



# CARNATIONS

**Center for Assured & Resilient Navigation  
in Advanced Transportation Systems**

## **Center Data Management Plan**

**U.S. DOT Grant Number:** 69A3552348324

**Grant Period:** 06/01/2023 – 05/31/2029

**Lead Institution:**

Illinois Institute of Technology

**Partner Institutions:**

Virginia Polytechnic Institute and State University

Stanford University

Chicago State University

University of California, Riverside

**Version 1.1**

Effective Date: 12/6/23

**For any questions or further information, please contact:**

Boris Pervan, Center Director

[pervan@iit.edu](mailto:pervan@iit.edu)

# Center Data Management Plan

The Center for Assured and Resilient Navigation in Advanced Transportation Systems (CARNATIONS) is a Tier-1 University Transportation Center (UTC) created to address the USDOT's research priority area of *Reducing Transportation Cybersecurity Risks*. CARNATIONS is specifically focused on Resilient Positioning, Navigation, and Timing (R-PNT) and resilient vehicle-to-everything (V2X) PNT-relevant communications in multimodal transportation. The Center will benefit existing and emerging transportation systems by addressing PNT's evolving vulnerabilities. Our education and workforce development programs will support underserved communities. We will collaborate with stakeholders to develop demonstrable R-PNT performance metrics, rigorous standards, and open evaluation methods to enable seamless technology transfer.

CARNATIONS is a consortium of partner institutions including Illinois Tech, Chicago State University, Stanford University, the University of California-Riverside, and Virginia Tech.

The purpose of this Data Management Plan (DMP) is to assist CARNATIONS faculty, staff, students, and partners in producing research outputs in accordance with U.S. DOT and UTC requirements. The CARNATIONS UTC mandates that the UTC center must submit a Data Management Plan (DMP) (for an example, see <https://doi.org/10.21949/1520571>) and ensure compliance with it. This DMP details the general requirements for all activities funded by the U.S. DOT UTC, including research, education and professional training, or outreach with the board of directors including industry, government, and academia.

In the pursuit of excellence and adherence to best practices in data management, all Principal Investigators (PIs) associated with the CARNATIONS project are required to follow the guidelines and policies outlined in this DMP. Recognizing the diverse and unique nature of each research project within CARNATIONS, PIs are empowered to tailor this DMP to the specific requirements of their individual projects.

Should a project's data management needs to deviate significantly from this Center DMP, PIs are encouraged to either amend this DMP with specific modifications pertinent to their project or, alternatively, develop a narrative Project DMP. This approach ensures that each project's DMP is finely tuned to its unique aspects, while still aligning with the broader principles and objectives of CARNATIONS and complying with U.S. DOT guidelines.

It is imperative for the PIs to detail how their project-specific DMPs diverge from the Center DMP, providing clarity and transparency. These individual project DMPs are envisioned as dynamic documents, evolving throughout the project lifecycle. Revised and newly developed DMPs will undergo review and approval by the CARNATIONS Executive Committee (EC), particularly when there are significant changes in the research scope, data collected, or

changes in project personnel. All DMP submissions, whether new or revised, will be sent to the CARNATIONS DOT Grant Manager for review and subsequent approval.

For further guidance and to ensure a comprehensive understanding of the U.S. DOT Public Access Plan, PIs can refer to the detailed guidelines available at U.S. DOT Public Access Plan at <https://ntl.bts.gov/public-access>.

By adhering to this flexible yet structured approach, CARNATIONS aims to foster a culture of responsible and effective data management, ensuring that our research endeavors are both impactful and compliant with the highest standards of data stewardship.

The project DMPs must include the specified sections outlined below, each addressing the respective prompts.

### **Data Description**

CARNATIONS anticipates collecting a diverse array of data types, formats and sources, including but not limited to:

- Publications, such as published papers, project reports, working papers, and conference briefs. These are typically in text and multimedia format.
- Data collected in experiments or simulations environment, provided by external sources, publicly available data sets. This data ranges from textual numerical data sets to binary and/or even proprietary formats.
- Center developed tools, such as models and software that process the data and provide the outputs in the publications. These typically are textual codes in specific programming language, or configuration and settings for commercial toolsets.
- Secured, sensitive, confidential, or protected data shared by agency or private entity. This can fall under the previously mentioned categories, but are different in that they cannot be publicly shared without permission or restrictions. Any restrictions for release, if any, should be clearly documented and marked, and relayed to the Center director for access restrictions details.
- Educational data: this includes course notes, syllabi, presentation slides, videos, webcasts, summary of attendees, outcomes, survey results, and internship reports.
- Outreach activities, such as seminar slides, videos, or webcasts, conference proceedings, meeting minutes and outcomes.

Each dataset will be meticulously cataloged in our data management system, ensuring a clear record of its relevance to our research objectives focused on improving urban transportation efficiency and safety. Designated team members will be responsible for this crucial task, ensuring strict adherence to both Center and project-specific DMPs. For secured, sensitive, or confidential data, stringent protocols are in place to ensure

compliance with privacy laws and contractual agreements. This includes proper anonymization of personally identifiable information (PII) and adherence to any specific restrictions on data sharing.

1. Name the data (or data producing program).
2. Name the person(s) responsible for managing the data and checking for adherence to the Center and project-specific DMPs.
3. Describe the relevance of the data to the research.
4. Describe the nature of the data (e.g., numerical, image, text, video, audio, database, source code, etc.).
5. Describe the methods for creating the data (e.g., simulated, observed, experimental, software)
6. If requesting permission not to make data publicly accessible, explain the rationale.

### **Data Formats and Standards**

In the CARNATIONS project, our approach to data storage formats is tailored to the unique characteristics of each dataset. We anticipate a spectrum of data types, each necessitating its optimal storage format.

Our choice of file formats, such as CSV for numerical data and script files for code, is driven by the need for both machine readability and ease of use across platforms. For media files like images and videos, we opt for universally compatible formats like txt and pdf, prioritizing open-source options wherever feasible.

While we strive for open-source and interoperable formats, we recognize that certain aspects of our research may necessitate the use of proprietary software. In these instances, we will undertake a careful evaluation to ensure compatibility and accessibility as much as possible.

Our commitment to standardized metadata, particularly through the use of .json files, aligns with our goal of ensuring that our datasets are easily discoverable and usable. The adoption of the Project Open Data Metadata Schema (<https://resources.data.gov/resources/dcat-us/>) is a testament to our alignment with U.S. DOT standards. This machine-readable .json metadata file format will be implemented for each primary source data to ensure interoperability and ease of discovery.

To uphold the integrity of our data, we implement rigorous quality control measures. This includes both automated validation processes and peer reviews, forming a robust system that guarantees the accuracy and reliability of our datasets. A detailed version control system will also be implemented, with clear documentation of data changes over time.

Code can be written in any language, with test instances designed to run specific input data to achieve expected output. This ensures that the tools and models developed can be applied and validated across different consortium institutions or partners. To facilitate replication and understanding, codebooks or README files will accompany the code.

To maintain data integrity and track its evolution, a detailed version control system will be implemented, with clear documentation of data changes over time. Quality control measures will include automated data validation checks and peer review of data, ensuring the reliability and accuracy of our datasets.

As general guidance, Principal Investigators (PIs) may consider addressing the following in their project-specific DMPs:

1. Describe the metadata that will be provided.
2. Provide the format(s) for the data to be collected.
3. Indicate what tools or software are required to read or view the data.
4. Describe data quality control measures to be implemented.

### **Data Access and Sharing**

The CARNATIONS project is committed to fostering open and accessible research. Accordingly, data generated by our research projects, supported by the Center and funded in whole or in part by U.S. DOT, will typically be made publicly accessible following the conclusion of each study. This approach aligns with our overarching goal of contributing to the wider scientific community and adhering to U.S. DOT's open-access mandates.

Throughout the research lifecycle, Center-wide data will be securely stored on a dedicated central server at the Illinois Tech CARNATIONS Center. Access to this data is stringently controlled through authentication methods linked to users' university-provided credentials. This ensures that only authorized personnel, such as principal investigators, co-principal investigators, and designated project team members, can access the data. Any exceptions to this policy, especially in cases involving non-disclosure agreements or other confidentiality clauses, will be diligently observed.

For each individual research project, data storage and access protocols will be tailored to meet the specific needs of the project and its PI. This bespoke approach allows for flexibility while ensuring data security and integrity.

We recognize, however, that certain datasets may contain sensitive information, such as PII, confidential business information (CBI), or restricted data. In such instances, we will take necessary measures to either restrict public access or anonymize the data to a degree that

balances confidentiality with the utility of the dataset. The decision-making process behind the level of accessibility for each dataset will be meticulously documented.

In cases where our research involves human subjects, adherence to the highest ethical standards is paramount. All researchers will comply with the Institutional Review Board (IRB) policies of their respective institutions. This includes obtaining IRB approval before commencing the study and ensuring informed consent for all participants. Special attention will be paid to protecting the privacy and confidentiality of participants, with clear strategies outlined in our data handling procedures.

Additionally, in research involving Indigenous populations or Tribal communities, the CARNATIONS team will align with the CARE Principles for Indigenous Data Governance, available at <https://www.gida-global.org/care>, acknowledging the unique considerations such studies entail.

To ensure a comprehensive approach to data access and sharing, PIs and project-specific DMPs developed by our researchers should address the following critical aspects:

1. Indicate whether proprietary data from external (e.g., commercial) sources that cannot be publicly shared will be used in the research project and implications of data sharing.
2. Describe whether there are any personal privacy, ethical, or confidentiality issues with sharing any of the data publicly.
3. Indicate whether there will be any expected embargo periods for any of the data, and clearly state their duration and rationale.

By adopting these principles and practices, CARNATIONS aims to strike a balance between open-access ideals and the responsible handling of sensitive data, ensuring our research is both impactful and ethically sound.

### **Re-use, Redistribution, and Derivative Products**

All datasets, barring proprietary restrictions, will be made available under the Creative Commons Attribution 4.0 International License to facilitate reuse and redistribution, in accordance with federal guidelines (

<https://www.transportation.gov/sites/dot.gov/files/docs/Official%20DOT%20Public%20Access%20Plan%20ver%201.1.pdf>). Furthermore, the PIs may consider addressing the following in their project-specific DMPs:

1. Indicate who holds the intellectual property rights and copyrights to the data. (The Illinois Institute of Technology or the home institution(s) of the PI(s) responsible for the research project will hold the IP and copyright for data and other materials created by the project.)
2. State that the US DOT 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.
3. State that project data will be uploaded to an open archive, except proprietary data from external sources and data under temporary embargo. However, no change of ownership is implied and no property rights are transferred. All uploaded content will remain the property of the parties listed in item 1 above.
4. Describe which data will be designated in the archive under open license for reuse, redistribution, and derivative products.

### **Archiving and Preservation**

To ensure the longevity and accessibility of our research data, CARNATIONS has implemented a comprehensive data preservation strategy. This strategy includes the periodic migration of datasets to current formats in line with industry standards and regular backups. Data is stored both locally and at off-site locations, providing robust safeguards against potential data loss.

Upon the completion of each project, the resulting data will be archived on Zenodo (<https://zenodo.org/>), a data repository conformant with U.S. DOT standards as outlined at <https://ntl.bts.gov/ntl/public-access/data-repositories-conformant-dot-public-access-plan>. This archiving process includes the assignment of a Digital Object Identifier (DOI) to each dataset, enhancing data discoverability and ease of reference. The data will be accessible via Zenodo's DOI-enhanced URI, as affirmed by Principal Investigators (PIs) in their project-specific DMPs.

In addition to Zenodo, PIs are encouraged to utilize other publicly available platforms for data archiving, as appropriate for their specific project needs and as identified in their DMPs.

At the CARNATIONS central facility at Illinois Institute of Technology, additional backups will be maintained. Here, the data is stored on physically secured hard drives, protected from unauthorized access and potential threats through robust security measures, including firewalls and project-specific authentication systems.

The metadata accompanying each dataset, as detailed in the "Data Formats and Standards" section of this DMP, will ensure data discoverability on Zenodo and efficient retrieval from CARNATIONS resident storage.

It is important to acknowledge Zenodo's commitment to data security and preservation:

- Zenodo's procedures and policies for backup, data recovery, retention, security, and integrity are comprehensively detailed in their policy documentation (<https://zenodo.org/policies>).
- The repository employs strategies such as backup, disaster recovery, offsite data storage, and other redundant storage measures to ensure the security and integrity of the data.
- Zenodo guarantees data retention for the lifetime of the repository, with a minimum commitment of 20 years, as articulated in their Frequently Asked Questions (<https://zenodo.org/faq>).

Through these measures, the CARNATIONS project is dedicated to maintaining the highest standards of data archiving and preservation, ensuring that our research outputs remain accessible and secure for the foreseeable future.

To reinforce the responsibilities of the Principal Investigators (PIs) in the context of archiving and preservation, the following points should be addressed in their individual project DMPs:

1. Archiving Platform Selection: PIs should specify their choice of archiving platforms, with Zenodo as the primary repository. If additional platforms are used, they should be publicly accessible and conformant with U.S. DOT standards.
2. DOI Assignment and Management: PIs are responsible for ensuring that a DOI is assigned to each dataset upon upload to Zenodo or any other chosen repository. This includes maintaining accurate and up-to-date records of these DOIs for easy referencing.
3. Metadata Provision: Ensure the provision of comprehensive metadata for each dataset, as outlined in the "Data Formats and Standards" section. This metadata should facilitate easy discoverability and accessibility of the data on Zenodo and other platforms.
4. Backup Procedures: PIs should detail the procedures for backing up data both at the CARNATIONS central facility and any other chosen storage locations, including frequency and methods of backup.
5. Data Security and Integrity: Outline the measures taken to protect the data from unauthorized access, malicious modification, or deletion. This includes details of security protocols and authentication mechanisms employed.



6. Compliance with Zenodo's Policies: Adherence to Zenodo's backup, data recovery, retention, security, and integrity policies should be documented. PIs should familiarize themselves with these policies and incorporate relevant aspects into their DMPs.
7. Long-term Data Retention: PIs should acknowledge the commitment to long-term data retention, in line with Zenodo's guarantee of a minimum of 20 years, and describe any additional steps taken to ensure the longevity of the data.
8. Review and Update of Archiving Strategies: Regular review and updates to the archiving strategies should be planned, in accordance with evolving project needs and technological advancements.

By addressing these points in their project-specific DMPs, PIs will contribute to the cohesive and effective archiving strategy of CARNATIONS, ensuring the long-term preservation and accessibility of valuable research data.

### Reference

United States. Department of Transportation. (2022). *Creating Data Management Plans for Extramural Research*. <https://doi.org/10.21949/1520571>

### Changelog

<u>Version</u>	<u>Date</u>	<u>Description</u>
<u>1.1</u>	<u>12/6/2023</u>	<u>Updated all sections and cover page to address DOT comments and change requests.</u>
<u>1.0</u>	<u>10/16/2023</u>	<u>Original Draft</u>