

CARMEN+: Center for Automated Vehicles Research with Multimodal AssurEd Navigation

**Highly Automated Transportation Systems Research
University Transportation Centers Program**

U.S. Department of Transportation Grant # 69A3552348327

Data Management Plan

Effective Date: June 1, 2023

PI: Dr. Zak M. Kassas, kassas.2@osu.edu

Consortium Members: The Ohio State University
Columbus, OH 43210

University of California, Irvine
Irvine, CA 92697

University of Texas at Austin
Austin, TX 78712

North Carolina Agricultural and
Technical State University
Greensboro, NC 27411



THE OHIO STATE UNIVERSITY

COLLEGE OF ENGINEERING

UCI University of
California, Irvine



TEXAS
The University of Texas at Austin



**NORTH CAROLINA
AGRICULTURAL AND TECHNICAL
STATE UNIVERSITY**

1 Overview

This data management plan (DMP) describes how the CARMEN+ University Transportation Center (UTC) will comply with U.S. Department of Transportation (USDOT) [policies](#) on the dissemination and sharing of research results. The DMP defines how CARMEN+ researchers will handle digital data both during and after a research project.

The CARMEN+ UTC will require each individual Project principal investigator (PI) or Faculty Lead, referred to hereafter as an Investigator, to submit to the Center detailed data descriptions and a data management plan for their individual research activities and projects per this master plan. A DMP may evolve with the research project and should be reviewed for possible revision whenever a data management procedure is changed. Modifications and changes to a DMP should be submitted to CARMEN+ for review and approval.

In general, project/research DMPs will include the following elements:

1. Data Descriptions: the anticipated final research data to be produced during the entirety of the active project period
2. Data Standards: the standards that will be used for data and metadata format and content
3. Data Access Policies: the policies for access and sharing the final research data
4. Data Derivatives: the policies and provisions for re-use, re-distribution, and the production of derivatives
5. Data Archiving and Preservation: the plans for archiving and preserving access to final research data and other research product

Guidance on these DMP elements are discussed herein.

2 Data Description

Investigators will describe the nature and source of data generated or used in their research activities in their individual DMPs, which will provide the following information:

1. Name of the data, data collection project, or data producing program.
2. Brief 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 over which the data will be collected and frequency of update.
6. If using existing data, a description of the source of the data and the relationship between any additional data to be collected and the existing data.
7. Potential users of the data, both internal and external to the Center.
8. The potential value of having the data over the long-term for your institution, the Center, and also for the public and/or research community in general.
9. Whether the data will be publicly accessible. If requesting permission not to make data publicly accessible, explain the rationale for lack of public access.
10. The party(s) responsible for managing the data.

3 Data Format and Metadata Standards

To the maximum extent practicable, and in accordance with the generally accepted practices in the field, CARMEN+ Investigators will use platform-independent and non-proprietary formats to ensure maximum utility of the data in the future. If unable to use platform-independent and non-proprietary formats, Investigators should specify the standards and formats that will be used and the rationale for using those standards and formats. Investigators must also identify the metadata standards you will use to describe the data.

Project-specific DMPs should address the following:

1. List in what format(s) data will be collected and stored. Indicate if they are open or proprietary.
2. Indicate what tools or software are required to read or view the data.
3. If using proprietary data formats, discuss the rationale for using those formats and how you will document the proprietary format. The Center may require Investigators to provide software to read the proprietary data format.
4. Describe the data processing log and version control procedures.
5. Describe the quality control measures to be used.
6. Indicate what documentation will be created in order to make the data understandable by other researchers. This should include a data dictionary.

7. Indicate what metadata schema will be used to describe the data. If the metadata schema is not one standard for the field, provide the rationale for using that scheme. The minimum metadata is author, contributors, title, abstract, keywords, project, and data source or data producing program including the [DCAT-US schema](#).
8. Describe how the metadata will be managed and stored.

4 Data Access and Sharing Policies

All data needed to validate CARMEN+ research findings, and/or useful to conduct new studies or for teaching products will be preserved for long-term access. These datasets are required deliverables. Following USDOT guidelines, the Center is committed to the principle that data collected using public funds should be available to the public and other researchers.

Investigators will be required to release their data, whenever possible, in an open license for reuse, redistribution and derivative products. Investigators are encouraged to release programming code as well. If this is not possible, Investigators will describe the reasons and rationale for not releasing the data and code. Data-specific restrictions for access or release, if any, should be clearly documented and submitted as part of the Investigator's DMP.

For additional details and requirements for individual DMPs, see the section on archiving and preservation, below.

For projects involving human subject research, Investigators 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 data). If necessary, they will describe any division of responsibilities for stewarding and protecting the data among other project staff. If Investigators will not be able to de-identify the data in a manner that protects privacy and confidentiality while maintaining the utility of the dataset, they will describe the necessary restrictions on access and use. If an individual research project includes human subject research, Investigators will be required to use The Ohio State University institutional review board (IRB) or their home institutions IRB procedures. Further, when working with, or conducting research that includes Indigenous populations or Tribal communities, CARMEN+ researchers will adhere to the [CARE Principles for Indigenous Data Governance](#).

Investigators shall address the following in their project DMP when they submit to the Center:

1. Describe what data will be shared, how data files will be shared, and how others will access them.
2. Indicate whether the data contain private or confidential information. If so;
 - (a) Discuss how you will guard against disclosure of identities and/or confidential business information.
 - (b) State the party responsible for protecting the data.

- (c) List what processes you will follow 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 you will de-identify your data before sharing. If not:
 - (a) Identify what restrictions on access and use you will place on the data.
 - (b) Discuss additional steps, if any, you will use to protect privacy and confidentiality.

5 Policies for Re-use and Re-distribution of Research Data

In general, the home institution of the Investigator is assumed to hold the IP for data created by the project, subject to the Terms and Conditions of the grant including those terms which grant certain rights and license to USDOT. The USDOT also reserves a royalty-free, nonexclusive and irrevocable license to reproduce, publish, or otherwise use and to authorize others to use the work for government purposes. Investigators will be required to describe if they are transferring rights to the data archive or another recipient. The data archive may require the transfer of some rights. For data reused or acquired from another source, the Investigator will be required to state the data source and the license under which the data was used in their research.

If an investigator intends to enforce specific terms for data use or a requirement for data citation through a license, for example some variant of the Creative Commons or GPL licenses, this must be indicated in the project-specific DMP and must be consistent with the Terms and Conditions of the grant. Describe any other legal requirements that might need to be addressed. Also include the following information as appropriate:

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

6 Plans to Archive and Preserve Research Data

Plans for archiving will follow and support the provisions of the USDOT Public Access Plan (<http://ntl.bts.gov/public-access>), including the archiving of final datasets and the creation and maintenance of persistent identifiers (e.g., Digital Object Identifiers (DOIs), handles, etc.) and will provide for maintenance of those identifiers throughout the preservation lifecycle of the data.

- The CARMEN+ UTC will archive all final datasets on a CARMEN+ community of CERN's Zenodo, <https://zenodo.org/>, which is an approved site for the USDOT.
- When a project submits a final report, the Investigator will have 60 days to archive their data on Zenodo.
- Investigators will maintain and backup the data until it is uploaded to Zenodo.
- Zenodo's procedures and policies for back-up, data recovery, retention, security and integrity are outlined in <https://zenodo.org/policies>
- Zenodo provides how back-up, disaster recovery, off-site data storage, and other redundant storage strategies will be used to ensure the data's security and integrity.
- Zenodo will retain data for the lifetime of the repository. Data in Zenodo is guaranteed for at least 20 years as articulated via Zenodo's Frequently Asked Questions, <https://zenodo.org/faq>
- Each data upload in Zenodo gets a DOI to make them easily and uniquely citable. Researchers should submit this Zenodo DOI to DOT as part of research outputs.
- Project reports and other products will be publicly available on the Center's website.

In addition, Investigator and project DMPs must include the following:

1. Identify where the data will be stored at the Investigators home institution (or other location).
2. Describe how back-up, disaster recovery, off-site data storage, and other redundant storage strategies will be used to ensure the data's security and integrity.
3. Describe how data will be protected from accidental or malicious modification or deletion prior to receipt by the archive.

Change Log:

2023-08-28: Original Draft

2023-10-12: NTL recommendations added.