FTA RESEARCH

FEDERAL TRANSIT ADMINISTRATION

Automated Driving Systems (ADS) for Rural America Demonstration Data Management Plan (DMP)

OCTOBER 2020

FTA Report No. 0174
Federal Transit Administration

PREPARED BY
University of Iowa
National Advanced Driving Simulator

U.S. Department of Transportation
Federal Transit Administration

COVER PHOTO

Image courtesy of the University of Iowa

DISCLAIMER

Automated Driving Systems (ADS) for Rural America Demonstration Data Management Plan (DMP)

OCTOBER 2020

FTA Report No. 0174

PREPARED BY

Stephen Cable
Omar Ahmad
Greg Wagner
Cher Carney
Dan McGehee
University of Iowa
National Advanced Driving Simulator
2401 Oakdale Boulevard
Iowa City, IA 52242

SPONSORED BY

Federal Transit Administration
Office of Research, Demonstration and Innovation
U.S. Department of Transportation
1200 New Jersey Avenue, SE
Washington, DC 20590

AVAILABLE ONLINE

https://www.transit.dot.gov/about/research-innovation

Metric Conversion Table

SYMBOL	WHEN YOU KNOW	MULTIPLY BY	TO FIND	SYMBOL
		LENGTH		
in	inches	25.4	millimeters	mm
ft	feet	0.305	meters	m
yd	yards	0.914	meters	m
mi	miles	1.61	kilometers	km
		VOLUME		
fl oz	fluid ounces	29.57	milliliters	mL
gal	gallons	3.785	liters	L
ft^3	cubic feet	0.028	cubic meters	m^3
yd^3	cubic yards	0.765	cubic meters	m^3
		NOTE: volumes greater than 1000 L shall be shown in m^3		
		MASS		
oz	ounces	28.35	grams	g
lb	pounds	0.454	kilograms	kg
T	short tons (2000 lb)	0.907	megagrams (or "metric ton")	Mg (or "t")
		TEMPERATURE (exact degrees)		
^oF	Fahrenheit	5 (F-32)/9 or (F-32)/1.8	Celsius	^oC

REPORT DOCUMENTATION PAGE	*Form Approved* *OMB No. 0704-0188*

1. AGENCY USE ONLY	2. REPORT DATE October 2020	3. REPORT TYPE AND DATES COVERED Report, April 2020–May 2023

4. TITLE AND SUBTITLE Automated Driving Systems (ADS) for Rural America Demonstration Data Management Plan (DMP)	5a. FUNDING NUMBERS
6. AUTHOR(S) Stephen Cable, Omar Ahmad, Greg Wagner, Cher Carney, Dan McGehee	5b. CONTRACT NUMBERS IA-2020-005
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) University of Iowa National Advanced Driving Simulator 2401 Oakdale Boulevard Iowa City, IA 52242	8. PERFORMING ORGANIZATION REPORT NUMBER FTA Report No. 0174
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Department of Transportation Federal Transit Administration Office of Research, Demonstration and Innovation East Building 1200 New Jersey Avenue, SE Washington, DC 20590	10. SPONSORING/MONITORING AGENCY REPORT NUMBER FTA Report No. 0174

11. SUPPLEMENTARY NOTES [https://www.transit.dot.gov/about/research-innovation] [https://doi.org/10.21949/1518346]
Suggested citation: Federal Transit Administration. Automated Driving Systems (ADS) for Rural America Demonstration Data Management Plan (DMP). Washington, D.C.: United States Department of Transportation, 2020. https://doi.org/10.21949/1518346

12A. DISTRIBUTION/AVAILABILITY STATEMENT	12B. DISTRIBUTION CODE
Available from: National Technical Information Service (NTIS), Springfield, VA 22161. Phone 703.605.6000, Fax 703.605.6900, email [orders@ntis.gov]	TRI-30

13. ABSTRACT

Nationwide, rural communities and their roadways are disproportionately affected by road safety issues. Multiple complex factors contribute to transportation-related fatalities and injuries on rural roadways, including their infrastructure (sharp curves, steep grades, limited sight distances, sharp pavement drop-offs) and unique environments (slow-moving vehicles, animals, high speeds, extreme weather conditions). Whereas Automated Driving System (ADS) technologies bring the promise of even greater safety on our roadways, rural roads are underrepresented in ADS research today. Also, a relatively large and growing portion of rural and small-town residents are older Americans. As the US population significantly ages over the next few decades and Americans continue their desire to "age in place" (continuing to live in their own home regardless of age, income, or ability level), increased rural mobility options will be needed to maintain and improve the safety on these roadways. This 2.5-year demonstration project will use a custom, mobility-friendly vehicle with advanced technologies to gather publicly-available data for analysis. Data collected will help identify challenges, opportunities, and insights relevant for USDOT safety and rulemaking priorities. The project also will focus on studying ADS applications for aging, transportation-challenged rural populations. This report documents the University of Iowa's Data Management Plan (DMP) for this project and provides details regarding how UI will securely and reliably manage the data and provide access to the data to users inside and outside the organization.

14. SUBJECT TERMS	15. NUMBER OF PAGES
Rural, Mobility-impaired, Automated Vehicles, Automation, Demonstration, Data	37

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Abstract

Nationwide, rural communities and their roadways are disproportionately affected by road safety issues. Multiple complex factors contribute to transportation-related fatalities and injuries on rural roadways, including their infrastructure (sharp curves, steep grades, limited sight distances, sharp pavement drop-offs) and unique environments (slow-moving vehicles, animals, high speeds, extreme weather conditions). Whereas Automated Driving System (ADS) technologies bring the promise of even greater safety on our roadways, rural roads are underrepresented in ADS research today. Also, a relatively large and growing portion of rural and small-town residents are older Americans. As the US population significantly ages over the next few decades and Americans continue their desire to "age in place" (continuing to live in their own home regardless of age, income, or ability level), increased rural mobility options will be needed to maintain and improve the safety on these roadways. This 2.5-year demonstration project will use a custom, mobility-friendly vehicle with advanced technologies to gather publicly-available data for analysis. Data collected will help identify challenges, opportunities, and insights relevant for USDOT safety and rulemaking priorities. The project also will focus on studying ADS applications for aging, transportation-challenged rural populations. This report documents the University of Iowa's Data Management Plan (DMP) for this project and provides details regarding how UI will securely and reliably manage the data and provide access to the data to users inside and outside the organization.

1 Introduction

Iowa has a bold vision to lead the nation in improving rural roadway safety while providing significant public benefit to our communities and transportation-challenged populations such as the exponentially-growing older American population segment. By leveraging our previous on-road demonstrations and research as well as financial investment through the US Department of Transportation (USDOT), Automated Driving Systems (ADS) for Rural America will develop and execute a demonstration project that gathers and generates a wealth of publicly-available data on rural roadways that can address a variety of questions among a diverse set of end-users to safely integrate ADS into all US roadways.

Our region provides an ideal location to demonstrate the safe deployment of ADS technologies in rural environments that have variance in climate, weather conditions, road surface, traffic, and much more. What is encountered in Iowa throughout ADS for Rural America will be transferrable to rural areas across the US that are not addressed by current urban-centric demonstrations. The University of Iowa (UI) has a proven record of experience in ADS demonstration and a wealth of expertise at one of the most world-renowned driving research centers, the National Advanced Driving Simulator (NADS), here in the heart of rural America.

1.1 Project Goals

ADS for Rural America developed three major goals in addressing these issues and challenges:

- Improve safety on US roadways by beginning to lay the groundwork for the safe integration of ADS.
- Work to address disparities in US roadway system by focusing demonstrations and ADS data gathering on rural roadways.
- Demonstrate how ADS can be used to enhance mobility for transportation-challenged populations such as the aging populations in our rural communities.

To begin to address these goals and the complex, multifaceted safety challenges rural roadways pose, ADS for Rural America developed a demonstration project to gather publicly-available data for analysis that will help to identify risks, opportunities, and insights relevant for USDOT safety and rulemaking priorities. At the same time, ADS for Rural America will focus on testing ADS applications for the aging, transportation-challenged rural population.

2 Project Overview

Project Title	Project Goals and Objectives	Project Description	Project Lifecycle Phase	Project Performance Measurements
ADS for Rural America	(1) Improve safety on the US roadway system by beginning to lay the groundwork for the safe integration of ADS (2) Work to address disparities in the US roadway system by focusing demonstrations and ADS data gathering on rural roadways (3) Demonstrate how ADS can be used to enhance mobility for transportation-challenged populations such as older adults in rural communities. ADS for Rural America distinctly aligns with the USDOT's goals of safety, data for safety analysis and rulemaking, and collaboration.	ADS for Rural America will take place in eastern Iowa, driving a loop from a mid-sized city (Iowa City) through rural areas and small towns, collecting data on challenges associated with driving a highly-automated vehicle on rural roadways in a variety of traffic, weather, and lighting conditions. A custom, mobility-friendly automated vehicle (AV) built on a commercially-available platform will be used for the demonstration and will provide a use case for how ADS can connect rural populations. The demonstration will consist of eight phases that add additional automation and increase in complexity for ADS technologies. In total, 10 drives will take place during each phase to increase the breadth of data collected, including driving during varying lighting and weather conditions. The route will be driven in its entirety each phase. This approach will produce a dataset that provides comparative data across phases as the level of automation increases.	Post-Award	• Ability of vehicle to perform in particular situations (e.g., on gravel, at night, during snow) and as level of autonomy increases. Measures will include information from each vehicle's many sensors, status of automation, and number of takeovers required by the safety driver. • Metrics of roadmanship will be a leading measure of safety and will include safety-critical events, near- misses and crashes, abrupt accelerations or decelerations, time to collision (TTC), and other applicable metrics. • Data will be gathered from passengers of the AV through surveys and from biometric sensors (e.g., wearables) that will examine trust and anxiety related to their experiences. More details will be documented in the Project Evaluation Plan.

2.1 Change Control

This DMP will be a living document that will be updated over time. The preliminary DMP, submitted with the project proposal, has been revised to address feedback from USDOT provided after the kickoff meeting. Moving forward, UI anticipates the next revision to take place around the following milestones:

- After completion of Project Evaluation Plan (07/13/20)
- After receiving approval from Institutional Review Board (IRB) (est. 11/01/2020)
- After completion of Data Dictionary (est. 04/29/21)
- After completion of first phase of data collection (est. 08/10/21)

Version	Date	Description
1.0	13-JUN-2020	Revised DMP based on feedback from USDOT, 60 days after award
2.0	22-JUL-2020	Revision based on feedback from USDOT

2.2 Relevant Documents

2.2.1 Vehicle Automation Component Documentation

The following are technical documents describing computing modules and sensors used on the vehicle to capture data described in Section 5.3.3, Metadata Collection Examples:

- By-Wire Kit, PACMod: https://autonomoustuff.com/product/pacmod/
- Computing, Spectra: https://autonomoustuff.com/product/astuff-spectra/
- Mapping, Mandli Communications High Definition (HD) Maps: https://autonomoustuff.com/product/mandli-communications/
- Sensing – Radar, Continental ARS 408-21: https://autonomoustuff.com/product/continental-ars-408-21/
- Sensing – Vision, Mobileye Camera Development Kit: https://autonomoustuff.com/product/mobileye-camera-dev-kit/
- Sensing – Laser Imaging, Detection, and Ranging (LIDAR); Velodyne Puck Hi Res: https://autonomoustuff.com/product/velodyne-puck-hi-res/
- Sensing – LIDAR, Velodyne HDL 64E: https://autonomoustuff.com/product/velodyne-hdl-64e/
- Communications – Wireless, Cohda Wireless MK5 OBU: https://autonomoustuff.com/product/cohda-mk5-obu/
- Positioning, Novatel Level 2.5 Kit: https://autonomoustuff.com/product/novatel-vehicle-kits/

- Video, Leopard Imaging USB 3.0 Cameras: https://leopardimaging.com/product/usb30-cameras/usb30-box-cameras/li-usb30-ar023zwdrb/

2.2.2 Supplemental Sensors

The following links describe sensor devices used to capture physiological data from the safety driver and passengers on each drive:

- Vaisala Mobile Detector MD30: https://www.vaisala.com/en/products/instruments-sensors-and-other-measurement-devices/weather-stations-and-sensors/md30
- Biovotion Everion Physiological Armband: https://biovotion.zendesk.com/hc/en-us/categories/201377109-Everion-Device-
- Empatica E4 Physiological Wristband: https://www.empatica.com/research/e4/

3 Data Overview

ID	Dataset Title	Description	Type / Scale	Collection Method	Data File Format(s)
https://orcid.org/0000-0001-9299-0555 https://orcid.org/0000-0001-6741-4433 https://orcid.org/0000-0002-7019-3317 https://orcid.org/0000-0002-7704-2132 https://orcid.org/0000-0001-5340-7398 https://orcid.org/0000-0002-9759-4743	ADS for Rural America	Time-correlated data from 80 drives of AV outfitted with a suite of sensors over 2.5-year span. Additional recorded data include safety driver and participant (rider) physiological data, participant user interface interaction data, survey data from participants.	Numerical data, text sequences, positional data (e.g., latitude, longitude), image data (external and internal cameras), recorded for duration of each vehicle drive.	Robot Operating System (ROS) bag files, custom video data collection software (synchronized with ROS recordings), OEM physiological data recording, multiple Qualtrics surveys.	SQL
	ADS – Robot Operating System (ROS) Bag File(s)	Default logging format for storing ROS messages in files; efficient way to record accurately timestamped sensor and vehicle data; widely used in automation and simulation or automated vehicles.	Numerical data, text sequences, positional data (e.g., latitude, longitude), image data (external cameras).	Default logging format for storing ROS messages in files; efficient way to record accurately timestamped sensor and vehicle data; widely used in automation and simulation or automated vehicles.	.BAG (ROS)
	ADS – Physiological Sensor Data	Sensor exports (from OEM device) are natively in .CSV format, including timestamps.	Numerical data, text sequences.	Exports from data collections using armband physiological sensors; recordings from individual riders and driver during each drive, including baseline observation recording from each rider/driver made pre-drive.	.CSV
	ADS – Rider Survey Data	Qualtrics survey software native exports (.CSV).	Numerical data, text sequences.	Pre- and post-drive surveys administered to riders.	.CSV
	ADS – Driver/ Rider State Monitoring	Video recordings of passenger compartment and driver, capturing input video to be used for post-processing to determine driver and rider state.	Video files.	Multiple camera viewpoints within cabin recorded during each drive.	.MP4

4 Data Stewardship

4.1 Data Owner and Steward

Dataset Title	Data Owner	Data Steward	Federal Sponsor
ADS for Rural America	USDOT	University of Iowa	Steve Mortensen

4.1.1 Data Steward Roles

Member Name	Project Role	Responsibilities
Stephen Cable	IT & Data Quality Lead	Ensure vehicle-side (and external data sources) data collection and transfer to cloud infrastructure.
Matthew McLaughlin	IT Project Manager	Manage cloud infrastructure (storage, SQL, other back-end infrastructure).
Alan DenBleyker	Application Developer	Write/manage data collection ingestion scripts, write/manage Data Access portal (front end for Data Dictionary, Playground, general statistics).

4.2 Access Level

4.2.1 Data Sharing

Can all data from this project be shared with the public or is controlled-access required for at least some of the data?

☒ All Public Access ☒ Some/All Controlled-Access

4.2.2 Datasets Requiring Controlled-Access

This section is required if "Controlled-Access" is selected above.

Dataset Title	Reason(s) for Controlled-Access	Safeguarding Methods and Processes
ADS for Rural America (video data of participants)	Video data of participant faces is personally identifiable data (PII) and protected from general access.	Access to PII data will require requesting organization to enter into a data usage agreement with the University of Iowa.

4.2.3 Informed Consent

All potential research participants will be provided with an Informed Consent document to review before their scheduled drive. Upon arrival at the pick-up location, a researcher will review the Informed Consent document with the potential participant and answer any questions. If he/she wishes to participate, both parties will sign the Informed Consent document. The following language pertaining to privacy and confidentiality is in the consent form currently being reviewed by the UI IRB:

Data Storage for Future Use

As part of this study, we are obtaining passenger riding behavior, video, physiological measures, and questionnaire data from you. We would like to study these patterns and responses in the future, after this study is over. Other qualified researchers who obtain proper permission may gain access to your data for use in approved research studies that may or may not be related to the purpose of this study. This process could occur without additional informed consent from you.

These future studies may provide additional information that will be helpful in understanding how driving performance varies among individuals, but it is unlikely that what we learn from these studies will have a direct benefit to you. It is possible that your riding behavior, video, physiological data, and questionnaire data will be used to develop products, tests, or discoveries that could be patented and licensed. In some instances, these may have potential commercial value and may be developed by the investigators, University of Iowa, commercial companies, organizations funding this research, or others that may not be working directly with this research team. There are no plans to provide financial compensation to you should this occur.

Your riding behavior, video, physiological data, and questionnaire data will be stored without your name. Your data will be linked by a subject identification number to the date of your participation. Your riding behavior, video, physiological data, and questionnaire data will be stored and available for use in future research studies for five years and cannot be removed.

What about Confidentiality?

We will keep your participation in this research study confidential to the extent permitted by law. However, it is possible that other people such as those indicated below may become aware of your participation in this study and may inspect and copy records pertaining to this research; some of these records could contain information that personally identifies you:

- Federal government regulatory agencies,
- The study sponsor, USDOT, and
- Auditing departments of the University of Iowa and the University of Iowa Institutional Review Board (a committee that reviews and approves research studies)

Additionally, organizations or individuals that sign a Data Usage Agreement with the University of Iowa will have access to the data. This may include data that could indirectly identify you, such as your video data. A major component of this research effort is providing a dataset for public use, and we will have protections in place so that only vetted parties can access this data.

To help protect your confidentiality, we will separate your name from your data. A coding scheme will be used to identify your data by a subject number only (Driver 1). The list linking your name and your study subject number will be stored in a secure location that is accessible only to the UI researchers. Any video files, as well as other data, will be stored in a data storage area located at a university-owned, physically secure data center.

The video data (video image data recorded during your drive) recorded in this study includes your video-recorded likeness. Video will be used to examine your behavior as a passenger while riding in the research shuttle. Video image data (in continuous video or still formats) may be publicly released, either separately or in association with the appropriate engineering data for scientific, regulatory, educational, outreach, legislative, or research purposes (as noted above). It will also be shared with the organizations or individuals that complete a Data Usage Agreement. Your name will never be shared, but other potentially identifying information may be linked to your video (e.g., age group, gender).

Data collected from the video recordings and a report of the investigators' findings may be shared with the UI General Counsel Office. The data will include information that could identify you. Under extreme and extraordinary circumstances and under advisement from the General Counsel, law enforcement may be provided information that includes your identity. Authorities could subpoena the video recordings if accidents or illegal activities are recorded.

If we write a report or article about this study or share the study dataset with others, we will do so in such a way that you cannot be directly identified.

Upon approval of the IRB application, UI will submit an updated DMP that will contain a link to the approved Informed Consent document.

4.2.4 Access Requests

The main point of access to ADS project data will be through the ADS Data Access portal, a public website that takes project data from the ADS Digital Clearinghouse Database and represents it in two ways—the Data Dictionary, which defines all metadata terms and displays bulk statistics, and the Playground, which allows users to view parts of drive(s) depending on queries against terms from the Data Dictionary. The URL of the ADS Data Access website is TBD.

Collecting data and making that data available is the key goal of this project. UI wants to make data accessible but also to safeguard data with PII. UI would like to have metrics on who is using the data and how they are planning to use it (so UI can support them if needed), but also not make it onerous for users to request access to the data. Table 4-1 shows how UI will provide access to four different levels of users and what UI will make available to each class. UI will create a web-based front end, accessible from anywhere in the world, that will be the interface that users will request access from.

Table 4-1

User Levels and Access to Data

User Level	Data Dictionary	"Playground"	All Data/PII
Unregistered	✓		
Registered	✓	✓ (max 2 variables)	
Qualified	✓	✓ (max 10 variables) (reverts to registered two months after approval)	
Qualified (with Data Usage Agreement [DUA])	✓	✓	✓ (ROS bag files, custom queries/ output)

Blurring out PII such as participant faces is beyond the scope of this project due to the large volume of data. Furthermore, UI is interested in enabling research on driver/passenger state detection, which would be better served by not altering the recorded images of participant faces. For these reasons, our approach to protecting PII is to provide access to it through the approach described below. This approach enables greater utility for researchers that desire access to PII data and find it valuable to enter into a DUA.

4.2.4.1 User Levels

Unregistered – Unregistered users are default viewers of the ADS Data Access website. There is no login information required and no information gathered (outside of Google Analytics).

Registered – Registered users are those who have completed a login form on the ADS Data Access website to create personalized credentials. Default information gathered from this form will include name, email, organization, organization affiliation (academic, governmental, consulting, etc.), and area of expertise in addition to a chosen, complex password. Registered users will be allowed to query for two specific variables (terms selected from the Data Dictionary). Data presented are de-identified and contain no video. The registration form includes basic CAPTCHA (Completely Automated Public Turing Test to tell Computers and Humans Apart) or equivalent challenge. All forms will be reviewed by designated personnel at the National Advanced Driving Simulator (NADS) before Playground access is granted.

Qualified – Qualified users are people (registered users) who specifically ask for deeper access to the dataset. Via an online form within the ADS Data Access website, a registered user can request temporary elevation. The form must include a detailed reasoning/plan for data derived, which is reviewed by NADS staff before approval. Qualified users will be allowed to query the dataset by 10 (or fewer) variables. Data presented may include forward- and rear-facing vehicle camera views but not participant-facing identifiable video. Two months after granting qualified access, the account automatically will revert to registered levels.

Qualified (with Data Usage Agreement) – Qualified users who need the greatest access (raw bag files or custom queries that might include video) will need to consult directly with NADS staff. These types of access require a formal data-sharing agreement and may require IRB training and/or IRB modification to avoid privacy, ethical, or confidentiality concerns.

4.2.4.2 Access Definitions

Data Dictionary – This is a menu-driven responsive web page that lists and defines the available recorded information in a dataset. Any given sensor might have multiple messages (or sub-messages) attached. An ROS message is a data structure with one or more fields.

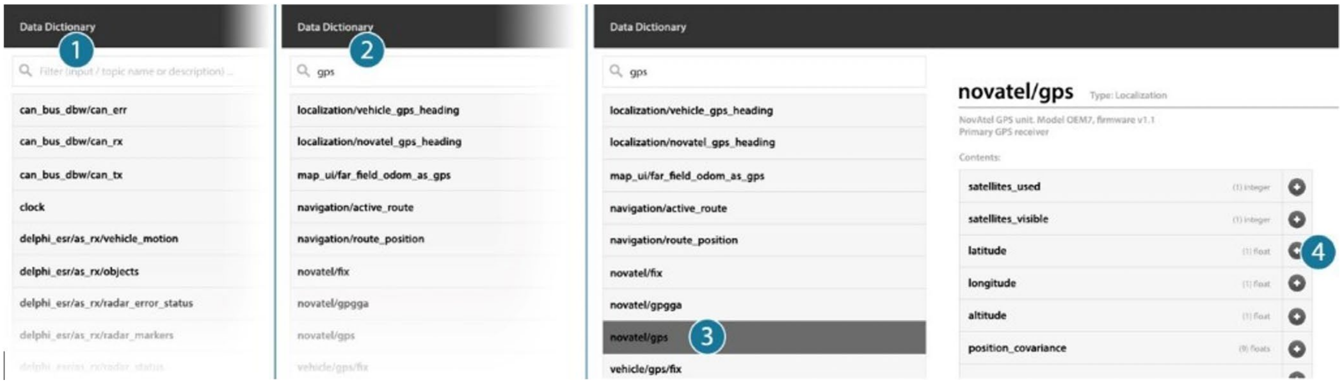

Figure 4-1 *Data Dictionary Message Name Search / Discovery Example*

Also available in the Data Dictionary is a set of global (summary) statistics for the dataset. This list of statistics would be defined by NADS staff and viewable by all (registered or unregistered). Summary statistics would include:

- Number of drives completed
- Total miles driven
- Total number of takeovers / engagements of automation
- Demographic range of passengers

"Playground" – This refers to the portion of the ADS Data Access website that will allow users to make queries against the dataset based on variable names used in the Data Dictionary. The number of variables available to be used in any given query would be set by the user's access level (see Table 4-1). Output is a visual table and/or timeline that shows data for the variables requested. Examples of queries include the following:

- How many takeovers occurred in drive X?
- What did the road surface "look" like when takeovers occurred in drive Y?
- How did the heart rates of the passengers change when the automation disengaged?
- What was the safety driver's state during takeovers in drive Z?

All Data/PII – In ADS terms, this refers to the highest set of access credentials. This would be done via specific request (coordinated and vetted by appropriate NADS staff) and would involve:

- Custom scripts (MATLAB, SQL, etc.) (output TBD)
- Access to raw ROS bag files, which would include all data include unredacted video. Transfer could be by internet but is most likely by shipped drives (HDD, SSD, etc.) due to the size of the data.

4.2.5 Related Tools, Software, and/or Code

Data Tool/ Software	Tool/Software Accessibility	Pertains To	Availability
Robot Operating System (ROS)	https://www.ros.org/	Accessing raw ROS bag files (qualified with DUA-only access)	Free

4.2.6 Relevant Privacy and/or Security Agreements

An Informed Consent document (Appendix A) is required for all research involving human subjects conducted at the University of Iowa. This document describes the demonstration project to those being recruited as passengers of the AV and helps potential subjects to decide whether they want to participate. The Informed Consent document provides important information about what subjects will be asked to do, the risks and benefits of the research, and their rights as a research subject. This document is currently under review by the UI IRB.

The Data Usage Agreement (DUA) (Appendix B) is a contractual document that will be used for the transfer of all non-public data that are subject to some restriction on their use (i.e., PII). When sharing data housed at UI with an outside organization, UI must consider multiple security and compliance criteria. All DUAs will be negotiated and signed by the office responsible for the particular dataset to be shared. The DUA has been reviewed and approved by the UI Division of Sponsored Programs, although it may be necessary to negotiate each individually, as the type and amount of information requested will likely be different for each data request.

4.2.7 Data Rights

Data generated in this project will be jointly owned by UI and USDOT with the exception(s) noted below. UI will manage the data, and it has budgeted for resources for five years after completion of the project to store the data and provide access to users as described in this document. USDOT will have access to all data in a controlled-access environment.

After five years, the data will be deleted from UI's system, and the ADS Data Access website will be decommissioned, barring future funding to allow for the data to be retained. Data may also be transferred to USDOT systems through arrangement.

IP rights to a subset of the data will belong to the companies whose IP was used to generate the data. UI has identified one specific instance with HD map data, which will remain under ownership of Mandli Communications. However, Mandli is licensing the map data in perpetuity to UI with restricted uses; non-commercial research and development (R&D) with the ability to share data with users as long as they agree to similar terms. UI believes this is a reasonable approach to protecting the company's IP while providing access to relevant data to USDOT and all users of data generated from this project. UI will pursue a similar approach to balance proprietary rights and public access should more issues like this be encountered.

All data stored on the cloud will be available to users for research (non-commercial) purposes.

4.3 Re-Use, Redistribution, and Derivative Products Policies

Dataset Title	License Used	Reason(s) for Non-Open License
ADS for Rural America	Creative Commons BY-SA 4.0	N/A
ADS for Rural America (containing video data of participants)	Proprietary	Files containing PII (participant video) cannot be redistributed without consent/redaction

Data Citations – In the DUA, UI will ask all recipients to agree to recognize the contribution of the provider (UI) as the source of the data in all written, visual, or oral public disclosures concerning the recipient's research using the data, as appropriate in accordance with scholarly standards. UI will also ask all recipients to acknowledge that funding was provided by the USDOT under grant agreement #IA-2020-005.

Other Legal Requirements to be Addressed – UI is reviewing the protection of privacy and ethical use of the data balanced with streamlined data-sharing. With a tiered sharing approach, UI hopes to make basic de-identified data available without an agreement or possibly a click-accept acknowledgement of ownership and security. For sharing of identifiable data or sharing with for-profit entities, UI will execute a Data Transfer and Use Agreement modeled after the national Federal Demonstration Partnership (FDP) template (http://thefdp.org/default/committees/research-compliance/data-stewardship/).

4.4 Data Storage and Retention

4.4.1 Storage Systems

Data Storage System Name	Data Storage System Type	Dataset Title(s)	Initial Storage Date	Frequency of Update	Archiving and Preservation Period
University of Iowa Research Data Storage	UI-managed – Controlled-Access System	ADS for Rural America (raw data)	13 months after award	As data collection drives occur (10+ every 3 months for 2.5 years)	Five years
University of Iowa MSSQL	UI-managed – Controlled-Access System	ADS Digital Clearinghouse Database	13 months after award	As data collection drives occur (10+ every 3 months for 2.5 years)	Five years

4.4.2 Data Storage System Description

The UI Research Data Storage (RDS) comprises two dedicated data center facilities located on the UI campus. Equipment is housed in an environmentally-sound, restricted access (both physically and electronically) data center location that is monitored. Data are stored on redundant servers, each with multiple drive fault redundancy. Data are synced across campus to another UI data center nightly. Physical access is prohibited except for local IT technicians. RDS is used to store raw ROS bag files, sensor data exports, and survey exports.

UI Microsoft SQL (MSSQL) data services are located within the UI RDS data centers and staffed by full-time database administrators (DBA). Data are stored on redundant servers, each with multiple drive fault redundancy.

All servers are registered with the UI Information Security & Policy Office (UI-ISPO) so they can be monitored for potential security problems. In addition, this supplies name and contact information for the systems as well as the services, criticality, and sensitivity of the data. User accounts and access lists are reviewed at least quarterly and kept up to date by deactivating invalid/terminated users and removing unnecessary access rights. All computers are kept up to date with current operating system and application updates/patches by reviewing, testing, and installing them at least monthly. Critical security patches for active exploits are installed within five business days of release. Successful patch installation is verified manually or by using a management service such as System Center Configuration Manager (SCCM) for Windows.

Long- and short-term data storage for this project will comply with the National Transportation Library Guidelines for Evaluating Repositories for Conformance with the DOT Public Access Plan.

4.4.3 Cybersecurity Policies

All systems are located inside UI Information Technology Services' high security firewall context and require explicit firewall rules for any connection outside the high security VLAN.

Privileged access account sessions are encrypted using industry standard "strong" encryption algorithms and key lengths—a minimum of 256-bit symmetric or 2048-bit asymmetric for remote connections, administration tasks, and file transfers containing confidential data. Transparent Data Encryption is enabled (encrypted at-rest), which also automatically encrypts the backup files.

Failed login attempts are audited, including source IP, account name used, failure reason, and database name. Internal IT pages are sent out when a significant number of failed login attempts occur within a five-minute period. Object change (DDL) and server-level modification events are audited. Hourly reports of atypical events on production systems are sent during business hours. Both successful and failed attempts are audited. DBAs always connect to database systems via an RDS "jump server" with DUO enforced for administrative/elevated tasks.

4.4.4 Data Security Policies and Procedures

Data storage security is managed by UI IT personnel in conjunction with UI-ISPO. Data storage employs the "least privilege" principle by granting users only the minimum necessary authorization level to do their work.

System/audit logs are maintained for a minimum of 90 days and are analyzed regularly for security vulnerabilities or intrusions. Regular security/risk evaluations are performed by a third party (e.g., collegiate IT manager, Security Office, internal audit, UI-ISPO) at least annually.

A dedicated system is used for ITAR/Export Control databases, and non-US citizen DBAs are unable to connect to it. UI's TCP has been signed by all members of the SQL DBA team.

4.4.5 Back-up and Recovery Policies and Procedures

The backup schedule is full backups weekly and differential backups nightly. Transactional backups are hourly for systems in FULL recovery model; default backup file retention is 60 days. Backup files reside on NetApp file shares that are replicated across UI data centers (ITF & LC) several times per day.

Databases in FULL recovery model can be restored to any point in time covered by the available backups (down to the second, if required). Databases in a SIMPLE recovery model can be restored to any full or differential backup point in time. Restore requests are handled by the UI SQL DBA team, and non-urgent restore requests typically are completed within one business day.

Raw data (bag files, device exports, raw survey data, etc.) are stored on redundant servers with multiple levels of backup. File snapshots are kept at regular intervals (hourly, daily, monthly). Restore requests are handled by the UI HPC systems administrators group, and non-urgent restore requests typically are completed within one business day.

4.4.6 Data Creation Team Access Restrictions

Member Name	Project Role	Restrictions
Stephen Cable	IT & Data Quality Lead	Administrative full r/w access to all incoming data. PII (participant video) accessed only to ensure data quality.
Cher Carney	Data Analysis & Research Lead	No access to PII video data.
Cheryl Roe	Safety Lead	No access to PII video data.
Greg Wagner	Vehicle Operations Lead	No access to PII.

5 Data Standards

5.1 Data Standards

Dataset Title	Data Standard(s)	Data Standard(s) Digital Object Identifier(s) (DOI[s])	Open or Proprietary?	Data Standard(s) Rationale
Robot Operating System (ROS) Bag File	ROS bag v2.0	http://wiki.ros.org/Bags/Format	Proprietary	ROS bag files are default logging format for storing ROS messages in files; efficient way to record accurately timestamped sensor and vehicle data; widely used in automation and simulation or automated vehicles.
Physiological Sensor Data Export	RFC 4180 (.CSV)	https://www.loc.gov/preservation/digital/formats/fdd/fdd000323.shtml	Open	Sensor exports (from OEM device) are natively in .CSV format, including timestamps.
Rider Survey Data Export	RFC 4180 (.CSV)	https://www.loc.gov/preservation/digital/formats/fdd/fdd000323.shtml	Open	Qualtrics survey software natively exports survey data to .CSV.
Driver/ Rider State Monitoring/ Workload	ISO/IEC 14496-14:2020 (MP4), ISO/IEC 23008-2:2017 (H.265), RFC 4180 (.CSV)	https://www.iso.org/obp/ui/#iso:std:iso-iec:14496:-14:ed-3:v1:en https://www.iso.org/obp/ui/#iso:std:iso-iec:23008:-2:ed-3:v1:en https://www.loc.gov/preservation/digital/formats/fdd/fdd000323.shtml	Proprietary	Realtime H.265 compression implementation using GPU resources requires use of close-source compression technique. Use (decompression) of files available using widespread ISO standards.
ADS Digital Clearinghouse Database	(Custom)		Proprietary	Custom relational database, combines and correlates all dataset titles above into searchable format.

5.2 Versioning

Versioning of ROS datasets will be handled by internal file versioning following major/minor changes in ROS bag formats.

Ingestion scripts (scripts reading ROS bag files and/or external data sources and correlating them into the ADS for Rural America Dataset database) will be managed via version control and will follow semantic versioning practices (https://semver.org/). Major versions will be marked by fundamental changes in the ROS bag structure (such as fundamental Application Programming Interface [API] structure). Minor versions will be marked by minor changes (such as adding or removing topics). Patch versions will be defined by bug fixes.

AV software (authored by UI) versioning will be managed via version control (Git) and follow semantic versioning practices.

5.3 Metadata and Data Dictionary

Dataset Title	Metadata Standards Used	Metadata Discoverable (Y/N)	Data Dictionary Discoverable (Y/N)	Metadata and Data Dictionary Access
ADS Digital Clearinghouse Database	Project Open Data	Y	Y	URL TBD

5.3.1 Metadata Description

The ADS Digital Clearinghouse Database will be stored as a relational database, correlating sensor data from the vehicle (contained within ROS bag files) with external data sources (such as physiological sensor data from riders). The ROS bag file timestamp data serves as the master clock to any individual drive data. Network Time Protocol (NTP) time services will be hosted from the main automation computer node and connected to all external sensors where possible. Metadata from the ADS Digital Clearinghouse Database will be available to registered users via the online ADS Data Access web application containing the Data Dictionary and Playground. The Data Dictionary is a web application serving a browsable/searchable list of data source names/topics, defining sensor output sources from the vehicle, and ingested data from external devices. The Playground is a web application on which verified users will be able to search and view individual vehicle drive data (organized by topics defined by the Data Dictionary) and statistics from individual and/or total drives. It will organize drive data into tables and visualizations of data recorded from all recorded aspects drive data, minus PII. Small sets of targeted drive information will be available for export via CSV.

Remote data reduction will be a process by which clients will upload their custom MATLAB, R, or SASS script—to be run on cloud computing infrastructure. The first step will be requesting a job (using the web front end application or email). A data curator at NADS will consult privacy/export guidelines and, if acceptable, approach the data analysis cloud computing manager (DACCM), who will evaluate the script functionality and estimate the cost to execute the job. The data curator will inform the client of the projected cost. If acceptable, the client will begin working directly with the DACCM to further tailor the script (if necessary) and execute the job. Once complete, the DACCM will work transfer the job output to the client. CPU-based processing will occur using UI's High-Performance Computing (HPC) cluster.

Individual ROS bag files will be provided to professional end-users (research entities, industry, etc.) upon request. As these will include raw, unredacted, identifiable data (including video), users will be required to comply with applicable privacy and/or export laws and a negotiated DUA, which may or may not require IRB approval.

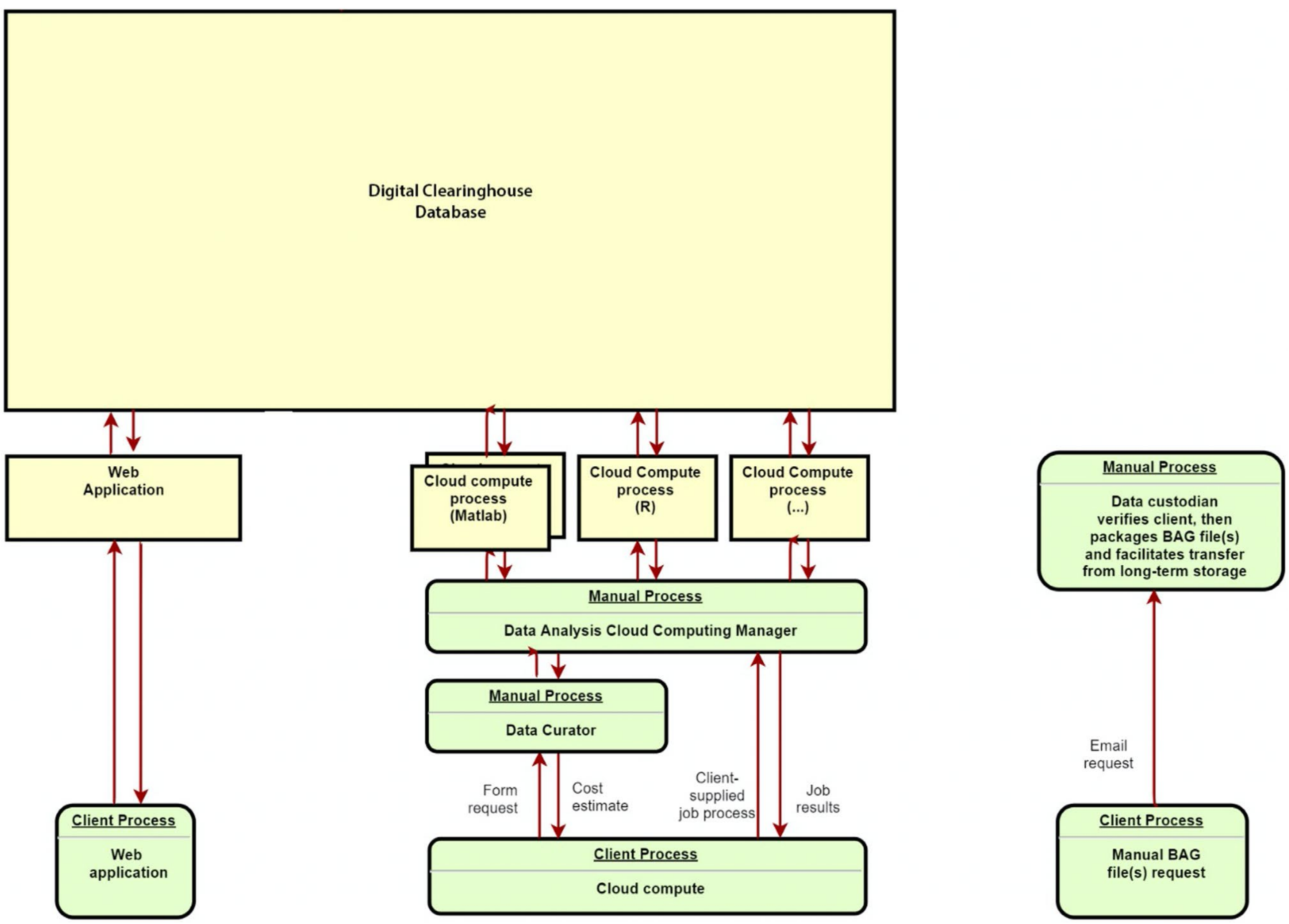

Figure 5-1 *ADS Digital Clearinghouse Database Data Interactions*

5.3.2 Metadata Collection Process

ADS for Rural America will produces a large amount of data collected from the project's on-road vehicle. The on-road vehicle is a custom-built, drive-by-wire, Ford Transit with a suite of sensors, hardware, and automation software.

The vehicle uses the open source Robotic Operating System (ROS) as part of its software. Data collected from the vehicle is organized into ROS bag files. Bag files are well-documented binary recordings, used at length in the ROS and greater automation community. The bag file format contains message data from ROS with the range of variables that UI expects to collect as part of this project. At the end of each project drive, these bag files will be stored in their raw form. At the same time, these bag files will be processed to generate a set of indexed, synchronized point-in-time data. The bag file data includes local vehicle information, sensor data, video, LIDAR, and ingested external data (V2V/X, third party APIs, etc.). By collecting this data, UI will generate an open data pool to further vehicle automation. Our data will not only provide detailed information on where UI was successfully able to use automation but also on where we weren't able to implement automation. The bag files will show which sensors and sources of data were available for the automation along each part of the drive and where they weren't available (i.e., automation system disengagement) by variables dedicated to sensor health. This will enable us to better understand the reasons for automation disengagements and link them to specific roadway conditions, weather conditions, and/or connectivity issues. All of these points of data will be recorded in real-time on the vehicle by internal ROS capability (bag file recording).

External data sources (such as wearable devices or other sensors, questionnaires, etc.) will collect timestamped datasets and participant-correlated user experience feedback.

As the vehicle drives, all vehicle data points (sensor data, video channels, external API data) would be written in real time to a local Quantum solid state storage array using a solid state disk (SSD) cartridge. The cartridge securely slots into a ruggedized chassis and allows for the throughput necessary for the high volume/high velocity data streams.

Once the vehicle docks back at NADS, the SSD cartridge can be removed from the vehicle's chassis and placed in a second chassis connected to a workstation located in the vehicle garage. A user-initiated process checks the volume automatically for new files. New files are assigned a checksum "digital thumbprint" (tamper-proofing) then transferred via a high speed wired Ethernet link to a University of Iowa cloud services, leveraging UI Research Services Large Scale Storage (LSS) cloud storage. LSS storage is mirrored between two physically secure and geographically separated data centers (operated by the

University of Iowa), protected against local disk failure by appropriate RAID application, as well as scalable for the expected large volumes of data.

Once the data is transferred, the SSD cartridge can be swapped back into the vehicle, making it ready for another round of data collection.

An automated process (running in the UI's cloud architecture) would digest each individual bag file, converting the binary file into correlated, indexed data to be included in the digital clearinghouse database. Video frame data (and other non-conformant data) would be written to disk and linked symbolically. External data source exports will be digested in the same manner, contextually placing their data inline with the vehicle data.

Additionally, video frame data will be streamed to a secondary automated process to calculate passenger posture vectors. GPU-based processing will occur using the University of Iowa's High-Performance Computing (HPC) Argon cluster. The Argon cluster consists of 21 machines with Nvidia P100 accelerators, 2 machines with Nvidia K80 accelerators, 11 machines with NVidia K20 accelerators, 2 machines with Nvidia P40 accelerators, 13 machines with 1080Ti accelerators, and 18 machines with Titan V accelerators. Most of those in the ITGF data center are connected with the InfiniPath fabric.

Per each project drive, these bag files (and external data source files/exports) will be stored in their raw form. At the same time, bag files and external data source files/exports will be processed to generate a set of indexed, synchronized point-in-time data which will then be stored into a comprehensive relational database in the cloud (ADS Digital Clearinghouse Database).

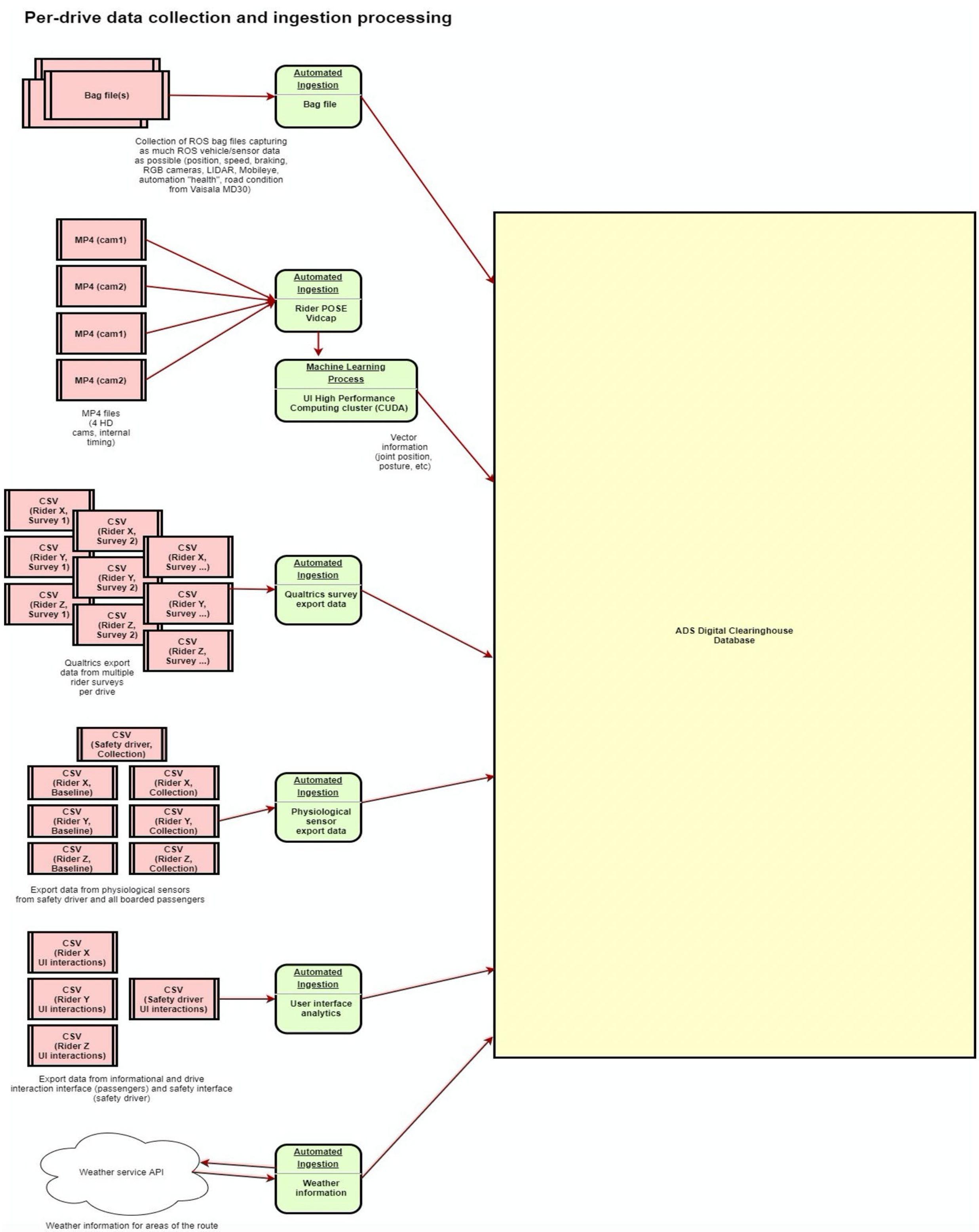

Figure 5-2 *ADS Digital Clearinghouse data ingestion processing*

Data will be recorded for 10 drives per phase of the autonomous vehicle across the ADS for Rural America period of performance (2.5 years), with thought given to trying to vary time-of-day, season, driving conditions, etc. in order to provide greater insight into ADS technologies.

5.3.3 Metadata Collection Examples

General Data To Be Collected

- Unique data frame identifier (unique across entire dataset)
- Current time of day
- Current weather
- Relevant vehicle CAN bus (e.g., head lights, windshield wipers, blind spot monitor, adaptive cruise control, lane keeping assist, anti-lock braking system, electronic stability control, odometer)
- Road surface state
- Physiologic sensors attached to safety driver and riders; measuring heart rate, electrodermal activity (EDA), etc. (devices and technical specifications listed in Section 2.2.2, Supplemental Sensors)
- Forward facing camera video
- LIDAR data

Automation and Human Safety Driver Data To Be Collected

ADSs, for the foreseeable future, will require close oversight by human drivers for their safe operation. Therefore, UI will collect data that will provide insight on the automation, the vehicle's operational design domain (ODD), and the human safety driver. All raw and video data (where appropriate DUAs are accepted) will be made available for post-study analysis. UI will also process data (in real-time and post-drive) to provide additional information about driver state. Some of processing will use publicly available algorithms and other processing will use algorithms developed by the UI.

Today, many vehicle developers use the primary metrics of miles accumulated on public roads and number of times a safety driver has to intervene. Although these metrics are important, they provide a limited view of ADS testing since they do not offer a way to measure the diversity of situations encountered. No single measure at any stage or setting can tell the entire story. "Roadmanship" is a new but not entirely defined integrated leading measure of driving abilities that goes beyond these basic safety metrics to suggest that an ADS behaves similar to human drivers. UI's plan is to collect data that provides building blocks to helping define and build roadmanship metrics. UI will collect data on the operation of the automated vehicle under challenging conditions, which will demonstrate and provide data on the vehicle's ODD. Data collected regarding the automation and the human safety driver will include:

- Automation status: current automation level
- Takeover request
- Automation system re-engaged in driving
- Driver engaged in driving
- Driver control inputs:
  - Steering wheel position
  - Brake pedal position
  - Accelerator pedal position
  - Current gear/transmission status
  - Turn signal status
  - State of various switches on dashboard
- Safety driver state:
  - Vital signs – heartbeat, respiration rate
  - Steering wheel reversals
  - Hand position
  - Arm/joint location/posture
  - Video of driver's face
  - Video of driver's hand on steering wheel
  - Additional driver state information such as:
    - Eye blink, glance, fixation
    - Gaze direction and angles
    - Percentage eye openness tracking (PERCLOS)

Examples of data to be collected regarding geographic location:

- Vehicle latitudinal/longitudinal coordinates, heading and odometer from GPS
- Vehicle lateral position within lane from camera-based sensor
- Desired path as defined by waypoints from an HD map/HD map centerline data (where available)
- Distance to left/right lane marking and curb
- Distance to signs and signals
- Current speed limit based on signage
- Current speed limit based on HD map (where available)
- Desired traveling speed

- Distance to lead vehicle
- Type of lane marking on the left/right

Examples of data to be collected regarding roadway type:

- Type of roadway (interstate, highway, local, etc.)
- Pavement type (paved, gravel)
- Surface coefficient of friction (estimated)
- Left and right lane marker data:
  - Quality (visual quality of marker)
  - Marker type (solid, dashed, etc.)
  - Curve model kind (order of equation)
  - Offset (lateral distance from the sensor to marker)
  - Heading angle (angle of marker relative to sensor)
  - Curvature (curvature of lane marker at camera)
- In construction/work zone (if applicable):
  - Curvature derivative (amount of curvature changes as one moves away from camera)
  - Marker width (width of painted marker)
  - View range (physical view range of marker)

Examples of data to be collected regarding other vehicles around demo vehicle:

- Location, heading, and speed of demonstration vehicle
- Connected vehicle information:
  - Vehicle ID/type
  - Vehicle speed
  - Vehicle heading
- Distance to conflict vehicle
- Time to arrival to conflict vehicle
- Speed of other vehicles on or approaching roadway

Examples of data to be collected regarding environmental conditions:

- Exterior video showing lighting condition
- Current weather
  - Clear/rain/snow/fog
  - Wind velocity
  - Visibility

  - Grip
  - Surface state
    - Dry
    - Moist
    - Wet
    - Frost
    - Snow
    - Ice
    - Slushy
    - Streaming water
    - Slippery
    - Ice patch
- Surface layer thickness
- Surface temperature
- Air temperature
- Dew point and frost point temperature
- Relative humidity
- Sensor status

Examples of LIDAR data to be collected:

- Start and end angle of scan
- Angular distance between measurements
- Range data
- Intensity data (when available)
- Point cloud

Examples of participant survey data to be collected:

- Pre-drive and post-drive survey responses
  - Trust/acceptance
  - Perceptions of safety/usefulness
  - Anxiety
  - Intention to use
- Experience with and attitude toward technology
- External locus of control
- Targeted demographics

5.3.4 Examples of Potential Research Questions to be Addressed using Metadata

The goal of this project is to collect data that can be useful for researchers and practitioners from government, industry, and academic sectors in answer to several research questions. Following are sample questions based on discussions with colleagues around the US for which the data could provide insight:

- What is the impact of lighting/weather conditions on vehicle sensors and the resulting automation?
- Does the quality or type of lane marking affect the vehicle's ability to maintain lane position? How does that change as lane markings degrade over time?
- What affect does the safety driver state have on takeover timing?
- How does the information displayed to the safety driver affect the timing or quality of takeovers?
- What type of environmental conditions affect the vehicle's ability to connect to infrastructure (e.g., Signal Phase and Timing [SPaT])?
- How is the vehicle's ability to identify surrounding vehicles affected by speed?
- How does automation handle different types of construction zones and associated markings or speed changes?
- What type of data can be obtained by the sensors when driving on gravel roads?
- How do rider experiences in the vehicle affect their ratings of trust/acceptance?
- Do anxiety levels of riders decrease with exposure?

6 AV Maintenance

6.1 Sensor Calibration

RVIZ (a visualization package integrated into ROS) will validate each sensor working independently. In general, most sensors simply output data, so calibration either is done internally by the sensor or is not required. However, the general sensor array will have some "calibration," as it will measure and input the lever arm information for each sensor into a reference file on the computer. If the sensors are not moved, checking and editing this file should not be needed regularly.

6.2 Software Updates

AV software updates (tuning/improvements) will be performed by AutonomouStuff for each data collection during Phase 1 through 8. Updates typically will be documented using version control with each release that would contain notes regarding the release. "Data Review" is a UI task that will involve reviewing the data (and feedback from the safety driver) for how well the vehicle fulfilled the goals of each phase from an autonomy performance point of view.

GLOSSARY

ADS	Automated Driving System
API	Application Programming Interface
AV	Automated vehicle
CAN	Controller Area Network (electronic vehicle bus standard)
CAPTCHA	Completely Automated Public Turing test to tell Computers and Humans Apart (https://en.wikipedia.org/wiki/CAPTCHA)
DACCM	Data Analysis Cloud Computing Manager
DUA	Data Usage Agreement
EDA	Electrodermal Activity
FDP	Federal Demonstration Partnership
ITF	University of Iowa Data Center Site 1
IRB	Institutional Review Board
LC	University of Iowa Data Center Site 2
LIDAR	Laser Imaging, Detection and Ranging Sensor
NADS	National Advanced Driving Simulator
NTP	Network Time Protocol
ODD	Operational Design Domain
PERCLOS	Percentage eye openness; a drowsiness detection measure, referred to as the percentage of eyelid closure over the pupil over time. (https://en.wikipedia.org/wiki/Fatigue_detection_software#Percentage_eye_openness_tracking_(PERCLOS))
PII	Personal Identifiable Information
ROS	Robot Operating System
SCCM	Microsoft System Center Configuration Manager
SPaT	Signal Phase and Timing
SQL	Structured Query Language
TTC	Time to Collision
UI	University of Iowa
UI-ISPO	University of Iowa Information Security & Policy Office

U.S. Department of Transportation

Federal Transit Administration

U.S. Department of Transportation
Federal Transit Administration
East Building
1200 New Jersey Avenue, SE
Washington, DC 20590
https://www.transit.dot.gov/about/research-innovation