

## Data Management Plan University Transportation Center for Railway Safety (UTCRS)

The University Transportation Center for Railway Safety (UTCRS) has developed the following Data Management Plan for research data produced or acquired by UTCRS under US Department of Transportation (USDOT) sponsorship during the five-year grant period, beginning June 01, 2023, and ending May 31, 2028. This plan covers UTCRS activities at the lead institution, the University of Texas Rio Grande Valley (UTRGV), as well as the consortium members South Texas College (STC), Texas A&M University (TAMU), University of California Riverside (UCR), University of Nebraska Lincoln (UNL), and the University of South Carolina (UofSC).

### 1. Data Description

Data Collection Projects: All projects will be conducted through the University Transportation Center for Railway Safety (UTCRS).

Purpose of Research: The UTCRS supports a wide range of projects aimed at improving the safety and reliability of rail transportation. The projects will be conducted by the consortium member institutions in collaboration with rail industry partners and state and federal agencies. All projects are subject to review and approval by an external advisory board, composed of rail industry professionals, to ensure (a) direct relevance to the Center's mission, and (b) likelihood of leading to improved practices, policies, and technologies.

We, therefore, anticipate that the data, reports, publications, algorithms, and software produced by the Center will bring benefit to several categories of potential users: rail industry management and technical staff; state, federal, and local government agencies; researchers at peer institutions; and members of the public with an interest in rail safety.

Data Types and Methods Used: We anticipate collecting or producing a wide range of final data including, but not limited to:

- a. Raw and processed data collected from sensors on laboratory testers, rolling stock, track, and wayside systems,
- b. Video data collected from trains, grade crossings, unmanned vehicles, and laboratory tests,
- c. Simulation results resulting from physics-based analyses and design, as well as research in rail-relevant artificial intelligence and machine learning,
- d. Software in both source code and executable formats,
- e. Mathematical models,
- f. Design documentation including mechanical drawings, electrical/electronic schematics, CAD and EDA files, embedded code, and application programmer interfaces (APIs),
- g. Maps and geographic information system (GIS) data,
- h. Photographic images of rail components, infrastructure, and accident sites,
- i. Radar and lidar images, and
- j. Surveys.

Time Period: June 01, 2023 through May 31, 2028.

Potential Users of Data:

- a. Rail industry management and technical staff.
- b. Public transportation management and technical staff.
- c. State, federal, and local transportation agencies.
- d. Transportation researchers at academic institutions.
- e. Members of the public interested in rail safety.
- f. Data not covered by this plan includes raw data that has not been validated as correct, intermediate calculations and designs, intermediate versions of software, results that are found to be incorrect, and similar information that is not final. Personal information that is protected by federal or state regulations (e.g., FERPA, HIPPA) or Institutional Review Board (IRB) restrictions will also not be made available.

Potential Value of the Data: All UTCRS projects support rail transportation safety, reliability, and sustainability. The data produced may support those engaged in new equipment and infrastructure design, enable development of new advanced safety technologies, and inform management and regulatory decision making.

Limitations on Public Accessibility: Our intention is to make publicly available all information that could be of value to the potential users listed above. The following exclusions from this policy are requested:

- a. Intermediate or non-final results such as raw data that has not been validated as reliable; intermediate calculations; incomplete or tentative versions of software and hardware designs.
- b. Personal information restricted by federal or state privacy regulations (e.g., FERPA, HIPPA) or by Institutional Review Boards (IRB).
- c. Proprietary business information provided by, or collected in collaboration with, industry partners.
- d. For a temporary period not to exceed six months, information may be embargoed while manuscripts are prepared for publication or applications for patents or other intellectual property protection are filed.

Management Responsibility: The Center Director, Dr. Constantine Tarawneh, will be responsible for ensuring that the Data Management Plan is implemented. He will be assisted in management by Associate Directors at each of the consortium institutions.

Compliance Monitoring: The UTCRS is monitored by

- a. an External Advisory Board that will review periodic reports, and
- b. the UTRGV Office of Research, which reviews the Center's operation for compliance with contract requirements and federal regulations.

## 2. Standards Used

Data Format: The UTCRS will support a wide range of projects producing field data, laboratory data, simulation data, and theoretical results. A wide variety of sensors, instruments, and platforms will be used. It is not possible to list all possible formats that will be used; however, we are committed wherever feasible to using standard, widely recognized formats that can be opened using commonly available software, for example:

Documents:	TXT or PDF
Images:	JPEG, PNG, TIFF, EMF, or BMP
Raw Data:	CSV
Video:	MP4 or MOV

Design files produced with commercial software will be exported/archived to open-access formats. For wireless sensor work, we intend to place all data in an SQL-compatible database.

Internally Developed Formats: Part of our research involves the development of new sensors which may use new communication protocols. Very large data sets may need to be stored in raw binary form rather than CSV or TXT for efficiency. In such cases, we are committed to fully and publicly documenting the file format and making the format clear in the metadata.

Metadata: At a minimum, the metadata will include Title, Creators (e.g., PI), Contributors (e.g., co-PIs), Identifiers (e.g., researcher ORCID IDs, digital object identifier (DOI), grant number), Publishers (e.g., university performing research, sponsor), Description, Types, Formats, Subjects, Date of Collection, and Date of Submission. The established metadata will contribute to the discoverability and accessibility of the research data. To meet federal data discoverability guidelines, researchers will also create a DCAT-US schema (<https://resources.data.gov/resources/dcat-us/>) .JSON metadata file.

## 3. Access Policies

UTRGV will maintain a UTCRS Data Repository that will include final versions of data produced by all consortium partners and all projects. The Repository will have an online, public, free access section with no login requirement. This section will allow complete, single site access to all data and metadata, excepting items listed in the “Limitations on Public Accessibility” section above. Confidential data will be appropriately indicated in the database and stored for the purpose of preservation and will not be publicly accessible. Only the research team responsible for the confidential data may be granted access to it by UTCRS.

Individual PI’s will be responsible for protecting the identity and privacy of research participants, conducting their experiments according to the specific ethical codes and procedures of their institutions, and protecting proprietary information before submitting final data to UTCRS staff for archiving.

In addition, individual consortium partners and researchers may choose to make parts of the same data publicly available through GitHub, Dryad, the Open Science Framework (OSF), or local

archives such as the California Digital Library (CDL). However, regardless of any alternative sources, the final copies will be made available through the UTCRS Data Repository

#### 4. Re-Use, Redistribution, and Derivative Products Policies

Copyrights: Intellectual property rights are governed by the policies at each of the consortium institutions. Depending on these policies, copyrights (if any) may be held by individual researchers, students, or the institution. Items submitted for publication in conferences, journals, or books will typically have copyrights transferred to the publisher.

Further researchers are reminded:

1. Data, as a collection of facts, cannot be copyrighted under US copyright law.
2. UTCRS research carried out under a US DOT University Transportation Centers (UTC) program grant is federally funded. As stated in grant language and referenced in the University Transportation Centers (UTC) Grant Deliverables and Reporting Requirements: For Grants Awarded in 2023 Funded by the Bipartisan Infrastructure Law (BIL):
  - a. Researchers must comply with the US DOT Public Access Plan, meaning, among other requirements, research data must be shared with the public, either by the researchers or by US DOT;
  - b. That by accepting US DOT funding through this grant, researchers have granted to US DOT a comprehensive non-exclusive, paid-up, royalty-free copyright license for all research outputs (publications, datasets, software, code, etc.). This includes all rights under copyright, including, but not limited to the rights to copy, distribute, prepare derivative works, and the right to display and/or perform a work in public; and,
  - c. In accordance with Chapter 18 of Title 35 of the United States Code, also known as the Bayh-Dole Act, where UTCRS researchers elects to retain title to any invention developed under this UTC grant, US DOT retains a statutory nonexclusive, nontransferrable, irrevocable, paid-up license to practice or have practiced for or on behalf of the United States any such invention throughout the world.

None of the general IP provisions above negate US DOT's non-exclusive rights nor UTCRS's obligations.

Re-use and Redistribution Policy: Items made available in the public portion of the UTCRS Data Repository may remain copyrighted but are intended to be available, at no cost, for any fair use of copyrighted material. Users will be asked and expected to conform to accepted academic ethical standards, which require proper acknowledgement of original sources. Software and datasets will be covered under a version of the Gnu Public License or similar open licensing system, allowing redistribution and derivative products, but requiring notification of any modifications that have been made.

## 5. Archiving and Preservation Plans

Archive, URL, and Backup: The UTCRS Data Repository will be hosted at UTRGV and available through the center website at: <https://www.utrgv.edu/railwaysafety/>

The site is maintained by Center staff, with technical support and guidance from the UTRGV Information Technology (UTRGV IT) department. UTRGV IT provides standard services for all UTRGV sites, including redundant backup through secure, off-site storage. The department has extensive experience operating university websites with public-facing interfaces that protect and control access to critical data.

Collection and Archiving Intervals: Researchers will be encouraged to submit data for archiving as soon as it is finalized. At a minimum, they will be required to submit data as part of the mandatory reporting at the end of each project period (one year).

Protection from Modification/Deletion: Users of the public interface of the Data Repository will have *read-only access*. The ability to *add* content will be restricted to UTCRS personnel, using UTRGV provided credentials, and requiring two-factor authentication for off-campus access. The ability to *delete* content will be restricted to a single site manager and the Center Director.

Preservation: UTRGV intends to maintain the Data Repository contents throughout the life of the UTCRS, and to preserve availability of its contents for an additional period of five years beyond the life of the center.

### Change Log

September 18, 2023; Original Draft

October 15, 2023; Revised Document Based on USDOT Feedback