that will be continuously reviewed and updated throughout the research project's data lifecycle. Additionally, the researcher will submit a compliance statement. The compliance statement will include any deviations from the DMP framework set forth in subsequent sections. Furthermore, the researcher and everyone on their team must have an ORCID identifier. Lastly, the DMP must be approved by the center before work can begin on the project.

This plan only applies to those research projects that are federally funded. Research projects from in-kind or cost share funds are exempt from the plan but will be tracked the RiP system. The data from those research projects will only be made publicly available if it is not in conflict with the sponsor for the project and the IP will remain with the sponsor or as is stipulated in the agreement with the sponsor.

1. Data Description

The Center for Connected and Automated Transportation will conduct research spanning a broad array of topics related to connected and automated vehicles and infrastructure. The data collected will vary by project but may include:

- Vehicle state data from vehicle data bus and/or sensors
- Traffic data
- Weather data
- Video
- Audio
- DSRC messages and pcap files (SPaT, MAP/GID, TIM, BIM, BSM, etc.)
- Subjective survey responses

The CCAT UTC will require each individual researcher to submit detailed data descriptions for their individual research projects that will include:

1. Name of the data, data collection project, or data producing program.
2. Description of the purpose of the research.
3. Description of the data that will be generated in terms of nature and scale (e.g., numerical data, image data, text sequences, video, audio, database, modeling data, source code, etc.).
4. Description of the methods for creating the data (e.g., simulated; observed; experimental; software; physical collections; sensors; satellite; enforcement activities; researcher-generated databases, tables, and/or spreadsheets; instrument generated digital data output such as images and video; etc.).

5. The period of time data will be collected and frequency of update.
6. If using existing data, the description of the relationship between the data being collected and existing data.
7. List potential audiences and uses of the data.
8. The potential value of the data over the long-term for the home institution and for the public and/or research community in general.
9. Rationale for lack of public access, if appropriate, and name of person with primary responsibility for data stewardship.
10. Description of how the researcher will check for adherence to this data management plan.
11. Description of how the data will be protected from accidental or malicious modification or deletion prior to uploading it to PURR (see section 5). It is the researchers' responsibility to maintain the data until it is uploaded to PURR (see section 5).

2. Data format and metadata standards

The researcher shall adhere to the following:

1. Use the standard file formats for your field, utilizing non-proprietary formats such as CSV whenever possible.
2. If using proprietary data formats
 - a. Discuss rationale.
 - b. Software and version needed for a user to open and view the data or documentation files.
3. Describe the data process log to clarify the final version of data shared to the public.
4. Describe how alternative formats will be documented and why they are being.
5. List what documentation will be created in order to make the data understandable by other researchers.
6. Indicate what metadata schema is being using to describe the data. If the metadata schema is not one standard for their field, discuss rationale for using that scheme.
7. Describe how the metadata will be managed and stored.
8. Indicate what tools or software is required to read or view the data.
9. Describe the quality control measures employed.

3. Access Policies

The researcher will be required to address any access restrictions in the project DMP they submit to the CCAT Center DMP.

For project DMPs, the researcher will address issues and outline the efforts they will take to provide informed consent statements to participants, the steps they will take to protect privacy and confidentiality prior to archiving their data, and any additional concerns (e.g., embargo periods for the data). If necessary, the researcher will describe any division of responsibilities for stewarding and protecting the data among other project staff.

If the researcher will not be able to de-identify the data in a manner that protects privacy and confidentiality while maintaining the utility of the dataset, the researcher will describe the necessary restrictions on access and use.

If an individual research project includes human subject research, the researcher will be required to go through the University of Michigan IRB or their home institutions IRB, if they have one.

The researcher will be required to address the following:

1. Describe what data will be shared, how data files will be shared, and how others will access them.
2. Indicate whether the data contains private or confidential information. If so
 - a. Discuss how the data will be guarded against disclosure of identities and/or confidential business information
 - b. State the party responsible for protecting the data.
 - c. List what processes will be followed to provide informed consent to participants.
3. Describe what, if any, privacy, ethical, or confidentiality concerns are raised due to data sharing.
4. If applicable, describe how data will be de-identified before sharing. If not:
 - a. Identify what restrictions on access and use will be placed on the data.
 - b. Discuss additional steps, if any, that will be used to protect privacy and confidentiality.

4. Policies for re-use, redistribution, derivatives

For federally funded research, the University of Michigan or the home institution of the researcher holds the IP for data created by the project.

The researcher will be required to describe if they are transferring rights to the data archive, if they do not describe this, the home institution maintains the rights.

The researcher will be required to cite the data source and license under which they used the data in their project DMPs.

The final data set will be archived using PURR (see section 5). The researcher will need to agree to the terms of deposit (<https://purr.purdue.edu/legal/termsofdeposit>) and choose a license (<https://purr.purdue.edu/kb/licensingrestrictions/licenses>). Both of these items should be documented in the project DMP. U.S. DOT strongly recommends CC-BY (<https://ntl.bts.gov/publicaccess/managingrights.html>), which PURR supports. If CC-BY is not appropriate for the project, the research must provide rationale.

The researcher may elect to store data in an alternate repository, but must be approved by the CCAT leadership team to assess the long-term viability of the data storage.

In general, the researcher will address the following in their project DMPs:

1. Name who has the right to manage the data.
2. Indicate who holds the intellectual property rights to the data.
3. List any copyrights to the data. If so, indicate who owns them.
4. Discuss any rights be transferred to a data archive.
5. Describe how your data will be licensed for reuse, redistribution, and derivative products.

5. Plans for Archiving and Preservation

The PI should include the following language in their project data management plan. If not using PURR, the PI should include equivalent information about the data archive facility.

1. The CCAT UTC will archive all data on PURR, <https://purr.purdue.edu>. PURR conforms to the U.S. DOT data repository requirements and is in the process of applying for the status “Conformant Repository.” It is also in the re3data.org registry.
2. When a project submits a final report, the researcher will have 45 days to archive their data on PURR. The Center Director will then send the required information to the Research Hub and the National Transportation Library, as well as change the status of the research project from “Active” to “Complete” in the RiP system within 2 months of the completed project.
3. PURR will assign a Digital Object Identifier (DOI) that should be referenced in any reports and/or publications.
4. Because the archived data is published with DOI’s it is considered a publication by the PURR system and publication rates apply. The CCAT team will be allotted 10GB of space in the PURR system. The researcher must request space from the Center

- Director. If insufficient space remains, the researcher will be notified and the storage costs should be included in the proposal.
5. All data within PURR is fully duplicated on a regular basis to prevent catastrophic loss of information. Information is backed up and mirrored at another site to provide a means of recovery in case of disaster. This file duplication also prevents data loss in case of data corruption detected through regular fixity checks. Purdue University Libraries is a member of the MetaArchive Cooperative, a Private LOCKSS Network. Data within PURR is geographically and redundantly stored within the network. Data is stored in ways which will facility easy recovery in the case of a catastrophic loss or server fail.
 6. The depositor is required to acknowledge that the work does not contain software viruses or any other computer codes, files, or programs that are designed or intended to disrupt, damage, limit or interfere with the proper function of any software, hardware or telecommunications equipment or to damage or obtain unauthorized access to any system, data, or other information in PURR or any third party. Furthermore, PURR reserves the right not to accept or preserve objects found to be corrupted or dangerous to the repository.
 7. PURR will preserve data for 10 years.
 8. PURR, through DataCite, works with data center to assign persistent identifiers to datasets and other research objects. DataCite is developing an infrastructure that supports simple and effective methods of data citation, discovery, and access. Purdue is a full member of DataCite and PURR is using DataCite services for mining DOI for dataset published in PURR.

6. Change Log

Version	Date	Description
1.0	10/5/23	Original Draft
1.1	10/5/23	<ul style="list-style-type: none"> • Added version number to cover page. • Updated opening paragraph to add language regarding the DMP is a living document and will be reviewed and updated regularly and that it must be approved by the center prior to starting the project. • Updated Section 2.2 to require the PI to include the software and version needed to open the data and documentation. • Added section 6 Change Log