TTI: 0-7128

Texas A&M Transportation Institute

Digitizing Traffic Control Infrastructure for Autonomous Vehicles (AV): Technical Report

Technical Report 0-7128-R1

Cooperative Research Program

TEXAS A&M TRANSPORTATION INSTITUTE
COLLEGE STATION, TEXAS

sponsored by the
Federal Highway Administration and the
Texas Department of Transportation
https://tti.tamu.edu/documents/0-7128-R1.pdf

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
FHWA/TX-23/0-7128-R1		

4. Title and Subtitle	5. Report Date
DIGITIZING TRAFFIC CONTROL INFRASTRUCTURE FOR AUTONOMOUS VEHICLES (AV): TECHNICAL REPORT	Published: January 2024
	6. Performing Organization Code
7. Author(s)	8. Performing Organization Report No.
Jason (Dayong) Wu, Matthew Miller, Gretchen Stoeltje, Minh Le, William Hwang, Tianchen Huang, Nanzhou Hu, Rym Zalila-Wenkstern, Behnam Torabi, and Xiao Li	Report 0-7128-R1
9. Performing Organization Name and Address	10. Work Unit No. (TRAIS)
Texas A&M Transportation Institute The Texas A&M University System College Station, Texas 77843-3135	11. Contract or Grant No. Project 0-7128
12. Sponsoring Agency Name and Address	13. Type of Report and Period Covered
Texas Department of Transportation Research and Technology Implementation Office 125 E. 11th Street Austin, Texas 78701-2483	Technical Report: September 2021–August 2023 14. Sponsoring Agency Code

15. Supplementary Notes

Project sponsored by the Texas Department of Transportation and the Federal Highway Administration.
Project Title: Digitizing Traffic Control Infrastructure for Autonomous Vehicles (AV)
URL: https://tti.tamu.edu/documents/0-7128-R1.pdf

16. Abstract

High precision road maps are a crucial component to facilitating autonomous driving techniques. Although current Autonomous vehicles (AVs) rely on vehicular sensing techniques (e.g., camera, light detection and ranging [LiDAR], radar), studies have suggested that creating high-quality road maps with traffic control infrastructures (TCIs) (e.g., traffic signs, signals, intersections) precisely digitized is necessary to enhance safe-driving operations of AVs. Meanwhile, digitizing TCIs is also of great importance for road assets planning and management. However, a readily available database with precisely digitized TCIs is still missing in most areas. Traditionally, TCIs are manually digitized by conducting field studies, which is time consuming and labor intensive. With the advancement of data collection and processing techniques, numerous emerging data sources are becoming available, posing great potential to capture and digitize TCIs more efficiently. In this project, the researchers developed an effective framework for the digitization, maintenance, and sharing of roadway assets, especially for TCIs. The research team evaluated available solutions (commercial, open-source, and public), investigated potential legal issues, and proposed new approaches by leveraging emerging data sources and techniques. Simulations based on various real-world scenarios were developed to evaluate the benefits of incorporating TCI digitized data in enhancing the safety and operational performance of AVs.

17. Key Words	18. Distribution Statement
Autonomous Vehicles, Traffic Control Infrastructure, High-Definition Map, Digitization, Simulation, Web GIS	No restrictions. This document is available to the public through NTIS: National Technical Information Service Alexandria, Virginia https://www.ntis.gov

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	196	

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

DIGITIZING TRAFFIC CONTROL INFRASTRUCTURE FOR AUTONOMOUS VEHICLES (AV): TECHNICAL REPORT

by

Jason (Dayong) Wu
Associate Research Scientist
Texas A&M Transportation Institute

Matthew Miller
Associate Research Scientist
Texas A&M Transportation Institute

Gretchen Stoeltje
Associate Research Scientist
Texas A&M Transportation Institute

Minh Le
Research Engineer
Texas A&M Transportation Institute

William Hwang
Research Scientist
Texas A&M Transportation Institute

Tianchen Huang
Graduate Student Worker
Texas A&M Transportation Institute

Nanzhou Hu
Graduate Student Worker
Texas A&M Transportation Institute

Rym Zalila-Wenkstern
Professor
University of Texas at Dallas

Behnam Torabi
Postdoctoral Researcher
University of Texas at Dallas

and

Xiao Li
Senior Research Associate
University of Oxford

Report 0-7128-R1
Project 0-7128
Project Title: Digitizing Traffic Control Infrastructure for Autonomous Vehicles (AV)

Sponsored by the
Texas Department of Transportation
and the
Federal Highway Administration

Published: January 2024

TEXAS A&M TRANSPORTATION INSTITUTE
College Station, Texas 77843-3135

DISCLAIMER

This research was sponsored by the Texas Department of Transportation (TxDOT) and the Federal Highway Administration (FHWA). The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the data presented herein. The contents do not necessarily reflect the official view or policies of FHWA or TxDOT. This report does not constitute a standard, specification, or regulation.

ACKNOWLEDGMENTS

This project received sponsorship from TxDOT and FHWA. The authors would like to express their gratitude to Joanne Steele, the TxDOT project manager, and extend their appreciation to the members of the Project Monitoring Committee: Jianming Ma, Ash Duong, Bethany Wyatt, Charles Tapp, James Kuhr, and Matthew McCarty, for their invaluable assistance and guidance throughout the project's duration. Special thanks are also extended to the data providers, Mobileye, Blncsy, Nexar, and Pathway, for generously providing their sample TCI datasets, which played a pivotal role in the successful completion of this project. Additionally, the authors wish to acknowledge the city of Arlington for their willingness to collaborate and share their asset dataset with the research team, facilitating the evaluation and validation process.

TABLE OF CONTENTS

LIST OF FIGURES

Page

LIST OF TABLES

Page

CHAPTER 1. INTRODUCTION

1.1 BACKGROUND

High precision road maps are a crucial component to facilitating autonomous driving techniques. Autonomous vehicles (AVs) are experiencing exponential growth. According to the latest forecast from IHS Markit, over 33 million AVs will be on the road globally by 2040, posing a higher requirement to ensure AVs' driving safety. Although current AVs rely on vehicular sensing techniques (e.g., camera, light detection and ranging [LiDAR], radar), studies have suggested that creating high-quality road maps with traffic control infrastructures (TCIs) (e.g., traffic signs, signals, intersections) precisely digitized is necessary to enhance safe-driving operations of AVs. Meanwhile, digitizing TCIs is also of great importance for road assets planning and management. However, a readily available database with precisely digitized TCIs is still missing in most areas. Traditionally, TCIs are manually digitized by conducting field studies, which is time consuming and labor intensive. With the advancement of data collection and processing techniques, numerous emerging data sources are becoming available, posing great potential to capture and digitize TCIs more efficiently.

In this project, the researchers developed an effective framework for the digitization, maintenance, and sharing of roadway assets, especially for TCIs. The research team evaluated available solutions (commercial, open-source, and public), investigated potential legal issues, and proposed new approaches by leveraging emerging data sources and techniques. Simulations based on various real-world scenarios were developed to evaluate the benefits of incorporating TCI digitized data in enhancing the safety and operational measures of AVs.

1.2 PROJECT GOAL AND RESEARCH TASKS

The research team outlined four goals for this study, which are summarized as follows:

1. Explore existing solutions, potential resources/datasets, and possible legal issues for TCI digitization and inventory development.
2. Understand the opinions from transportation agencies and AV companies on TCI data collection and utilization.
3. Propose a web GIS-based framework for TCI digitizing and sharing and perform a comprehensive evaluation of the framework and TCI datasets.
4. Develop simulations based on various real-world scenarios and evaluate the benefits of incorporating TCI digitized data in enhancing the safety and operational measures of AVs.

In order to achieve the project goals, the research team conducted four major tasks, summarized as follows:

- Task 2: The research team systematically reviewed the existing literature on TCI identification and digitization (e.g., intersection, traffic signs, traffic signals) and the interaction between traffic infrastructure and AVs.
- Task 3: The research team investigated the state of practice for TCI digitization and identified potential market available data sources for building TCI inventories. The team also identified the possible legal issues associated with the digitization, collection, and sharing of TCI data.
- Task 4: The research team obtained and evaluated sample digitized TCI datasets from multiple third-party data providers. A framework for digitizing and sharing these datasets was developed in this task.
- Task 5: The research team assessed the precision of the TCI datasets by comparing them with the ground truth data, constructed a web GIS-based platform specifically designed for the digitization of TCIs, simulated multiple real-world scenarios, and analyzed the advantages of integrating TCI digitized data to improve the safety measures of AVs.

1.3 REPORT ORGANIZATION

The remaining chapters of this report include the following:

- Chapter 2: Literature Review—literature review methodology, materials, results, and discussions.
- Chapter 3: State of Current Practices—existing practices, solutions, data sources, and potential legal issues.
- Chapter 4: TCI Digitization Framework—description of the conceptual TCI digitization framework.
- Chapter 5: Comprehensive Evaluation and AV Simulations—evaluation results on TCI digitization framework, TCI datasets, AV simulations, and the web GIS-based platform.
- Chapter 6: Conclusions and Recommendations—key findings and recommendations.
- Appendices—literature review materials, survey questions and results, data management plan, etc.

CHAPTER 2. LITERATURE REVIEW

2.1 SUMMARY

In Task 2, the research team conducted a systematic literature review on transportation asset management, especially for TCI digitizing and sharing.

2.1.1 Objective

The objective of this task is four-fold:

1. Explore what types of TCI had been digitized by previous research.
2. Summarize available data sources used for TCI digitization.
3. Examine the models and methods for TCI digitization.
4. Analyze the limitations in current progress and propose future research directions.

2.1.2 Overview of Chapter 2

This chapter presents the results of Task 2, where the research team systematically reviewed the existing literature on TCI identification and digitization (e.g., intersection, traffic signs, traffic signals) and the interaction between traffic infrastructure and AVs.

The rest of this chapter is organized as follows:

- Section 2.2 presents the introduction and background of this literature review.
- Section 2.3 introduces the review methods, processes, and materials.
- Section 2.4 shows the findings of data acquisition and processing methods in TCI digitization from the systematic literature review (SLR) and summarizes the modules and solutions for the inventory upkeep for inventory management identified through the narrative review (NR).
- Section 2.5 summarizes the key findings and further discusses the limitations and future research in TCI inventory management.
- References and Appendix A are provided at the end.

2.2 INTRODUCTION

Transportation Asset Management (TAM) has been widely acknowledged as an essential component of transportation planning and management [1]–[3]. As a data-driven decision support solution, the TAM system enables transportation agencies to make decisions regarding investment, maintenance, and replacement of roadway assets, which has significant implications for providing safe and efficient transportation services to all road users [4], [5]. In 2021, the Infrastructure Investment and Jobs Act (IIJA), with a historic $1.2 trillion infrastructure bill, was signed into law in the United States. The IIJA aims to provide significant funding for building resilient, safe, sustainable, equitable transportation infrastructures with a clear need for road asset digitization and inventorying initiatives, especially through adopting innovative technology (such as remote sensing, the internet of things, and artificial intelligence) for asset management [6]. Effectively establishing and managing an inventory database for roadway assets (such as asset types, locations, conditions, and dimensions) play an essential role in TAM [7].

Traditionally, road assets are manually digitized by conducting field studies, which is time-consuming and labor-intensive [8]. With the advancement of data collection and processing techniques, road asset digitization and management have become a multidisciplinary process involving the integration of knowledge and skills from multiple fields, such as geospatial data science, computer vision, transportation management. For example, geospatial data science is of growing importance since numerous emerging data sources have become available, such as street view images, aerial terrestrial/airborne light detection and ranging (LiDAR), aerial/satellite imagery, and crowdsourced asset data. These data sources are offering new avenues and are increasingly implemented to capture and digitize road assets[9], [10]. More importantly, by combining these datasets with computational advances, such as computer vision techniques, a variety of road assets can be recognized from different imagery and non-imagery data sources. It is also worth noting that road asset digitization and management are often driven by transportation planning and maintenance objectives, such as improving road safety, reducing congestion, and enhancing transportation efficiency. This requires an understanding of transportation management principles in order to produce actionable regulatory frameworks for the road asset inventory build-up and upkeep. Therefore, conducting a comprehensive literature review from an interdisciplinary point of view is greatly needed to integrate solutions and advances from relevant fields to guide the establishment and maintenance of road asset inventory more efficiently.

As an important component of road assets, traffic control infrastructures (TCIs), such as traffic signs, traffic signals, and pavement markings, play a vital role in managing traffic flows and improving road safety. These TCIs provide information about the current state of the road, restrictions, prohibitions, warnings, and other helpful information for driving guidance [11]. Building and maintaining a TCI inventory not only can benefit transportation management, but more importantly, the digitized and timely updated TCI data could lay a solid foundation for promoting the deployment of connected and autonomous vehicles as well as the establishment of intelligent transportation systems [10], [12]. However, compared to other road assets (e.g., pavement and bridges), the inventory establishment and maintenance for TCIs through mining emerging data are less explored. Therefore, there is a great need to review and summarize the current progress and challenges of using different geospatial data sources along with computer vision and data management techniques for TCI inventory buildup and upkeep.

To fill in this gap, this study used a mixed review approach by combining the systematic literature review (SLR) and narrative review (NA) methods to summarize the current state of practices, solutions, and challenges for TCI digitization and inventory development. We aim to provide a better understanding of how recent advances in spatial data science, computational science, and transportation management offer new opportunities for digitizing and inventorying TCIs and to help guide future research in this area. Please note that TCI includes a wide range of traffic assets, such as traffic signals/lights, traffic signs, pavement markings, lane markings, and intersections, among others. In light of the availability of existing studies, this paper primarily focuses on inventorying three main types of TCIs: traffic signals, traffic signs, and pavement markings.

The general workflow of TCI inventory establishment and management can be summarized into three modules: data acquisition, asset extraction, and inventory establishment and management [8], [14], as illustrated in Figure 1. Data acquisition aims to capture road assets using different data collection techniques. Asset extraction is the process of digitizing and localizing target objects from the collected imagery or non-imagery data. Inventory management is to establish and maintain a useful inventory tracking TCIs' information.

Guided by this general workflow, this review aims to address three specific research questions regarding TCI digitization and inventory development, including:

1. 1. Which data types have been utilized for TCI digitization?
2. 2. What methodologies have been employed for TCI digitization?
3. 3. What procedures have been implemented for the establishment, maintenance, and updating of TCI inventories?

This literature offers an interdisciplinary view of TCI inventory establishment by combining advances in multiple fields, such as data acquisition, computer vision, database management, and transportation management, among others. This study primarily contributes to transportation decision-makers and practitioners by providing a guideline on TCI inventory creation and maintenance. This study can also act as a valued reference helping data scientists to understand the landscape of available data sources and computer vision techniques (e.g., image classification, segmentation, object detection) and serves as a starting point to facilitate the digitization and inventorying of other road and urban assets.

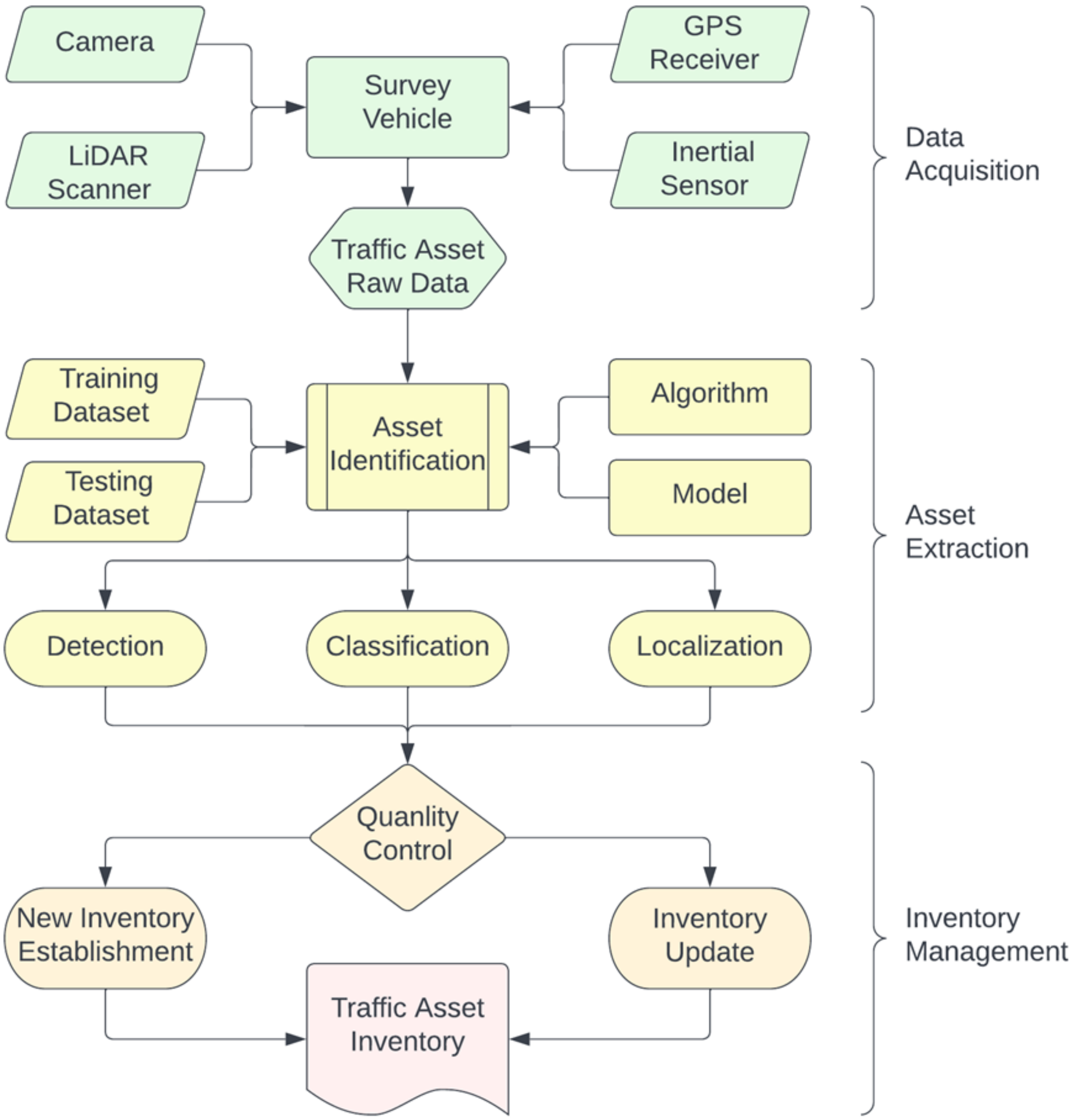


Figure 1. Conceptual Workflow for TCI Inventory Establishment and Management.

2.3 REVIEW METHODOLOGY

2.3.1 Review Method—Mixed Review Procedure

This study aims to achieve a comprehensive understanding of existing solutions supporting the TCI inventory build-up, summarized into three modules: data acquisition, asset extraction, and inventory establishment and management, as illustrated in Figure 1. We first conducted an SLR because of its transparent and rigorous review procedure and its well-recognized capacity to generate a comprehensive understanding of a given topic. However, most identified literature through the systemic literature search is related to data acquisition and road asset extraction, while only a few are inventory relevant. We noticed that many inventory establishment and management works were published as grey literature, such as research reports, working papers, or technique manuals, which are generally not indexed by mainstream research article databases. To complete the understanding of the inventory build-up, we then contacted an NR to summarize the methods used in inventory establishment and management from supplementary literature. By combining an SLR with an NR, this mixed review procedure can result in a more complete and nuanced understanding of the complete workflow of TCI inventory establishment and upkeep so as to conclude a framework for TCI inventory management.

SLR has been widely applied to build new frameworks and perspectives on a topic based on comprehensive understanding by reviewing, critiquing, and synthesizing representative literature on that topic [1]. Note that the specific SLR procedures could vary depending on the type of literature review, but in general, the whole literature review process could be accomplished through the following eight steps: (1) formulating the research problem; (2) developing and validating the review protocol; (3) searching the literature; (4) screening for inclusion; (5) assessing quality; (6) extracting data; (7) analyzing and synthesizing data; (8) reporting the findings [2]. In this study, the SLR is primarily used to summarize the data sources, methods, and models utilized to digitize TCIs.

Compared to SLR, NR was commonly used for obtaining the perspective on a topic of interest that usually does not involve a stated hypothesis. Unlike SLR which starts from a long list of all relevant articles, NR primarily search for pivotal papers known to the researcher[3], [4]. In this research, we performed a narrative search to review the inventory-related articles and summarize a framework for inventory management based on the review result.

2.3.2 Literature Search

2.3.2.1 Databases

This study performed a literature search on four research literature databases: IEEE Xplore (IEEE), Web of Science (WoS), Transport Research International Documentation (TRID), and Google Scholar. WoS and Google Scholar are among the most comprehensive academic literature databases[5]. Google Scholar is a free search engine that indexes academic publications across a wide range of publishing formats and disciplines. WoS is a paid-access platform with articles manually checked based on their defined scholarly and quality criteria, which has higher academic reliability than Google Scholar. IEEE and TRID focus more on specific themes: IEEE is a research database majorly for journal articles, conference proceedings, and other scholarly materials about computer science, electrical engineering, and electronics; TRID is more expertized in the literature on transportation studies. Since our review scope covers multidisciplinary components from data science, computer science, and transportation, IEEE and TRID fit the review scope well, which can help build a comprehensive collection of relevant publications.

2.3.2.2 Screening and Eligibility Check

The SLR establishes a literature inventory through keyword searching, which inevitably includes irrelevant or redundant papers. Therefore, a screening and eligibility check is needed to manually select the relevant literature aligning with the research topic, questions, and objectives. In this study, the screening and eligibility check was completed by two independent reviewers through three steps: title screening, abstract screening, and full-text screening. Some articles are accessible from multiple databases, resulting in duplications in the selected papers. The duplicated articles were manually removed through the title screening step.

To select the most relevant literature, the research team established two sets of selection criteria, as listed in Table 1. As defined in the inclusion criteria, the selected studies need to be either relevant to the digitization of TCIs or the establishment and management of road asset inventories. Meanwhile, the team proposed four exclusion rules for filtering articles that focus on specific topics, such as TCI detection or recognition under extreme conditions or only talking about algorithms for image-based asset recognition. Meanwhile, this review also excluded articles published before 2002 or written in non-English languages.

Table 1. Literature Selection Criteria.

Type	Criteria
Inclusion 1	The article is relevant to the establishment, maintenance, or update of the inventory database of TCIs.
Inclusion 2	The article is relevant to the detection, recognition, or positioning of TCIs.
Exclusion 1	The article focuses on the methods only under extreme environmental conditions.
Exclusion 2	The article is specifically about the establishment of algorithms.
Exclusion 3	The article is dated (before 2002), or the methods discussed in the article are too outdated (hardly appear in other articles after 2002).
Exclusion 4	The article is not written in English.

2.4 REVIEW RESULTS

This section presents the review results for both the SLR and NR. This section first presents the SLR result documenting the existing data sources, methods, and models utilized to digitize TCIs. Then, the section summarizes a framework for TCI inventory management based on the findings from the NR.

2.4.1 SLR Literature Search Results on TCI Digitization

Figure 2 illustrates the whole process of searching, screening, and eligibility check performed in the SLR. The literature search was accomplished in October 2021. The research team started with 6,079 articles, including 1,036 from IEEE, 2,137 from WoS, 2,300 from TRID, and

606 from Google Scholar. Through title screening, 5,289 articles were identified as irrelevant or duplicated and were thereby excluded. For the remaining 790 articles being passed to the eligibility check, 612 were identified as over-specific or too narrowed and thus were excluded as well. Finally, through an eligibility check based on full-text reading, 109 articles were identified as topic-relevant, technique-applicable, and up to date to our research.

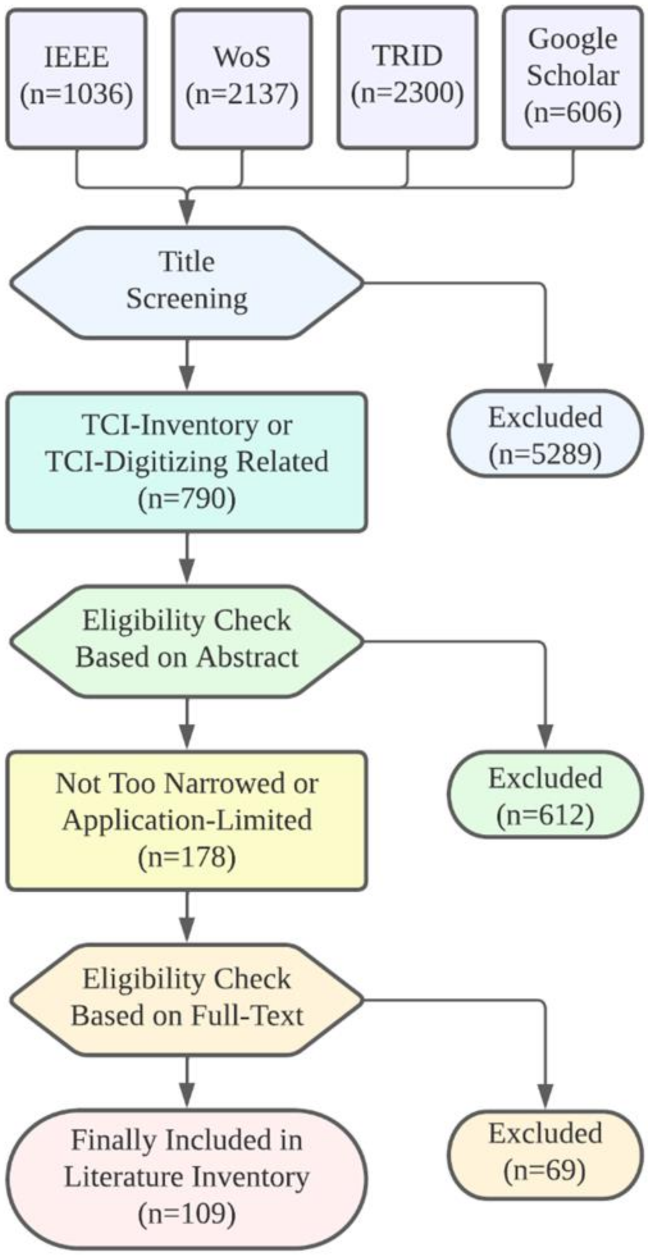


Figure 2. Workflow of Literature Screening and Eligibility Check.

2.4.2 SLR Result of Data Sources for TCI Digitization

According to the review result, two types of data were majorly captured for inventorying TCIs, including TCI information and TCI location.

TCI information is majorly derived from two kinds of raw data: imagery (including street view images and videos) and LiDAR point cloud [6]–[8]. Streetscape images provide 2D visual information for the road assets, which can be processed through computer vision algorithms to recognize and extract TCIs. However, they provide limited information about the depth and can have occlusions and shadows [9], [10]. LiDAR point clouds can provide detailed 3D information about a scene. Compared to streetscape images, LiDAR works better in night-time or low-light conditions for identifying TCIs with specific geometric shapes and characteristics. However, the data density depends on the sensor's resolution and can have missing or noisy data issues [6], [11]. Both types of data are usually captured using vehicle-based collectors (camera/LiDAR scanner). While vehicles run along the roads in the targeted area, the equipped camera captures images at a certain frequency or records videos or the LiDAR scanner scans points around the vehicles. Some researchers run the survey vehicles at different times of the day (e.g., morning, noon, dusk, and night) under various weather conditions (e.g., clear, rainy, froggy, etc.) to collect sufficient data that promote the performance of TCIs detection and classification (Feng et al., 2019). In addition, some survey vehicles are equipped with devices for in-vehicle data preprocessing, such as quality checks and data fusion, which can enhance the efficiency of developing TCI inventories [12], [13].

For the TCI location, GPS is the primary source for obtaining the location data. Some studies also implemented inertial measurement to enhance the location accuracy. The positioning equipment like GPS receivers and inertial measurement units (IMUs) are typically installed on survey vehicles [12]. Please note that the collected location data refers to the location of the survey vehicle, not TCIs; thus, a series of post-process is necessary to calculate the physical location of the targeted TCI [14]–[16].

2.4.3 SLR Result for TCI Extraction Methods

The general process for extracting TCIs from the collected data can be divided into three modules: detection, classification, and localization. TCI detection aims to identify the potential road sign regions from the images (a.k.a. regions of interests [ROIs]) and independently retrieve the pixel locations of all present road signs in each collected image. The classification module is performed to recognize the type of each detected TCI. The localization module retrieves/estimates the real-world positions of the detected signs. Some advanced methods that combine the detection and classification tasks can be implemented (see Section 2.4.3.3).

2.4.3.1 Image-based Methods for TCI Detection

Imagery is the most used data source for TCI detection. Imagery-based TCI detection methods can be classified into two categories based on their principles: color-based and shape-based.

Color-based detection relies on the fact that TCIs, such as traffic signs, usually have distinctive colors, which can be used to differentiate them from the surrounding environment. Many color-based segregation methods have been implemented based on the different color spaces to identify the ROIs within the input image [17]–[19]. The most used color spaces include Red Green Blue(RGB), Hue Saturation Value (HSV), Luma Chroma (YUV), among others. The commonly used color-based algorithms include color thresholding [20], [21], histogram-based thresholding [20], [21], region growing [20], [21], and machine learning algorithms using color features [20], [21], which are detailed in Table 2.

Shape-based methods rely on the distinctive geometric shape of TCIs (e.g., circle, triangle, rectangle, octagon) rather than their colors. Compared to color-based methods, shape-based methods are more advanced at tolerating environmental influence (e.g., weather conditions and times of day). Table 2 lists commonly used algorithms, which include template matching [22], [23], hough transform [14], [35], contour detection [37], distance transform matching [28], [29], and Harr feature-based cascade classifier [30].Note that the performances of the shape-based methods may not be satisfactory when the targeted objects are damaged or obscured.

To achieve a robust detection result, some studies combined color-based and shape-based methods where the candidate ROIs are extracted based on the color threshold and are filtered by shape template match [18], [31], [32]. In recent years, some learning-based methods have been implemented to enhance ROI extraction [12], [33].

Table 2. Image-based Traffic Sign Detection Methods.

Category	**Typical algorithms or models**	**Reference**
Category Color-based methods	Typical algorithms or models	Reference
	Color thresholding: It is a common technique for segmenting an image based on its colors. It processes the image by setting a threshold for the colors and then segment the image based on the pixels that meet that threshold.	[30]
	Histogram-based thresholding: It is a technique for segmenting an image into foreground and background regions based on pixel intensity values. It computes the image's color histogram and then sets a threshold based on the histogram.	[34]
	Region growing: Color-based region growing segments an image based on color similarity. This algorithm starts with a seed pixel or region and iteratively adds neighboring pixels wit similar color characteristics until the desired ROI is obtained.	[35], [36]
Shape-based methods Shape-based methods	Machining learning algorithms: Different supervised and unsupervised machining learning algorithms are also efficient for separating foreground and background pixels based on the color freaures. Commonly used methods include K-Means clustering, SVM, Rendom Forest, and CNN.	[20], [29]
	Template matching: It uses a pre-defined template/shape of the traffic sign to match against the image pixels. If the shape of the traffic sign in the image matches with the template, it is identified as a traffic sign.	[22], [23]
	Hough transform: It is a feature extraction algorithm. It converts the image space into a parameter space, where points that lie on the same line or curve in the image space are grouped together in the parameter space.	[14], [35]
	Contour detection: Contour detection algorithms can be used to distinguish the boundaries of traffic signs from images by implementing edge detection algorithms such as Canny edge detection or Sobel edge detection.	[26], [27]
	Distance transform matching: It first detects the object boundaries by performing edge detection algorithms. Then it assigns a distance value to each pixel based on its distance to the nearest object boundary.	[28], [29]

2.4.3.2 LiDAR-based Methods for TCI Detection

LiDAR data is commonly used to detect road markings. Two features of road markings make them suitable to be detected through LiDAR data. First, most road markings are decorated on asphalt concrete pavements with highly light-reflective coatings such as yellow or white, leading to higher reflected intensity values that can be easily recognized when scanned by LiDAR scanners [37], [38]. Second, similar to other traffic signs, road markings also show linear features with known width and length. The shapes and arrangement of road markings provide semantic information for target extraction and recognition [39].

Road marking detection using LiDAR data can be generally divided into two steps: road surface extraction and road marking extraction. The raw LiDAR data capture not only the road surface but also other objects like trees, cars, or buildings, and even some isolated LiDAR points in the air. Therefore, road surface extraction is needed to remove those non-ground points. There are different ways to accomplish this step, such as the elevation-based method and intensity-based method [40], [41]. Usually, the filtered points are then converted into 2D geo-referenced images based on the reflection intensity for road marking extraction and classification [42]. Once the 2D reflection intensity image generated, road markings can then be isolated from it based on their distinctively high values in terms of reflection intensity. In the resulted intensity images, the brightness indicates the reflection intensity and thereby, researchers could use various algorithms to identify the edges or shape boundaries of the road markings of interests [37], [39]. For example, Yang implemented a Progressive probabilistic Hough Transformation (PPHT) method [48], which is a variation of the standard Hough transform to extract the linear features from the georeferenced reflectance intensity image. [48] implemented an

Edge Detection and Edge Constraint (EDEC) to detect road markings [51]. They first organized the filtered LiDAR points clouds into scan lines. Then they extracted road marking points from road points by detecting the edges between road surface and road markings in scan lines. More commonly used algorithms for road surface detection and marking detection are listed in Table 3.

Table 3. LiDAR-based Detection for Road Surface and Markings.

Module	Typical algorithm or models	Reference
Road Surface Extraction	Positioning and orientation system (POS): record the height of LiDAR scanner to the scanned surface and thus identify if there are any off-ground points.	[38]
	Voxel-Based Normalized Cut Segmentation: partitions point cloud data into an octree structure with a voxel size and expand the octree until it reaches the top boundary; the point clouds with top voxel higher than a certain threshold will be filtered out from the road surface.	[40], [42]
Road Marking Extraction	Progressive probabilistic Hough Transformation (PPHT): a variation of the standard Hough transform to extract the linear features from the georeferenced reflectance intensity image; in addition to the orientation, it also computes an extent for individual line-shaped road marks.	[39]
	U-Shaped Capsule Network: a capsule network that consists of traditional convolutional layers, primary capsule layers, convolutional capsule layers, and deconvolutional capsule layers; it can derive the shape and position information of the road markings.	[37]
	Hybrid Capsule Network: this network consists of a convolutional capsule network and an FC capsule network, and it can categorize these road markings by encoding high-level and low-level features from input images.	[37]
	Edge Detection and Edge Constraint (EDEC): this method detect the edges of road surface and use these edges to constrain the search space for road markings. It organizes preprocessed LiDAR points clouds into scan lines. It then extracts road marking points from road points by detecting the edges between road surface and road markings in scan lines.	[41]

Although less common, some researchers have investigated the use of LiDAR-based methods for traffic sign detection. Similar to road marking detection, reflection intensity is the primary feature used to extract traffic signs. Due to the properties of LiDAR data, the detected traffic signs via LiDAR-based methods are always in three-dimensional space, which makes it possible to gain accurate information not only about the location and shape of the signs, but also the orientation and position of the sign bases or poles, which are much more complex to calculate via imagery-based methods [6], [43].

The most significant limitation of LiDAR-based methods is their inability to provide color information of traffic signs, making it necessary to pair LiDAR data with imagery data to obtain this information. It is worth noting that researchers could calculate the color deterioration in the imagery by examining the reflection intensity of the targeted traffic signs derived from LiDAR data [20]. Therefore, the combination of LiDAR and imagery is expected to maximize the performance of traffic sign detection in this regard.

2.4.3.3 Methods for TCI Classification

After the TCIs have been detected from the input data, the research team needed to recognize the content of these TCIs and recognize their types. As shown in Table 4, TCI classification methods could be broadly classified into two categories: methods performed on created features and deep-learning-based methods.

Table 4. TCI Classification Methods.

Category	Typical algorithms and models	Ref
Methods performed on created features	Support Vector Machine (SVM): a supervised learning method that constructs a hyperplane to separate data into classes. The "support vectors" are data points that define the maximum margin of the hyperplane. This lightweight classifier was intensively utilized by existing studies, which can potentially handle the classification in real-time.	[44]
	Adaptive Boosting (AdaBoost): a combination of multiple learning algorithms that can be utilized for regression or classification. AdaBoost assigns weights to weak classifiers based on their quality. The resulting strong classifier is a linear combination of weak classifiers with the appropriate weights.	[45]
	Tree-based models: a series of "if-then" rules to generate predictions from one or more decision trees. All tree-based models can be used for either regression or classification, and sometimes certain decision rules form a set of criteria could judge the traffic sign candidates by their various properties.	[32], [46]
	Template matching: this kind of methods are used to search for existing similar training samples. The training samples were pre-characterized by a set of features. The input is compared with different pre-coded samples to examine their similarity. The input type will be assigned as the same as the most similar sample.	[47]
Deep-learning-based methods	Convolutional Neural Networks (CNN): a computational processing network system inspired by biological nervous system, which comprised of neurons that self-optimize through learning, loading input image vectors and making decisions as outputs after being processed by multiple hidden operation layers. (O'Shea and Nash, 2015)	[51], [58], [60]
	Extreme learning machine (ELM): a single-hidden-layer feedforward neural network (SFNN) that encapsulates all classes of traffic signs. It only estimates the weight vector between the hidden and output layers using a least-square strategy. Therefore, its computational cost drops very much and is easy for parameter tuning.	[51]

The general classification methods are conducted based on a set of features to distinguish different objects. First, some methods can be applied to extract the targeted features, such as HOG, HOG variations (e.g., pyramid HOG [PHOG], HOGv) [52], [53], and Integral Channel Features (ICF) [54]. Based on these features, various models can be implemented to classify the traffic sign types, such as SVM [44], AdaBoost [45], Tree-based models [32], [46], Template Matching [47], among others. These models use learning algorithms to classify each input traffic sign image. For example, Figure 3 illustrates how a tree-based model distinguishes road sign types based on the generated shape and color features (Haar and Safran, 2012).

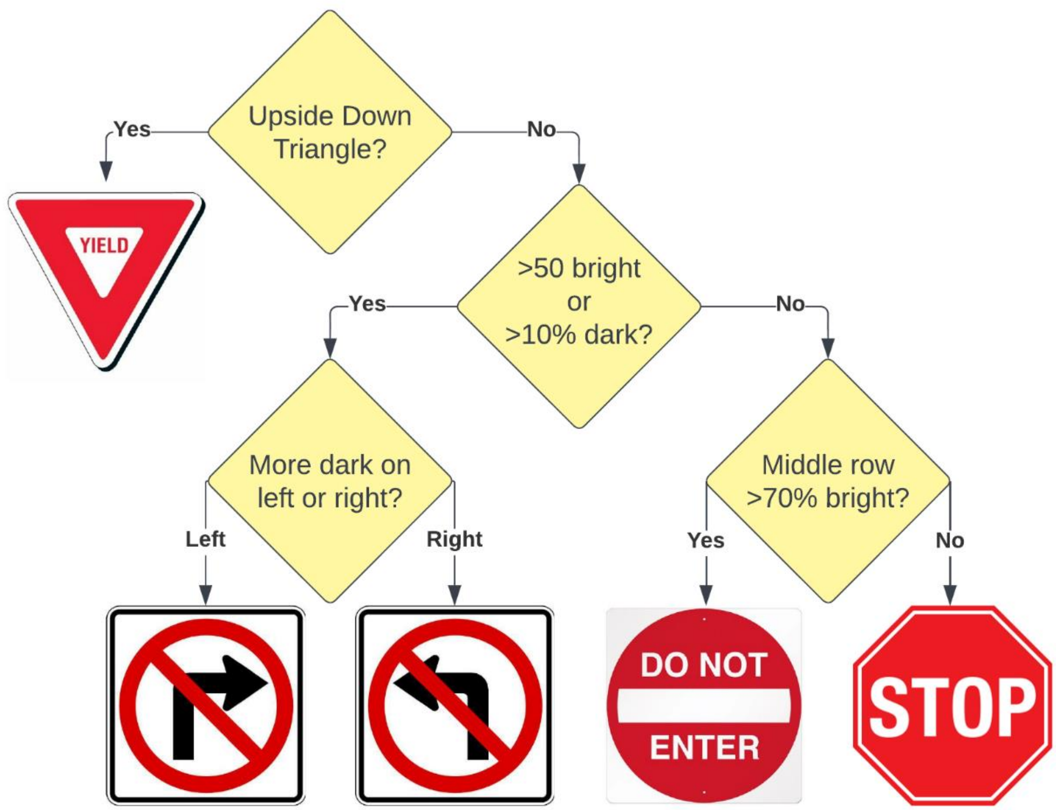


Figure 3. Tree-based Model in TCI Classification (adapted from [32]).

Deep learning methods also have been implemented for TCI classification, which has acquired a general interest in recent years because of its high performance in road asset classification and the power of representational learning from raw data. Deep learning methods use a cascade of hidden layers in the neural network for extracting and transforming features. The output from the previous one is used as the input for the successive layer. Higher-level features are derived from lower-level features to form a hierarchical representation. Among the deep learning models, convolutional neural networks (CNN) are widely used in image classification [48], [49]. CNN-based detectors prevail in object detection tasks. Unlike manually labeled features, CNN-based detectors use different convolutional layers to extract features directly from raw images [50]. In addition to CNN, some other similar neural network models are also popular in deep-learning-based TCI classification methods, such as the Extreme learning machine (ELM) [51].

2.4.3.4 Detection and Classification Integrated Deep Learning Methods

The advances in deep learning techniques offer plausible solutions to object detection and classification tasks. Some networks even could complete the detection and classification within the same model. Some state-of-the-art object-detection networks, such as Fast Region-Based CNN (Fast R-CNN) [55], Faster R-CNN [33], [34]. [66] Mask R-CNN [58], Single Shot Detector [59], [60], and You Only Look Once (YOLO) [34], [71], combined with various feature extractors (Resnet V1 50, Resnet V1 101, Inception V2, Inception Resnet V2, Mobilenet V1, and Darknet-19) are used for traffic sign recognition.

2.4.3.5 Methods for TCI Positioning

The geolocation of the detected TCIs is critical for establishing TCI inventory. The positioning procedure of TCIs could be broadly divided into three modules: GPS-based vehicle positioning, inertial-based correction, and image-based asset localization.

GPS-Based Vehicle Positioning: GPS is the most direct and straightforward method and also the most basic step for collecting position data of the targeted TCIs. Usually, the positioning data using GPS are collected at the same time as the imagery data when running the survey vehicles where the GPS receivers are installed along with the camera or LiDAR scanners. For every point where the road data is captured, the corresponding coordinate data would be recorded as well [34], [72], [73].

Inertial-Based Supplement and Correction: GPS receivers could sometime be blocked from the signals for some reasons (due to forests and high buildings). In this case, some inertial-based methods and instruments are utilized as supplements and corrections to GPS-based methods. When a GPS receiver loses the signals, the inertial instrument will record the movement of the vehicle, thus calculating the current position based on the last position data recorded by the GPS receiver. Thereby, the survey vehicles can keep collecting positioning data until GPS receivers regain the signals. Inertial-based methods include inertial measurement units (IMUs), distance measuring instruments, and an inertial reference system [34], [74].

Image-Based Asset Location Calculation: For both GPS-based positioning methods and inertial-based correction methods, the collected position data generally represents the location of the vehicle, not the targeted TCIs. To solve this issue, image-based methods could be used to calculate the physical position of the targeted TCIs. This kind of method mainly obtains the position data by connecting the same target in multiple images. After the TCIs are detected from the input images, we could combine their positions in different images, thereby retrieving their relative positions to the survey vehicles. Based on these relative positions and the geolocation data of survey vehicles we collected using GPS and inertial methods, we can finally estimate the actual location of those assets. For example, each combination of two detections could generate a hypothetical location of the target asset, and the subsequent captures could result in clustered points around the physical position of that asset, followed by a clustering algorithm to extract that physical location [14]–[16].

2.4.4 Narrative Review for TCI Inventory Establishment and Management

2.4.4.1 General Procedure for Inventory Establishment

To the best of our knowledge, there is no socially acknowledged process for building TCI inventory. In light of the review findings, the workflow proposed by Nima Kargah-Ostadi could serve as an initial framework for roadway asset inventory governance. With the technology of Artificial Intelligence (AI), Cloud, and internal network, Nima Kargah-Ostadi's team raised a framework for automated inventory establishment for roadway assets. As shown in Figure 4, in Nima Kargah-Ostadi's method, the imagery data are collected using survey vehicles and the asset extraction and localization are accomplished at the same time with AI technology. After the data is collected and delivered to the office, it is uploaded to the central network or online cloud to share across different organizations. Through a quality control process conducted on a subset of the data, the TAM department could decide to accept or reject the established inventory [13].

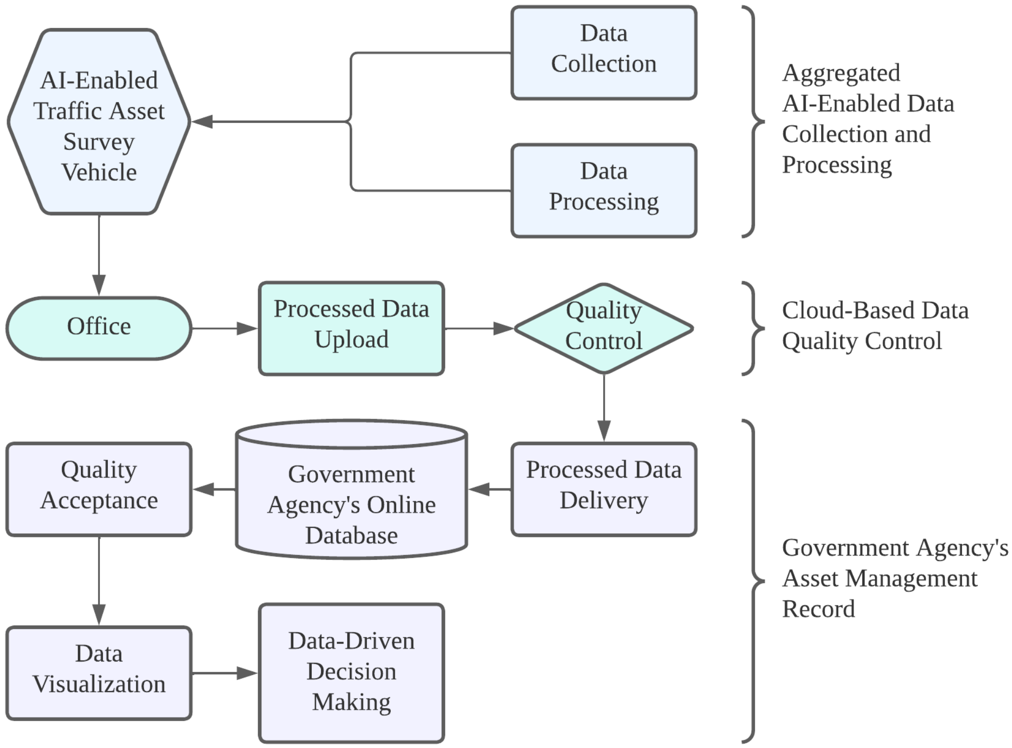


Figure 4. Nima Kargah-Ostadi's Framework for Roadway Asset Inventory Governance (adapted from [5])

TCI Inventory Update

Inventory updates are conducted to maintain a baseline inventory, which contains the existing asset records in a certain study area and has been examined to be corrected. As illustrated in Figure 5, when performing inventory updates, newly processed assets need to be compared with those in the baseline inventory one by one. If a newly digitized asset has the same type and location as the corresponding one in the baseline inventory, it would be labeled as "unchanged" in the baseline inventory [64]. After the asset comparison, all assets would be classified into three categories: (1) unchanged signs, which have matched type and position with an existing sign in the baseline inventory; (2) removed signs, which exist in the baseline inventory but do not exist in the new digitized dataset; and (3) newly found signs, which are newly digitized but cannot match with any existing assets in the baseline inventory [65]. The inventory updates need to be completed in two modules: asset location match and asset type match.

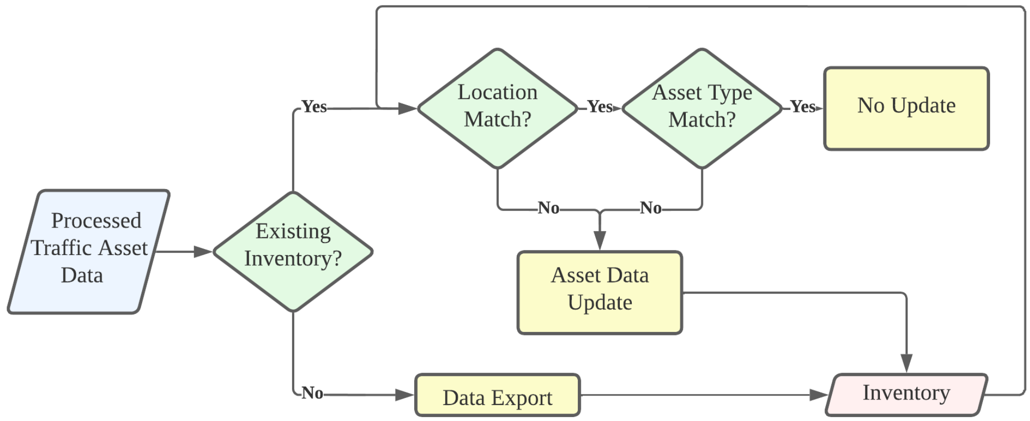


Figure 5. Workflow for TCI Inventory Update.

2.4.4.1.1 Asset Location Match

Asset location match**:** A critical challenge in inventory updates is to compare the location of collected road assets. When the location of a newly inventoried asset doesn't match with the baseline inventory, it could imply a location change of the asset or a GPS positioning error. For example, researchers noticed that the position offset usually happens when the GPS signal is blocked by trees in forest areas or by high buildings in urban areas. GPS disturbances at a certain asset's position could lead to the position offset for a series of subsequent TCIs. Studies have demonstrated that the position offsets among these TCIs are generally uniform in the same area, which can be identified and corrected through the window detection method. For example, Hazelhoff et al. proposed a context-based drift correction by leveraging the positions of all neighboring traffic signs within a 150-meter buffer to correct the position offset for a newly detected sign [15], [16].

2.4.4.1.2 Asset Type Match

Comparing the asset type along with the location is an important step in detecting changes or mutations in the new inventory. When comparing the newly inventoried assets with existing ones, if they are confirmed with the same asset type at the same location, they are considered matched assets. Please note that although TCIs of the same type usually have a similar appearance, small differences may still exist in some details. These minor differences may confuse the classification system and lead to a classification error. To prevent this kind of error, a cross-correlation of ideal templates of these asset types can be applied through the following procedure: (1) create a database of templates for teach traffic asset type; (2) conduct a correlation analysis between each pair of asset type templates; the asset types with a correlation coefficient greater than a certain threshold would be considered as "similar type"; (4) When a newly inventoried asset is classified as a different type, researchers need to check if these two types are "similar type". If they are, it would be considered as a match. Otherwise, the asset type would be considered not matched, and the asset information would be updated [15].

2.5 DISCUSSION AND CONCLUSIONS

TCIs play a vital role in guiding traffic flows. How to effectively establish and maintain a TCI inventory is an essential yet unanswered question for transportation agencies, not just in the United States but in all nations. To fill this gap, this study used a mixed review approach to summarize the current studies and practices for TCI digitizing and inventorying. The research team first performed an SLR to summarize the mainstream data sources and models for TCI extraction and detection. Then, the team performed an NR to collect the necessary modules and solutions for building and maintaining an effective TCI inventory. This study represents the first to systemically review the whole workflow for TCI inventory establishment by combining advances in multiple disciplines. This study makes unique contributions to transportation agencies by providing a valued guideline on TCI inventory creation and maintenance. Meanwhile, this study also helps data scientists understand the landscape of available data sources and computer vision techniques for digitizing and inventorying other road and urban assets.

2.5.1 Major Findings for the TCI Inventory Establishment and Management

As previously mentioned, the major findings of this study are centered around addressing three specific research questions regarding TCI digitization and inventory development, including:

- What types of data have been used for TCI digitization?
- What methods have been used for TCI digitization?
- What procedures have been implemented for establishing, maintaining, and updating the TCI inventory?

For the available data sources, this review shows that the primary data used in TCI digitization are collected via survey vehicles with equipment installed (e.g., camera, LiDAR, GPS) for obtaining two categories of data: TCI information and location. TCI information is obtained from imagery data collected by cameras, mainly for traffic signs and lights. LiDAR point clouds are more commonly used for detecting road markings. TCI location data are usually collected using GPS receivers with inertial-based instruments or units used as supplements.

For the TCI digitization methods, TCI data extraction generally involves three modules: detection, classification, and localization. The detection module differs for imagery and LiDAR data. For imagery data, color-based, shape-based, and learning-based methods are the most utilized for detection. For LiDAR data, road surface extraction is usually performed first, followed by the extraction of road markings based on the reflection intensity of points. Classification of detected assets into their respective types is typically accomplished using learning-based algorithms such as CNN and SVM. Some deep learning algorithms like Faster R-CNN, SSD, and YOLO can integrate detection and classification. To determine the actual locations of assets, GPS receivers and inertial-based instruments are usually integrated to derive the location data of the survey vehicle, and imagery data can be used to estimate the relative position of targeted assets with respect to the survey vehicle.

After digitation and localization, the data need to be sent to the TCI management offices or agencies for quality control. The checked data would then be used in new inventory establishment if there is not existing inventory yet, or in inventory update and mutation detection if there is already existing inventory.

2.5.2 Limitation in Current TCI Inventory Management

Based on the review, the research team noticed the following limitations existed in current studies for an effective and reliable TCI inventory establishment in real-world settings. These can be summarized into two categories: data limitation and technique limitation.

2.5.2.1 Limitation in Data

Two main data limitations were widely reported by existing studies, including data quality-related issues and lacking benchmark for model training.

The anomalies of TCI have posed great challenges for the image-based detection and classification process in reliability and accuracy. The anomalies may come from two major reasons: the circumstance of targeted TCI during data collection, such as extreme weather or environmental condition, low visibility or noticeability, or inappropriate position, and the condition of TCIs themselves, such as occlusion, damage, or deformation of the targeted TCI object. The captured images with those anomalies may not be able to provide enough information to interpret them; thus, the targeted TCIs may be misclassified or even fail to be detected.

The quality of the captured data is another challenge for TCI extraction. For example, motion blur causes sharp edges in the captured frame to weaken, which makes it difficult to detect and classify TCIs. Image noise can appear as black and white pixels randomly distributed throughout an image and greatly reduce the accuracy of TCI detection.

For most detection methods and classification methods, a reliable benchmarked dataset for training the model plays an important role in ensuring the model's accuracy. As summarized in Table 5, only a few counties have well-established benchmark datasets of TCIs. Since the format of TCIs varies across countries, the trained models based on one country's benchmark dataset may perform differently in other countries.

Table 5. Commonly Used Benchmark Datasets for TCI Detection.

Dataset Name	Regions	Sample Size	TCI Types	Reference
German Traffic Sign Recognition Benchmark	Germany	~52,000	sign	[66], [67]
German Traffic Sign Detection Benchmark	Germany	1,000	sign	[30], [66]
KUL Belgium Traffic Signs Dataset	Belgium	13,444	sign	[52], [68]
Laboratory for Intelligent & Safe Automobiles Datasets	California, US.	25,913	light	[50]
Swedish Traffic Signs Datasets	Sweden	3,488	sign	[31], [68]
Tsinghua-Tencent 100K dataset	China	33,071	sign	[69], [70]
Stereopolis Database	France	251	sign	[44]
RUG Traffic Sign Image Database	Netherlands	48	sign	[64]
The Challenging Unreal and Real Environments dataset	N/A	1,719,900	sign	[72]
Mapping and Assessing the State of Traffic InFrastructure	Croatia	~6,000 (TS2009), ~3,000 (TS2010), ~1,000 (TS2011)	sign	[73]
WPI traffic light dataset	Worcester, USA	10,034	light	[74]
Bosch Small Traffic Light dataset	N/A	~13,300	light	[75]
Cyber Identity Biometrics Traffic Sign Dataset	N/A	690	sign	[76]
DFG Traffic sign Dataset	Slovenia	6,957	sign	[58]
HERE map data	N/A		map	[77]
International Cybernetics Corporation	N/A		roadway image	[13]
Microsoft Common Objects in Context	N/A	200,000+	common objects	[78]
N/A	UK	1,200	sign	[79]
Russian Traffic Sign Dataset	Russia	80,000+	sign	[79]
Spanish Traffic Sign dataset	Spain	615+	sign	[80]
PASCAL Visual Object Classes	N/A	~3,000	visual objects	[56]

Note: N/A = not applicable.

2.5.2.2 Limitation in Techniques

The accuracy of detection, classification, and localization for road assets are all essential in TAM. However, most existing studies primarily focus on detection and classification accuracy; the positioning accuracy of the digitized TCIs is less discussed and underexplored. How to accurately map the digitized TCI information remains a vital research question for future TAM and high-precision mapping.

Data redundancy and integration are emerging issues in TCI digitization. Various data sources and methods have been used to digitize TCIs, and transportation agencies and commercial companies have created or already formed some TCI datasets. In future TAM, combining and integrating different existing datasets with newly created datasets to obtain an updated TCI database will be a challenge. Additionally, the inconsistency among different inventories may increase the difficulty of data integration.

Effectively capturing and updating all necessary information for TCI management are remaining challenges. T. Nguyen et al. introduced five key elements that a successful road asset inventory should contain: type, position, condition, installation date, and maintenance history [86]. However, existing studies have primarily focused on TCI type detection and recognition, with studies on detecting road asset conditions still underexplored. Additionally, an effective TCI management system is needed to record the installation date and maintenance history of TCIs for optimal management.

2.5.3 Future Work

To address the limitations and challenges discussed, further efforts should be made to reduce the influence of data quality and improve the performance of TCI detection and classification. For example, efficient data pre-processing techniques should be adopted to improve the quality of input data. Meanwhile, more efforts are needed to establish and enrich benchmark datasets for TCIs. This could facilitate the development of new solutions for reducing the impacts caused by the lack of benchmarks for model training and improving classification accuracy. Last but not least, a generalized standard and procedure are greatly needed to guide the TCI inventory establishment and maintenance, and more importantly, ease the data integration across different data sources and providers.

2.6 REFERENCES

[1] R. J. Torraco, “Writing Integrative Literature Reviews: Guidelines and Examples,” *Human Resource Development Review*, vol. 4, no. 3, 2005, doi: 10.1177/1534484305278283.

[2] Y. Xiao and M. Watson, “Guidance on Conducting a Systematic Literature Review,” *Journal of Planning Education and Research*, vol. 39, no. 1. 2019. doi: 10.1177/0739456X17723971.

[3] The University of Alabama at Birmingham, “Reviews: From Systematic to Narrative: Narrative Review.” Accessed: May 08, 2022. [Online]. Available: https://guides.library.uab.edu/c.php?g=63689&p=409774

[4] X. Li *et al.*, “COVID-19 impacts on mobility, environment, and health of active transportation users,” *Cities*, vol. 131, p. 103886, Dec. 2022, doi: 10.1016/J.CITIES.2022.103886.

[5] A. Martín-Martín, E. Orduna-Malea, M. Thelwall, and E. Delgado López-Cózar, "Google Scholar, Web of Science, and Scopus: A systematic comparison of citations in 252 subject categories," *J Informetr*, vol. 12, no. 4, 2018, doi: 10.1016/j.joi.2018.09.002.
[6] J. Landa and D. Prochazka, "Automatic Road Inventory Using LiDAR," *Procedia Economics and Finance*, vol. 12, no. March, pp. 363–370, 2014, doi: 10.1016/S2212-5671(14)00356-6.
[7] Y. Nie, Q. Chen, T. Chen, Z. Sun, and B. Dai, "Camera and Lidar Fusion for Road Intersection Detection," *2012 IEEE Symposium on Electrical & Electronics Engineering (EEESYM)*, pp. 273–276, 2012, doi: 10.1109/EEESym.2012.6258642.
[8] X. Li, H. Ning, X. Huang, Y. Kang, B. Dadashova, and A. Ma, "Urban infrastructure audit: an effective protocol to digitize signalized intersections by mining street view images," *Cartogr Geogr Inf Sci*, vol. 49, no. 1, pp. 32–49, 2021, doi: 10.1080/15230406.2021.1992299.
[9] P. Siegmann, "Traffic Sign Recognition System for Inventory Purposes," pp. 590–595, 2008.
[10] Y. Tsai, "A Generalized Framework for Parallelizing Traffic Sign Inventory of Video Log Images Using Multicore Processors," vol. 27, pp. 476–493, 2012, doi: 10.1111/j.1467-8667.2012.00765.x.
[11] H. Wang, B. Wang, B. Liu, X. Meng, and G. Yang, "Pedestrian recognition and tracking using 3D LiDAR for autonomous vehicle," *Rob Auton Syst*, 2017, doi: 10.1016/j.robot.2016.11.014.
[12] F. Almutairy, T. Alshaabi, J. Nelson, and S. Wshah, "ARTS: Automotive Repository of Traffic Signs for the United States," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 1, pp. 457–465, Jan. 2021, doi: 10.1109/TITS.2019.2958486.
[13] N. Kargah-Ostadi, A. Waqar, and A. Hanif, "Automated Real-Time Roadway Asset Inventory using Artificial Intelligence," *Transp Res Rec*, vol. 2674, no. 11, pp. 220–234, Sep. 2020, doi: 10.1177/0361198120944926.
[14] L. Hazelhoff, I. Creusen, and P. H. N. De With, "Robust detection, classification and positioning of traffic signs from street-level panoramic images for inventory purposes," in *Proceedings of IEEE Workshop on Applications of Computer Vision*, 2012. doi: 10.1109/WACV.2012.6163006.
[15] L. Hazelhoff, I. Creusen, and P. H. N. De With, "Mutation detection for inventories of traffic signs from street-level panoramic images," in *Video Surveillance and Transportation Imaging Applications 2014*, 2014. doi: 10.1117/12.2036732.
[16] L. Hazelhoff, I. Creusen, and P. H. N. De With, "Mutation detection system for actualizing traffic sign inventories," in *VISAPP 2014 - Proceedings of the 9th International Conference on Computer Vision Theory and Applications*, 2014. doi: 10.5220/0004793707050713.
[17] D. V. Filatov, K. V. Ignatev, A. V. Deviatkin, and E. V. Serykh, "Application of machine learning methods for traffic signs recognition," in *IOP Conference Series: Materials Science and Engineering*, Institute of Physics Publishing, Feb. 2018. doi: 10.1088/1757-899X/315/1/012008.
[18] S. B. Wali, M. A. Hannan, A. Hussain, and S. A. Samad, "An Automatic Traffic Sign Detection and Recognition System Based on Colour Segmentation, Shape Matching, and SVM," *Math Probl Eng*, vol. 2015, 2015, doi: 10.1155/2015/250461.

[19] J. D. Zhao, Z. M. Bai, and H. B. Chen, "Research on Road Traffic Sign Recognition Based on Video Image," in *Proceedings - 10th International Conference on Intelligent Computation Technology and Automation, ICICTA 2017*, Institute of Electrical and Electronics Engineers Inc., Oct. 2017, pp. 110–113. doi: 10.1109/ICICTA.2017.31.
[20] S. B. Wali *et al.*, "Vision-Based Traffic Sign Detection and Recognition Systems : Current Trends and Challenges," 2019.
[21] S. B. Wali, M. A. Hannan, A. Hussain, and S. A. Samad, "An Automatic Traffic Sign Detection and Recognition System Based on Colour Segmentation , Shape Matching , and SVM," vol. 2015, 2015.
[22] L. Jia, X. Shi, and L. Wei, "Design of Traffic Sign Detection and Recognition Algorithm Based on Template Matching," in *2020 IEEE 2nd International Conference on Civil Aviation Safety and Information Technology (ICCASIT*, 2020, pp. 237–240. doi: 10.1109/ICCASIT50869.2020.9368750.
[23] P. Pandey and R. Kulkarni, "Traffic Sign Detection Using Template Matching Technique," in *2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA)*, 2018, pp. 1–6. doi: 10.1109/ICCUBEA.2018.8697847.
[24] F. Almutairy, T. Alshaabi, J. Nelson, and S. Wshah, "ARTS : Automotive Repository of Traffic Signs for the United States," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 1, pp. 457–465, 2021, doi: 10.1109/TITS.2019.2958486.
[25] B. S. Shekar and G. Harish, "A Machine Learning Model For Detection and Recognition Of Traffic Signs," *2021 International Conference on Intelligent Technologies (CONIT)*, pp. 1–4, 2021, doi: 10.1109/CONIT51480.2021.9498270.
[26] G. Yu, A. Lei, H. Li, Y. Wang, and Z. Wang, "A Real-Time Traffic Light Detection Algorithm Based on Adaptive Edge Information," pp. 1–10, 2021, doi: 10.4271/2018-01-1620.Abstract.
[27] Y. Taki and E. Zemmouri, "An Overview of Real-Time Traffic Sign Detection and Classificaiton," pp. 344–356, 2021, doi: 10.1007/978-3-030-66840-2.
[28] D. M. Gavrila, "Traffic Sign Recognition Revisited," in *Mustererkennung 1999*, J. M. and F. A. and F. P. Förstner Wolfgang and Buhmann, Ed., Berlin, Heidelberg: Springer Berlin Heidelberg, 1999, pp. 86–93.
[29] S. Khalid, N. Muhammad, and M. Sharif, "Automatic measurement of the traffic sign with digital segmentation and recognition," *IET Intelligent Transport System*, vol. 13, no. 2, pp. 269–279, 2018, doi: 10.1049/iet-its.2018.5223.
[30] K. A. Pronchuk and P. Y. Yakimov, "Development of traffic signs recognition WebService based on convolutional neural networks," in *CEUR Workshop Proceedings*, 2018. doi: 10.18287/1613-0073-2018-2212-132-138.
[31] A. Ellahyani and M. el Ansari, "A new designed descriptor for road sign recognition," in *Proceedings - 3rd International Conference on Advanced Technologies for Signal and Image Processing, ATSIP 2017*, 2017. doi: 10.1109/ATSIP.2017.8075547.
[32] S. Haar and M. S. Safran, "A Method of US Traffic Sign Detection and Recognition," 2012.
[33] Y. Feng, D. Kong, P. Wei, H. Sun, and N. Zheng, "A Benchmark Dataset and Multi-Scale Attention Network for Semantic Traffic Light Detection," in *2019 IEEE Intelligent Transportation Systems Conference, ITSC 2019*, 2019. doi: 10.1109/ITSC45078.2019.9086430.

[34] Y. Daraghmi and A. M. Hasasneh, “Accurate Real-Time Traffic Sign Recognition Based on the Connected Component Labeling and the Color Histogram Algorithms,” vol. 4, no. 5, pp. 417–421, 2016, doi: 10.18178/ijsps.4.5.417-421.

[35] Y. L. Chang and X. Li, “Adaptive Image Region-Growing,” *IEEE Transactions on Image Processing*, vol. 3, no. 6, 1994, doi: 10.1109/83.336259.

[36] V. R. Deshmukh, G. K. Patnaik, and M. E. Patil, “Real-Time Traffic Sign Recognition System based on Colour Image Segmentation,” *Int J Comput Appl*, vol. 83, pp. 975–8887, 2013.

[37] L. Ma *et al.*, “Capsule-Based Networks for Road Marking Extraction and Classification from Mobile LiDAR Point Clouds,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 4, pp. 1981–1995, Apr. 2021, doi: 10.1109/TITS.2020.2990120.

[38] R. Yang, Q. Li, J. Tan, S. Li, and X. Chen, “Accurate road marking detection from noisy point clouds acquired by low-cost mobile LiDAR systems,” *ISPRS Int J Geoinf*, vol. 9, no. 10, Oct. 2020, doi: 10.3390/ijgi9100608.

[39] B. Yang, L. Fang, Q. Li, and J. Li, “Automated extraction of road markings from mobile lidar point clouds,” *Photogramm Eng Remote Sensing*, vol. 78, no. 4, pp. 331–338, 2012, doi: 10.14358/PERS.78.4.331.

[40] M. Cheng, H. Zhang, C. Wang, and J. Li, “Extraction and Classification of Road Markings Using Mobile Laser Scanning Point Clouds,” *IEEE J Sel Top Appl Earth Obs Remote Sens*, vol. 10, no. 3, pp. 1182–1196, Mar. 2017, doi: 10.1109/JSTARS.2016.2606507.

[41] L. Yan, H. Liu, J. Tan, Z. Li, H. Xie, and C. Chen, “Scan line based road marking extraction from mobile LiDAR point clouds,” *Sensors (Switzerland)*, vol. 16, no. 6, Jun. 2016, doi: 10.3390/s16060903.

[42] Y. Yu, J. Li, H. Guan, F. Jia, and C. Wang, “Learning hierarchical features for automated extraction of road markings from 3-D mobile LiDAR point clouds,” *IEEE J Sel Top Appl Earth Obs Remote Sens*, vol. 8, no. 2, pp. 709–726, Feb. 2015, doi: 10.1109/JSTARS.2014.2347276.

[43] M. Javanmardi, Z. Song, and X. Qi, “Automated traffic sign and light pole detection in mobile LiDAR scanning data,” *IET Intelligent Transport Systems*, vol. 13, no. 5, pp. 803–815, May 2019, doi: 10.1049/iet-its.2018.5360.

[44] S. B. Wali, M. A. Hannan, A. Hussain, and S. A. Samad, “Comparative survey on traffic sign detection and recognition: A review,” *Przeglad Elektrotechniczny*, vol. 91, no. 12. Wydawnictwo SIGMA-NOT, pp. 38–42, 2015. doi: 10.15199/48.2015.12.08.

[45] V. Balali and M. Golparvar-Fard, “Evaluation of Multiclass Traffic Sign Detection and Classification Methods for U.S. Roadway Asset Inventory Management,” *Journal of Computing in Civil Engineering*, vol. 30, no. 2, p. 04015022, Mar. 2016, doi: 10.1061/(asce)cp.1943-5487.0000491.

[46] Y. (James) Tsai, P. Kim, and Z. Wang, “Generalized Traffic Sign Detection Model for Developing a Sign Inventory,” *Journal of Computing in Civil Engineering*, vol. 23, no. 5, 2009, doi: 10.1061/(asce)0887-3801(2009)23:5(266).

[47] A. U. Peker, O. Tosun, H. L. Akin, and T. Acarman, “Fusion of map matching and traffic sign recognition,” in *IEEE Intelligent Vehicles Symposium, Proceedings*, 2014. doi: 10.1109/IVS.2014.6856536.

[48] B. Smitha Shekar and G. Harish, “A Machine Learning Model for Detection and Recognition of Traffic Signs,” in *2021 International Conference on Intelligent Technologies, CONIT 2021*, 2021. doi: 10.1109/CONIT51480.2021.9498270.

[49] L. Yang, X. Wang, and J. Wang, “Deep Traffic Light Perception with Spatiotemporal Analysis for Autonomous Driving,” in *European Control Conference 2020, ECC 2020*, 2020. doi: 10.23919/ecc51009.2020.9143756.

[50] G. Yu, A. Lei, H. Li, Y. Wang, Z. Wang, and C. Hu, “A Real-Time Traffic Light Detection Algorithm Based on Adaptive Edge Information,” in *SAE Technical Papers*, SAE International, Aug. 2018. doi: 10.4271/2018-01-1620.

[51] Z. Huang, Y. Yu, and J. Gu, “A novel method for traffic sign recognition based on extreme learning machine,” in *Proceedings of the World Congress on Intelligent Control and Automation (WCICA)*, 2015. doi: 10.1109/WCICA.2014.7052932.

[52] Z. Huang, Y. Yu, J. Gu, and H. Liu, “An Efficient Method for Traffic Sign Recognition Based on Extreme Learning Machine,” *IEEE Trans Cybern*, vol. 47, no. 4, pp. 920–933, Apr. 2017, doi: 10.1109/TCYB.2016.2533424.

[53] H. Li, F. Sun, L. Liu, and L. Wang, “A novel traffic sign detection method via color segmentation and robust shape matching,” *Neurocomputing*, vol. 169, pp. 77–88, Dec. 2015, doi: 10.1016/j.neucom.2014.12.111.

[54] A. Mogelmose, D. Liu, and M. M. Trivedi, “Detection of U.S. Traffic Signs,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 6, pp. 3116–3125, Dec. 2015, doi: 10.1109/TITS.2015.2433019.

[55] J. Bengtson *et al.*, “Deep learning methods for recognizing signs/objects in road traffic,” 2018.

[56] Z. Deng, “DETECT TRAFFIC SIGNS FROM LARGE STREET VIEW IMAGES WITH DEEP LEARNING,” 2019.

[57] W. G. Pan, Y. H. Chen, and B. Liu, “Traffic Light Detection for Self-Driving Vehicles Based on Deep Learning,” in *Proceedings - 2019 15th International Conference on Computational Intelligence and Security, CIS 2019*, Institute of Electrical and Electronics Engineers Inc., Dec. 2019, pp. 63–67. doi: 10.1109/CIS.2019.00022.

[58] D. Tabernik and D. Skocaj, “Deep Learning for Large-Scale Traffic-Sign Detection and Recognition,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 4, pp. 1427–1440, Apr. 2020, doi: 10.1109/TITS.2019.2913588.

[59] S. M. M. Ahsan, S. Das, S. Kumar, and Z. La Tasriba, “A Detailed Study on Bangladeshi Road Sign Detection and Recognition,” in *2019 4th International Conference on Electrical Information and Communication Technology, EICT 2019*, 2019. doi: 10.1109/EICT48899.2019.9068760.

[60] M. Lopez-Montiel, Y. Rubio, M. Sánchez, and U. Orozco-Rosas, “Evaluation of algorithms for traffic sign detection,” SPIE-Intl Soc Optical Eng, Sep. 2019, p. 20. doi: 10.1117/12.2529709.

[61] S. Song, Y. Li, Q. Huang, and G. Li, “A new real-time detection and tracking method in videos for small target traffic signs,” *Applied Sciences (Switzerland)*, vol. 11, no. 7, Apr. 2021, doi: 10.3390/app11073061.

[62] A. Hata and D. Wolf, “Road marking detection using LIDAR reflective intensity data and its application to vehicle localization,” in *2014 17th IEEE International Conference on Intelligent Transportation Systems, ITSC 2014*, Institute of Electrical and Electronics Engineers Inc., Nov. 2014, pp. 584–589. doi: 10.1109/ITSC.2014.6957753.

[63] M. Yang *et al.*, "Laser data based automatic recognition and maintenance of road markings from MLS system," *Opt Laser Technol*, vol. 107, pp. 192–203, Nov. 2018, doi: 10.1016/j.optlastec.2018.05.027.
[64] T. Strain, R. E. Wilson, and R. Littleworth, "Computer vision for rapid updating of the highway asset inventory," *Transp Res Rec*, vol. 2674, no. 9, pp. 245–255, Jul. 2020, doi: 10.1177/0361198120928348.
[65] L. Hazelhoff, R. O. H. Veld, I. Creusen, and P. H. N. De With, "Subsign detection and classification system for automated traffic-sign inventory systems," in *VISAPP 2014 - Proceedings of the 9th International Conference on Computer Vision Theory and Applications*, 2014. doi: 10.5220/0004654402620268.
[66] Y. Taki and E. Zemmouri, "An Overview of Real-Time Traffic Sign Detection and Classification," in *Lecture Notes in Networks and Systems*, Springer Science and Business Media Deutschland GmbH, 2021, pp. 344–356. doi: 10.1007/978-3-030-66840-2_26.
[67] D. KANG, X. CAI, X. GUO, J. ZHENG, and X. ZHOU, "An OpenCL Parallelized Traffic Sign Recognition," *DEStech Transactions on Computer Science and Engineering*, no. mmsta, 2018, doi: 10.12783/dtcse/mmsta2017/19710.
[68] T. Chen and S. Lu, "Accurate and Efficient Traffic Sign Detection Using Discriminative AdaBoost and Support Vector Regression," *IEEE Trans Veh Technol*, vol. 65, no. 6, pp. 4006–4015, Jun. 2016, doi: 10.1109/TVT.2015.2500275.
[69] W. Zhang, Q. Wang, H. Fan, and Y. Tang, "Contextual and Multi-Scale Feature Fusion Network for Traffic Sign Detection," in *10th IEEE International Conference on Cyber Technology in Automation, Control and Intelligent Systems, CYBER 2020*, 2020. doi: 10.1109/CYBER50695.2020.9279180.
[70] Z. Yao, X. Song, L. Zhao, and Y. Yin, "Real-time method for traffic sign detection and recognition based on YOLOv3-tiny with multiscale feature extraction," *Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering*, vol. 235, no. 7, pp. 1978–1991, Jun. 2021, doi: 10.1177/0954407020980559.
[71] A. Møgelmose, M. M. Trivedi, and T. B. Moeslund, "Vision-based traffic sign detection and analysis for intelligent driver assistance systems: Perspectives and survey," *IEEE Transactions on Intelligent Transportation Systems*, vol. 13, no. 4, pp. 1484–1497, 2012, doi: 10.1109/TITS.2012.2209421.
[72] B. Bousarhane and D. Bouzidi, "Convolutional Neural Networks for Traffic Signs Recognition," in *Communications in Computer and Information Science*, Springer Science and Business Media Deutschland GmbH, 2020, pp. 73–91. doi: 10.1007/978-3-030-61143-9_7.
[73] Y. Yuan, Z. Xiong, and Q. Wang, "An Incremental Framework for Video-Based Traffic Sign Detection, Tracking, and Recognition," *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 7, pp. 1918–1929, Jul. 2017, doi: 10.1109/TITS.2016.2614548.
[74] Z. Ouyang, J. Niu, Y. Liu, and M. Guizani, "Deep CNN-Based real-time traffic light detector for self-driving vehicles," *IEEE Trans Mob Comput*, vol. 19, no. 2, pp. 300–313, Feb. 2020, doi: 10.1109/TMC.2019.2892451.
[75] A. D. Pon, O. Adrienko, A. Harakeh, and S. L. Waslander, "A hierarchical deep architecture and mini-batch selection method for joint traffic sign and light detection," in *Proceedings - 2018 15th Conference on Computer and Robot Vision, CRV 2018*, Institute

of Electrical and Electronics Engineers Inc., Dec. 2018, pp. 102–109. doi: 10.1109/CRV.2018.00024.
[76] E. B. Nuakoh, “NuNet: A Deep Learning Approach for U.S. Traffic Sign Recognition,” 2019.
[77] Z. Zhang, L. Stenneth, R. Marappan, Z. Sebastian, and P. S. Yu, “Insert beyond the traffic sign recognition: Constructing an autopilot map for autonomous vehicles,” in *GIS: Proceedings of the ACM International Symposium on Advances in Geographic Information Systems*, Association for Computing Machinery, Nov. 2018, pp. 468–471. doi: 10.1145/3274895.3274951.
[78] L. T. Torres *et al.*, “Effortless Deep Training for Traffic Sign Detection Using Templates and Arbitrary Natural Images,” in *Proceedings of the International Joint Conference on Neural Networks*, 2019. doi: 10.1109/IJCNN.2019.8852086.
[79] S. B. Wali *et al.*, “Vision-based traffic sign detection and recognition systems: Current trends and challenges,” *Sensors (Switzerland)*, vol. 19, no. 9. MDPI AG, May 01, 2019. doi: 10.3390/s19092093.
[80] A. Ellahyani, I. el Jaafari, and S. Charfi, “Traffic sign detection for intelligent transportation systems: A survey,” in *E3S Web of Conferences*, EDP Sciences, Jan. 2021. doi: 10.1051/e3sconf/202122901006.

2.7 APPENDIX A. DIMENSIONS FOR INFORMATION EXTRACTION DURING SLR

Table 6. Dimensions for Information Extraction

Dimensions	**Description**
Review (Y/N)	Whether or not this paper is a review/summary/survey of previous studies.
Inventory Related (N/Y)	Whether or not this paper is relevant to the establishment, management or update of an inventory database.
Digitized Infrastructures	Types of TCI are digitized in this paper.
Study Area	Country, city, etc.
Data Source	Data provider or method for data collection.
Method (Model)	Methods or models utilized in digitizing TCI.
Sample Size	Size of sample dataset.
Accuracy	Accuracy, precision, and recall, etc.
Application for AV driving	How the methods/models introduced in this paper could be utilized on AV driving.
Application for inventory	How the methods/models introduced in this paper could be utilized on the establishment, management, or update of an inventory database.
Limitation	Potential limitations.

- Review (Y/N)—The existing SLR works would be very helpful for our study and could save a lot of time and labor. For each article we collected, we identified whether it was a review or not.
- Inventory Related (N/Y)—Not all research on digitizing TCI involves the TCI inventory, while the ones related to inventory would be significantly more helpful for our study. Therefore, for each study we identified whether it is related to inventory, and thus determined which papers should have higher priority in the following SLR.
- Digitized Infrastructure—If certain types of TCI had ever been successfully digitized in a previous study, we could refer to the methods or models that are utilized or developed in those studies. On the other hand, for the TCI that had hardly or never been digitized, we would have to develop new techniques to deal with them.
- Study Area—Geographical analyses usually take place in various places around the world and based on different local conditions; the research on the same topic could have a lot of differences on design, methods, or results. For digitizing the traffic signs, such differences could be dramatic since different countries or regions could have very different systems of TCI. Therefore we derived the study area of each study so that we could identify whether these methods or models could performs as well in United States.
- Data Type—The data utilized in analyses have different types, and for the studies on TCI digitization, the major types of data include imagery (video or image), GPS, and point cloud. Different types of data could play different roles in the analysis on TCI digitization, and by comparing the recorded data types and the research outputs, we could identify the relations behind them.
- Data Source Type—In an experiment, the researchers could have two choices of getting data: derive the data they want from existing databases, especially those public databases, or collect their own data and build a new dataset. Data from the existing database could be more universal and are available for every other researcher, but the data from their own dataset could be more flexible on data format and study area. The selection of data source types can somewhat reflect the research objectives and lead to different applications.
- Database—There are already a lot of established databases for TCI analysis. These databases could be utilized in the training, testing, and validating of various TCI digitizing methods and models. We recorded the databases for the analyses that use existing databases, and thereby we could identify which are the most frequently utilized databases.
- Data Collection—For the analyses that used their own dataset, there are different ways to collect data. Since our subsequent analysis on TCI digitization needs to cover a relatively large area, it is likely that we also need to build our own datasets. Therefore, learning how previous research collected their data could help us figure out the most efficient and effective way to do this.

- Method (Model)—Usually different studies have their own methodologies on solving even the same problem, but similarities could exist when the topics are close. The common utilization of certain methods or models might imply their high performance or capacity on TCI digitization. Therefore, we recorded the methods or models developed or utilized in each study so we could identify which methods/models are widely applied on TCI digitization or if there are similarities among different methods.
- Sample Size—Among the methods or models developed for TCI digitization, machine learning is widely used for detecting or recognizing the TCI. To achieve an ideal level of accuracy, a certain size of sample data is required. We recorded the sample size (including training sample size and testing sample size, if any) of each study to identify what size of sample data we would need in our research.
- Accuracy—One of the major standards to evaluate the performance of a TCI-digitization method or model is to compare its accuracy with other methods or models, and the accuracy could be assessed in different aspects, including precision, recall, and mean Average Precision (mAP), etc. We recorded the accuracy of the methods or models in each study so we could identify which methods have the highest performance on TCI digitization.
- Application for AV Driving—Even if the TCI digitization succeeded, there is a gap between the digitization and the application on AV driving. Not every, but some papers introduced how the experiment results in their studies could be utilized for AV driving. We recorded these applications to figure out how our TCI digitization could help the safety of AV driving technology.
- Application for Inventory—Similarly, there is also a gap between TCI digitization and TCI inventory applications. We recorded how each method or model could be utilized in the establishment, management, or update of TCI inventory database (if any), so that we could learn how to build such an inventory database in our analysis.
- Limitation—In many academic literatures, the author will briefly describe the possible issues or potential limitations of their research. We recorded these limitations (if any) so that on one hand we could know what the biggest challenges might be before we start our experiments, and on the other hand we could look for possible solutions from other studies.

CHAPTER 3. EVALUATING EXISTING SOLUTIONS, POTENTIAL RESOURCES, AND POSSIBLE LEGAL ISSUES FOR TCI DIGITIZATION

3.1 SUMMARY

In Task 3, the research team reviewed the state of practice for TCI digitization and identified potential market available data sources for building TCI inventories. In this task, the research team also identified the possible legal issues associated with the digitization, collection, and sharing of TCI data.

3.1.1 Objective

The objective of this task is four-fold:

1. Evaluate the state of practice for TCI digitization.
2. Examine and identify existing market available data sources and solutions.
3. Identify possible legal issues associated with the collection and sharing of TCI data.
4. Develop a conceptual framework based on the findings of the above three objectives for TCI digitization and data management.

3.1.2 Overview of Chapter 3

This document presents the results of Task 3, where the research team reviewed the existing solutions adopted by different transportation agencies for the TCI digitization through conducting a focus group discussion with invited stakeholders. The research team also identified the possible legal issues arising from the acquisition of data for TCI digitization as well as those arising out of its dissemination to road users, especially to AV companies.

The rest of this chapter is organized as follows:

- Section 3.2 presents the focus group discussion results regarding the existing solutions and challenges.
- Section 3.3 summarizes the selected market-available data sources for TCI.
- Section 3.4 discusses the legal issue review results.
- Section 3.5 shows the conceptual framework of the proposed solution for TCI digitization and data sharing.
- References are provided at the end.

3.2 FOCUS GROUP DISCUSSION FOR EXISTING SOLUTIONS OF TCI DIGITIZATION

3.2.1 Purpose

This sub-task aims to (1) understand the availability of TCI data in different entities (e.g., regional and local transportation agencies), (2) understand the existing solutions utilized for TCI data collection by these entities, (3) identify their limitations, and (4) determine the directions for further improvement.

3.2.2 Focus Group Discussion

Focus group discussion is a commonly used qualitative research method, which typically involves a small group of people (8 to 12) from similar backgrounds or experiences together to discuss a specific topic of interest. It is led by a moderator (interviewer) with a list of questions to conduct a loosely structured discussion.

On April 19, 2022, the research team convened a focus group of practitioners who are in the early development stages of digitizing TCI data and had them share valuable information learned from their experiences. The group members were from the Texas Department of Transportation (TxDOT) districts, MPOs, and cities in the major urban areas in Texas who have extensive expertise or are identified as the lead in intelligent transportation systems (ITS) and connected vehicle (CV)/AV applications in their own agencies. Table 7 lists the detailed information of the focus group members.

Table 7. Focus Group Discussion Participants.

Focus Group Discussion Members—Infrastructure Owner Operators (IOOs)		
Name	**Agency**	**Title**
Thomas Bamonte	North Central Texas Council of Governments (NCTCOG)	Senior Program Manager/Transportation
Natalie Bettger	NCTCOG	Senior Program Manager
Craig Burgan	TxDOT Dallas District	Traffic Systems Administrator
Ugonna Ughanze	TxDOT Houston District	Director of Transportation Operations
Veronica Davis	City of Houston	Director—Transportation and Drainage Operations
Khang Nguyen	City of Houston	Multimodal Safety & Design Branch
Brian Moen	City of Frisco	Assistant Director Transportation
Marc Jacobson	City of San Antonio	Transportation System Management

3.2.3 Discussion Questions

The focus group discussion was mainly based on the survey questions designed by the research team. The questions were aimed to dive into the mind of a member about TCI digitization. They specially covered TCI digitization, data inventory and management, legal issues, and existing gaps. The questions consisted of single choice questions, multiple choice questions, ranking questions, and open questions for discussion. These questions and the responses from the group will be discussed in more detail in this section.

3.2.4 Discussion Results

3.2.4.1 Question 1

"Q1. Do you have established datasets for traffic control infrastructures (e.g., intersections, traffic signals, signs, etc.) in your agency?"

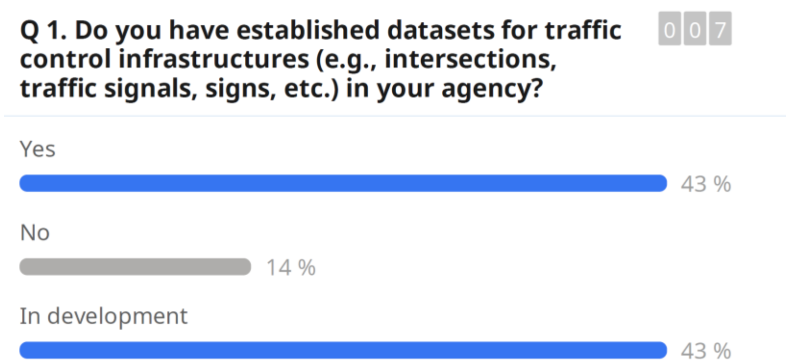


Figure 6. Results for Focus Group Discussion Question 1.

Most of the focus group members said their agencies have either established or are developing TCI datasets (86 percent). This clearly shows that this research project is timely and paves the path for safer and more efficient AV applications in Texas.

3.2.4.2 Question 2

"Q 2.1. What types of TCI data does your agency currently have? (Multiple choice)"

"Q 2.2. Any additional (not covered by Q 2.1) TCI data available in your agency?"

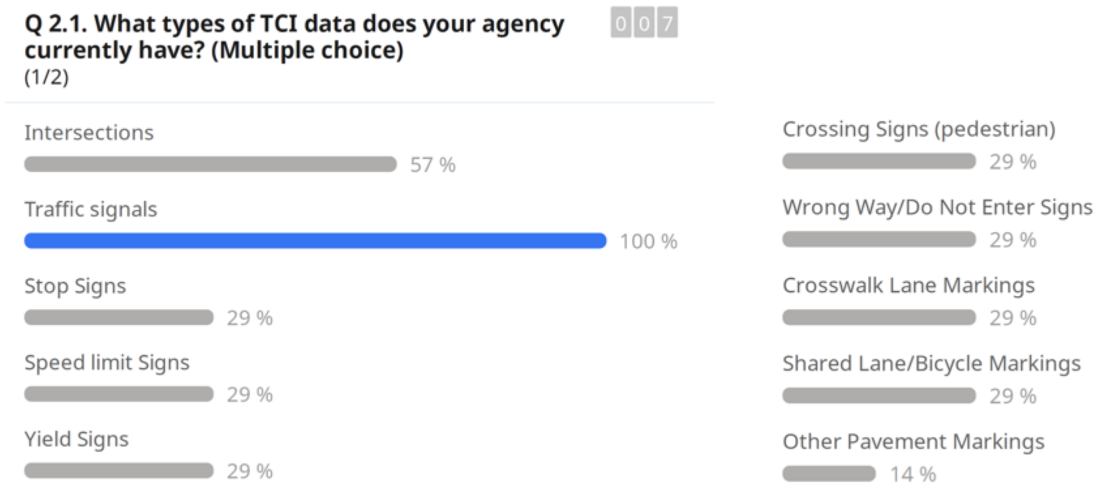


Figure 7. Results for Focus Group Discussion Question 2.1.

The focus group identified the types of TCI data listed in the question that are created by their agencies. Traffic signal data are the only type that all agencies currently have for their TCI data. In the next follow-up question (Q2.2), the group members also provided some additional TCI data types that are not covered by Q2.1: other ITS devices (cameras, Bluetooth, dynamic message signs), roundabouts, and speed feedback signs. One member specially indicated that his agency has tried to inventory all traffic control signs (not just a limited subset).

3.2.4.3 Question 3

"Q 3. How do you rank the importance or prioritize the digitization tasks for the following TCIs? (1 represents the most important, 9 represents the least important)"

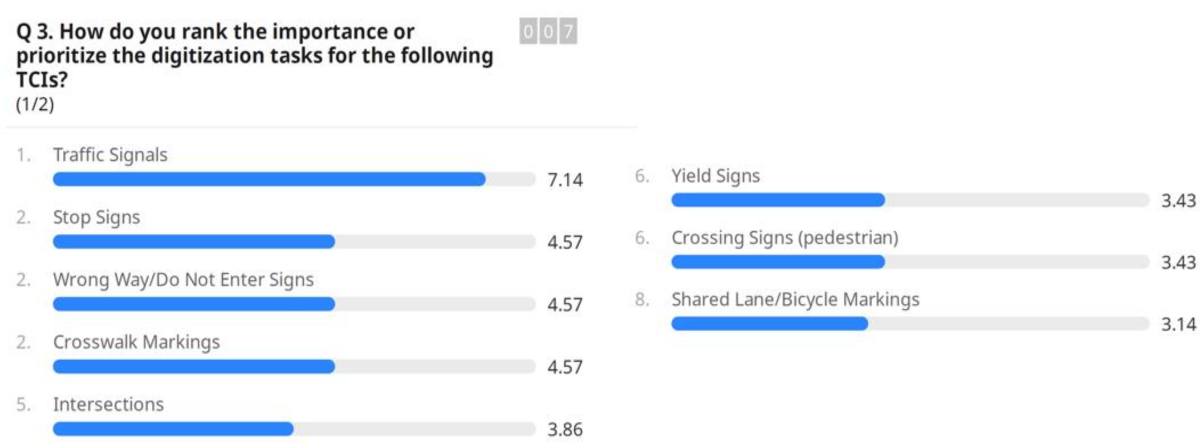


Figure 8. Results for Focus Group Discussion Question 3.

For this question, the group members ranked the importance or prioritized the digitization tasks for the TCIs listed by the research team. Traffic signals have been identified as the highest priority, followed by stop signs, wrong way driving signs, and crosswalk markings. The feedback from the group will greatly help the research team to develop and prioritize their digitization work in the next tasks.

3.2.4.4 Question 4

"Q 4.1. How does your agency obtain the digitized TCI data? (Multiple choice)"

"Q 4.2. If you chose 'Other' in the previous question, what other methods have you used to obtain the TCI data?"

"Q 4.3. If your agency had used the third-party providers' data, which data provider have you worked with? What TCI data have you received? And how are you satisfied with their data product (data quality & cost: 1-5, very bad, bad, fine, good, very good)?"

Figure 9. Results for Focus Group Discussion Question 4.1.

Currently, the focus group members said most of their agencies are collecting digitized TCI data from the fieldwork. About one-third of them use vehicle-based inspection and asset data collection methods (38 percent). This also shows the timeliness of this research project, which aims to develop an automated method to collect TCI data with AI/ML technologies and third-party traffic datasets.

In Question 4.2, some group members said they use other methods such as Google Maps and Google Street View. One member mentioned that his agency used a company to get pavement condition ratings. They also occasionally have used LiDAR to pin down the location of their assets.

In Question 4.3 (an open question), the members were also asked if they had used any third-party provider's data and if they were satisfied with these kinds of data products in terms of cost and data quality. A member from NCTCOG said that their agency does a regional procurement for pavement monitoring (Pavement Management Services—TXShare). Local governments then order services from the menu that NCTCOG has procured. NCTCOG may do more of this type of regional procurement to get one or more situational awareness app (e.g., Nexar, Blyncsy, Roadbotics) so its regional partners can monitor roadway conditions/TCI more effectively.

3.2.4.5 Question 5

"Q 5.1. How often does your agency update the dataset?"

"Q 5.2. Does your agency have a data management plan for your TCI assets?"

"Q 5.3. What strategies/solutions have you adopted for data quality control, assessment, and enhancement?"

Figure 10. Results for Focus Group Discussion Question 5.1.

In Question 5.1, the members were asked about the frequency of updating their TCI datasets. About two-thirds of them update the dataset more than once per year. Some agencies update the dataset once per year. Some agencies even update their datasets up to every three years.

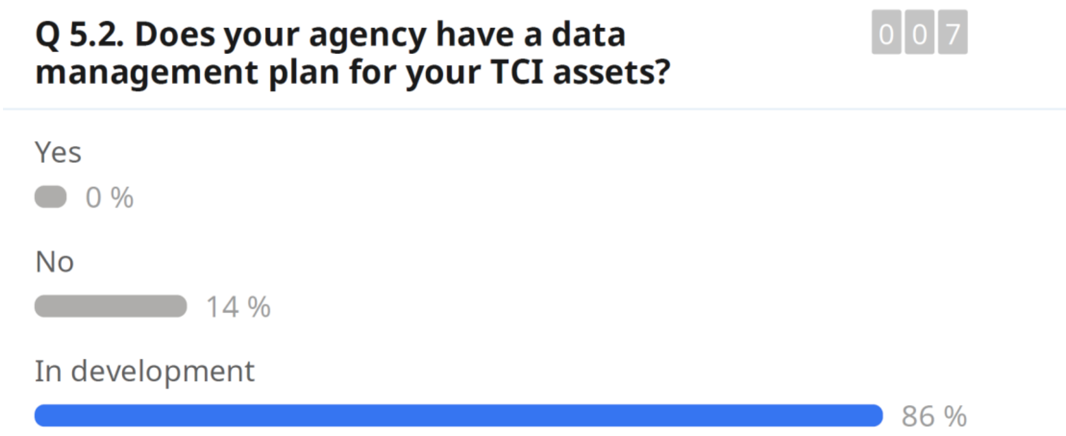


Figure 11. Results for Focus Group Discussion Question 5.2.

In Question 5.2, most of the members said a data management plan is under development for their TCI datasets. One member said their agency just simply does not have one.

In Question 5.3, the group members were also asked about what strategies/solutions they have adopted for data quality control, assessment, and enhancement when a data management plan is not in place. The City of Houston said its GIS group is developing a simple interface for technical and field staff to update and correct data in the field. The City of Frisco said that it is working with its business analysts in the IT department to design the workflow and architecture that can be maintained and is also compliant with its GIS system and work order system. The staff is trained on how to use the system to collect the data.

3.2.4.6 Question 6

"Q 6. In your current practices, how do you build up a digitized inventory for your traffic control infrastructures or assets?"

Question 6 is also an open question in which the group members were asked about how they build up a digitized inventory for their TCI assets. NCTCOG mentioned that they are using a consultant to build up an inventory of traffic signals in the Dallas-Fort Worth (DFW) area at the direction of NCTCOG staff.

3.2.4.7 Question 7

"Q 7. In your opinion, are there any identified gaps in the TCIs data? Can you specify any challenges in collecting and digitizing your asset inventory data? (Please add your name and affiliation to your answers)."

Question 7 asks if the group members have identified any gaps in the TCI data, especially for collecting and digitizing the data. The major gaps identified by the group are summarized as follows:

- By relying exclusively on asking jurisdictions for information about their traffic signals, the agency is proceeding more slowly than it might if it also uses crowdsourced video data (e.g., Blyncsy, Nexar) as a baseline data source to be supplemented via inquiries to local jurisdictions.
- It is inevitable that regions will want to digitize not just type/location of TCI but also make that information—and the information conveyed by the TCI—readily accessible to travel navigation services, original equipment manufacturers (OEMs), and automated vehicle developers. The agency should be planning for a future where physical signage decreases in importance—perhaps fades away—in favor of digitized TCI information, just like how physical maps have disappeared in favor of Google/Apple Maps.
- TCI will be digitized and consumed electronically by our mobility devices and our assistive devices (e.g., smartphone, smart AR glasses). This transition will be gradual, but it will happen in the next 30 years. The real game/gain is digitizing and sharing TCI data with the developer community. Many departments of transportation (DOTs) are still stuck in the 1990s mindset where they run a closed system of data collection and analysis, sharing what information they see fit with the public. That mindset would stop a TCI at the inventory stage. Travel navigation services and OEMs now run circles around DOTs in terms of what they know about roadway conditions. Firms like Wejo hint at the wealth of information OEMs are harvesting from their vehicles, and this information set will only grow with more and better sensors. Travel navigation services like Waze also know more about roadway conditions and incidents than DOTs.
- AVs are mapping roadways and tracking changes in much more detail and timeliness than DOTs. DOTs thus need to recognize that they are 'junior' partners in the roadway data game. Rather than inventorying and then buying and maintaining more roadway equipment to feed their closed ITS platforms, DOTs should focus on sharing what information they have about pertinent roadway features (e.g., speed limits, Signal Phase and Timing (SPAT) data) and feed this information to interested third parties to supplement the information that those third parties have compiled already. They can use processed information from connected vehicles to improve their situational awareness of roadway conditions and their understanding of the behavior of vehicles. In other words, do not confuse an inventory of TCI with effective utilization of TCI to improve roadway operations and safety.
- It is hard to have resources/labor to collect and digitize the data for an agency. Pavement markings are also difficult to collect.

3.2.4.8 Questions 8 and 9

"Q 8. Have you encountered any legal issues or obstacles in obtaining the data?"

"Q 9. Can you discuss any contractual provisions you've agreed to regarding liability or indemnification when obtaining third party data?"

Because the efforts led by these participants are still in the developmental stages, no one had encountered legal issues on these specific projects. However, some had encountered contractual questions or issues related to other projects. Most of these experiences had to do with use rights and restrictions that were applied to data the focus group members encountered. For example, agencies were prohibited from sharing third party data with partner research organizations without a sublicense agreement in place between the agency and the partner. The agencies also had to contend with other prohibitions related to data, including digitization, publication, sharing of proprietary work product that are costly to produce, and sharing data with law enforcement for purposes outside the scope of the contract (e.g., enforcing traffic laws).

Other issues that came up in the discussion related to agencies' internal use of data included:

- Whether knowledge of an asset, through digitizing TCI data, creates a legal duty or increased exposure to liability for the agency with increased knowledge about the asset.
- Whether possessing data exposes an agency to liability, even if the data are intended for internal purposes only.
- Whether notice requirements regarding infrastructure defects change at all if data about that defect are transmitted by an AV.

All of these issues are consistent with the questions and findings that researchers identified in the review conducted for this task.

3.2.4.9 Question 10

"Q 10. Please list your agency's name and contact info if you are willing to contribute some 'ground truth' data for our result comparison."

Question 10 asked if any group member would like to contribute some "ground truth" data to the researchers. NCTCOG and the City of Houston wanted to help with these datasets. This question also concluded the group discussion.

3.3 MARKET AVAILABLE DATA SOURCES FOR TCI

3.3.1 Purpose

With the advent of data acquisition techniques, various third-party TCI datasets are being generated based on various data sources and becoming commercially available. This section introduces the promising third-party data providers for TCI data and their data products that the research team selected. These data providers were identified from the Performing Agency's previous relevant projects or recommend by the stakeholders though the focus group discussion.

3.3.2 Third-Party TCI Data Providers

Through the discussion with different stakeholders, four companies were identified that provide well-developed data products based on their data sources and techniques for TCIs, including Otonomo, Mobileye, Nexar, and HERE.

3.3.2.1 Otonomo

- Data provider introduction: Otonomo is a vehicle data provider founded in 2015 that currently fuels a data ecosystem of 16 OEMs, fleets, and 100+ service providers. Their platform collects more than four billion data points from over 40 million CVs globally on a daily basis.
- Data source: The road sign data are a data product of Otonomo generated by the vehicle sensors. They can generate the road sign data from over 500,000 passenger vehicles in the United States and Canada, which are equipped with advanced driver assistance systems (ADAS).
- TCI data types: The digitized traffic signs include speed limit sign, yield sign, stop sign, passing restriction sign, mandatory passing sign, and no entry sign, among others. All the signs include location information with four decimal places of precision.

3.3.2.2 Mobileye

- Data provider introduction: Mobileye, an intel company, was founded in 1999. It has become a global leading company in the development of computer vision and machine learning, data analysis, localization, and mapping for ADAS and AV driving solutions.
- Data source: The Road Asset Network is a data product of Mobileye data services. Mobileye's technology uses a single camera to scan the road ahead. Mobileye's EyeQ chip and algorithms have been trained to identify, tag, and classify a variety of road assets as equipped vehicles (passenger and fleet vehicles) driving on their route. More than 30 vehicle manufacturers are using EyeQ to support the complex vision tasks of ADAS. The road asset data captured by the vehicles are processed and turned into GIS data layers, which are updated at a high refresh rate.
- TCI data types: Mobileye's Road Asset Network captures a wide variety of TCIs including pavement markings, traffic lights, and traffic signs. The pavement markings can be further classified into road markings, lane markings, crosswalks, and stop lines. Asset data are created through an aggregation of multiple detections of the same object as vehicles travel the roadway to increase both location and detection accuracy.

The detailed TCI types are listed below:

- Road markings: directional arrows, bicycle lane, diamond, selected words (e.g., Bus, Only, Stop), and speed limits painted on the pavement.
- Lane markings: solid white line, double solid white line, broken/dash white line, double broken/dash white line, and deceleration line.
- Crosswalks: solid crossing, dashed crossing, and zebra crossing.
- Stop lines: solid stop line, dashed stop line, double solid stop line, double dashed stop line, triangular stop line, and road bump.
- Traffic lights: vertical 2, 3, 4, 5 signal heads, horizontal 3, 4, 5 signal heads, pedestrian (horizontal and vertical), and others.
- Traffic signs: detection of traffic signs is primarily based on the Vienna Convention on Road Signs, which includes stop signs, speed limit signs, warning signs, two-way road signs, bicycle crossing signs, and winding signs, among others.

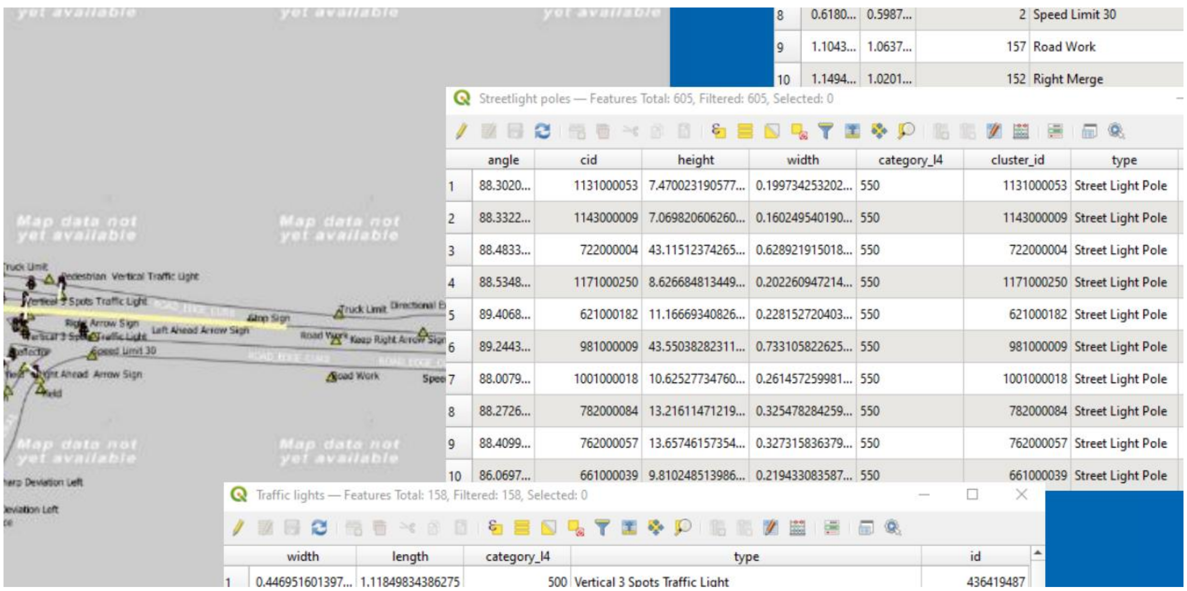


Figure 12. Screenshot of Mobileye TCI Data Sample.

3.3.2.3 Nexar

- Data provider introduction: Nexar is the leading dash cam company and was founded in 2015. Nexar-powered dash cams enable new vision-based applications for better driving. It has developed a portfolio of vision-based data services through its connected dash cam platform for both public and private sector partners to make roadways safer and more efficient.
- Data source: Nexar collects contribute images from its dash cam users to anonymously capture the streetscapes from their drives. Then, the collection of images is processed through Nexar's AI-powered algorithms to recognize and locate road assets automatically

and accurately. Nexar users cover more than 90 percent of the roads in the top 100 largest U.S. cities, which generate billions of new images monthly.

- TCI data types: The Road Inventory data provided by Nexar can detect, monitor, and locate traffic signs and signals, including speed limit signs, yield signs, stop signs, passing restriction signs, no entry signs, one way signs, pedestrian crossing signs, and school zone signs, among others. It also could monitor the state/condition of road assets.

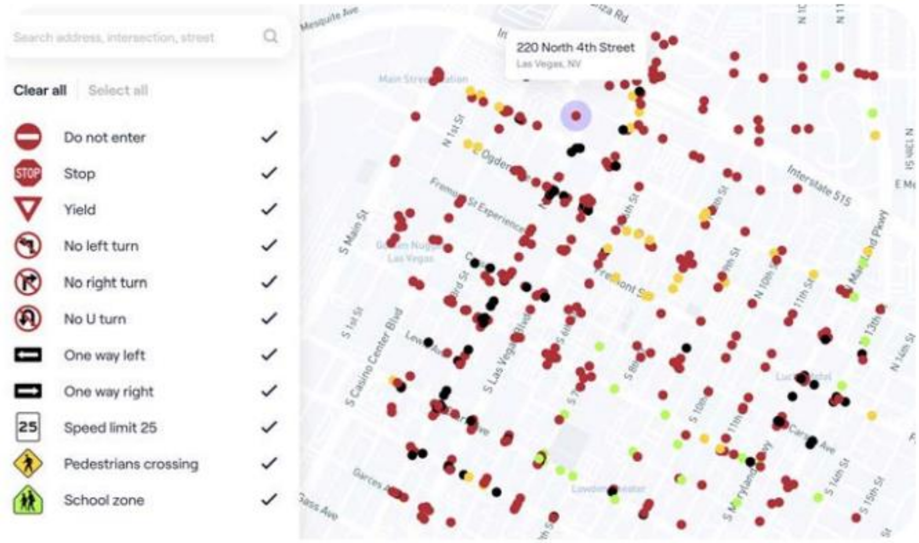


Figure 13. Example of Nexar TCI Data Platform.

3.3.2.4 HERE

- Data provider introduction: HERE Technologies is a multinational group dealing with mapping, location data, and related automotive services and was founded in 1985. HERE captures and digitizes location content such as road networks and buildings and sells its digitized data products.
- Data source: HERE utilizes multiple data sources, such as HERE map, LiDAR data, and street level imagery, to generate automotive-grade quality map data and refresh them regularly.
- TCI data types: HERE data layers contain the traffic sign layer, which represents signs along the road network used to inform the driver of specific road situations. The traffic sign layer includes the location of each traffic sign, the type of sign, and the sign category. The sign types include warning signs, priority signs, stop signs, yield signs, school zone signs, and pedestrian crossing signs, among others.

3.3.3 Companies for Building TCI Inventory

Instead of directly providing commercially available datasets for TCIs, some companies are providing products and services to help with the creation of solutions and systems for digitizing TCIs. Three companies were identified that provide solutions for building TCI inventory, including RoadBotics, Blyncsy, and VAISALA.

3.3.3.1 RoadBotics

RoadBotics was spun out of Pittsburgh's Carnegie Mellon University (CMU) Robotics Institute in 2016. RoadBotics transforms visual infrastructure data into meaningful maps using AI techniques.

RoadBotics' new platform, AgileMapper, could help with the development of asset inventories and assess the conditions of road assets. The customers can upload their existing visual infrastructure asset data collected through smartphone, GoPro, camera, or drone to the AgileMapper platform. The images appear on an interactive map and are geo-tagged, time-stamped, and user-referenced.

3.3.3.2 Blyncsy

Blyncsy is a movement and data intelligent company founded in 2014 and headquartered in Salt Lake City, Utah. Through the power of big data and location analytics, Blyncsy helps its customers understand how connected devices are moving throughout an environment.

Blyncsy' product, Payver, delivers instant insights by using AI to analyze crowd-sourced video data obtained through connected dash cams and smartphones. The AI algorithm extracts road assets information from the crowd-sourced video data, which is then hosted, aggregated, and shared through its cloud-based system, Payver.io.

3.3.3.3 VAISALA

VAISALA is a Finnish company that develops, manufactures, and markets products and services for environmental and industrial measurement. It was founded in 1931 and is currently headquartered in Vantaa, Finland.

VAISALA provides road asset management solutions through its RoadAI product. The RoadAI deployment uses several RoadAI-equipped mobile phones to capture real-time streetscape data. Through computer vision, RoadAI recognizes road assets, such as traffic signs, barriers, and bollards, adds them to the inventory databases, and records their conditions.

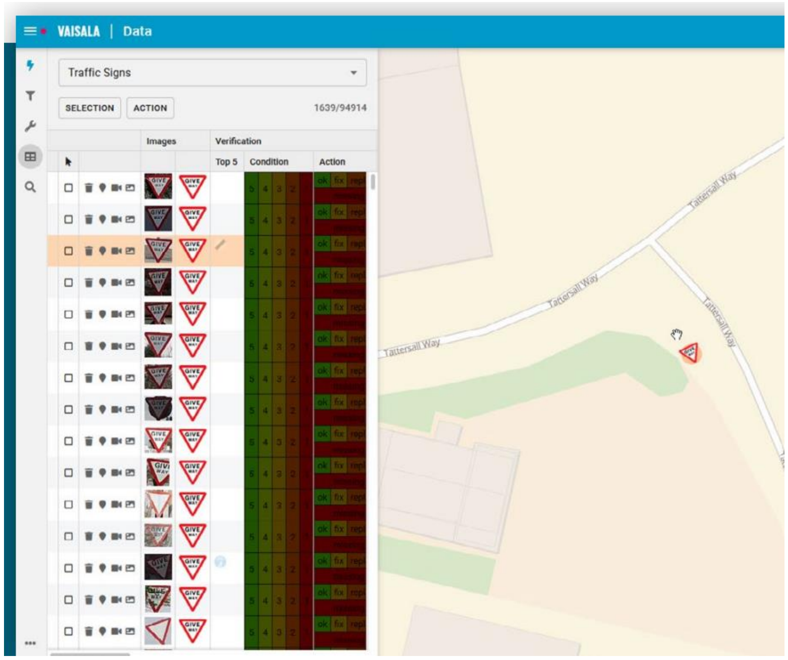


Figure 14. Example of VAISALA Inventory Data Platform.

3.3.4 Street View Imagery

Street view imagery (SVI) is an excellent resource that provides detailed urban street views as human vision and comprehensively depicts transportation infrastructures. Some leading SVI providers like Google, Microsoft, and Mapillary are serving millions of users globally by providing high-quality panoramic imagery captured at street level. Many transportation agencies conduct "virtual" audits of traffic infrastructures based on the SVI data instead of conducting field audits, which is more time- and labor-efficient.

Meanwhile, with the emergence of computer vision techniques and advanced machine learning methods, various geographic features can be automatically detected from SVIs, providing unprecedented opportunities to efficiently conduct a large-scale audit for TCIs such as traffic signs and traffic lights. Figure 15 illustrates an example of traffic lights detection using SVI images conducted by the authors.

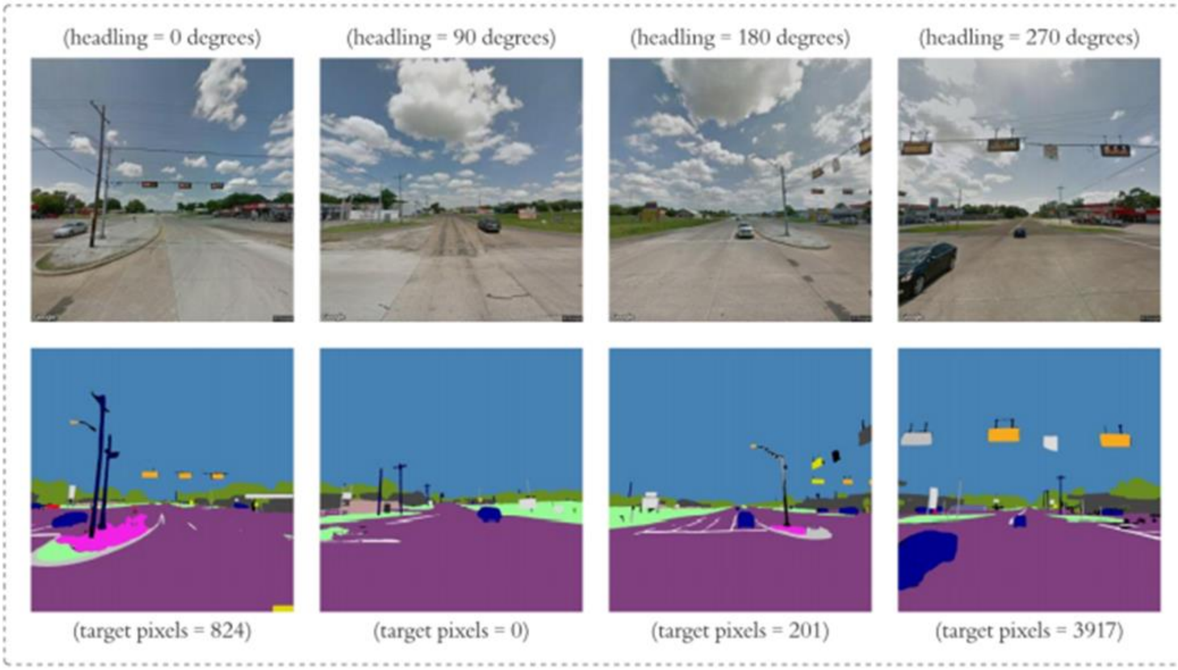


Figure 15. Detection and Digitization of Traffic Lights from SVI Images.

3.4 REVIEW OF LEGAL ISSUES ASSOCIATED WITH TCI DATA COLLECTION AND DISSEMINATION

3.4.1 Purpose

The purpose of this section is to review the results of the Task 2 Report (Literature Review) and identify possible legal issues associated with the digitization of TCI. Legal issues may arise from the acquisition of data for use in TCI digitization, as well as its dissemination to road users when TCI systems are implemented. These legal issues will vary depending on the type of data at issue (e.g., location data, photographic data, vehicle telematics), how the data are stored and processed, and how an agency plans to incorporate them into a publicly available database (e.g., data map, privacy impact analysis) for use by third parties and road users. In all these cases, the collection and sharing of TCI data poses legal questions about data protection, data privacy, data ownership, and agency liability, among others. Thus, whether collected directly by transportation agencies or purchased from third parties, the processes for acquiring, processing, and distributing TCI data may need to be reviewed to mitigate potential liabilities for transportation agencies, protect proprietary interests of private data owners that comprise trade secrets, and protect the privacy interests of those whose personally identifying data may be captured to create TCI.

3.4.2 Background

The Task 2 Report (Literature Review) provides that digitizing TCI data have the potential to greatly contribute to a greater efficiency in asset management and accommodate the rise of AVs but is a nascent practice. By moving away from manual digitization of TCIs (through field studies) to use of emerging data collection and processing innovations, transportation agencies can save time and labor associated with their TAM systems. TAM systems are data-driven decision support systems that use asset inventory data (e.g., asset type, location, condition, and dimensions) to determine how to best deploy road assets (e.g., traffic signs and traffic signals) and maintain existing infrastructure. TCIs "provide information about the current state of the road, restrictions, prohibitions, warnings, and other helpful information for guiding the driving," but "compared to other road assets (e.g., pavement and bridges), the [advanced] digitization and maintenance of TCI inventory data are less explored" [1].

Advanced digitization of TCI data could also enhance self-driving and safe AV operations in the future by allowing AVs to pre-compute roadway information before beginning the driving task. The literature review notes that "creating high-quality [road] maps with TCIs precisely recorded can significantly promote self-driving techniques and AV operation safety." Specifically, the maps "could allow AVs to pre-compute things before driving, which can solve some autonomy problems or add redundancy to some autonomy functions." This might include supporting the AV's ability to "recognize and locate static TCIs offline and in a highly accurate manner" [2].

The legal issues identified in this document will cover those that arise from the processes that this research project will simulate (i.e., creation of a TCI dataset), as well as those that could occur at later stages but not are included in this project (i.e., dissemination to AVs, maintenance of TCI data).

3.4.3 Research Methodology

For this portion of Task 3, the research team first met to describe and map out the process of digitizing TCI data (Figure 16). This map was used to clarify the process at a high level and identify points in the process where legal issues could arise.

Next, the research team reviewed the Task 2 report to gain a more detailed knowledge of the TCI digitization process. Finally, researchers conducted a very general scan of the literature to provide a snapshot of legal issues and best practices used to address them.

To assist in this last step, the research team utilized databases (e.g., TRID, Lexis Nexis, and Westlaw) to determine whether and how the identified legal issues have been encountered and how they have been managed or mitigated. The team also looked at Texas state law to frame legal implications of the proposed research activities.

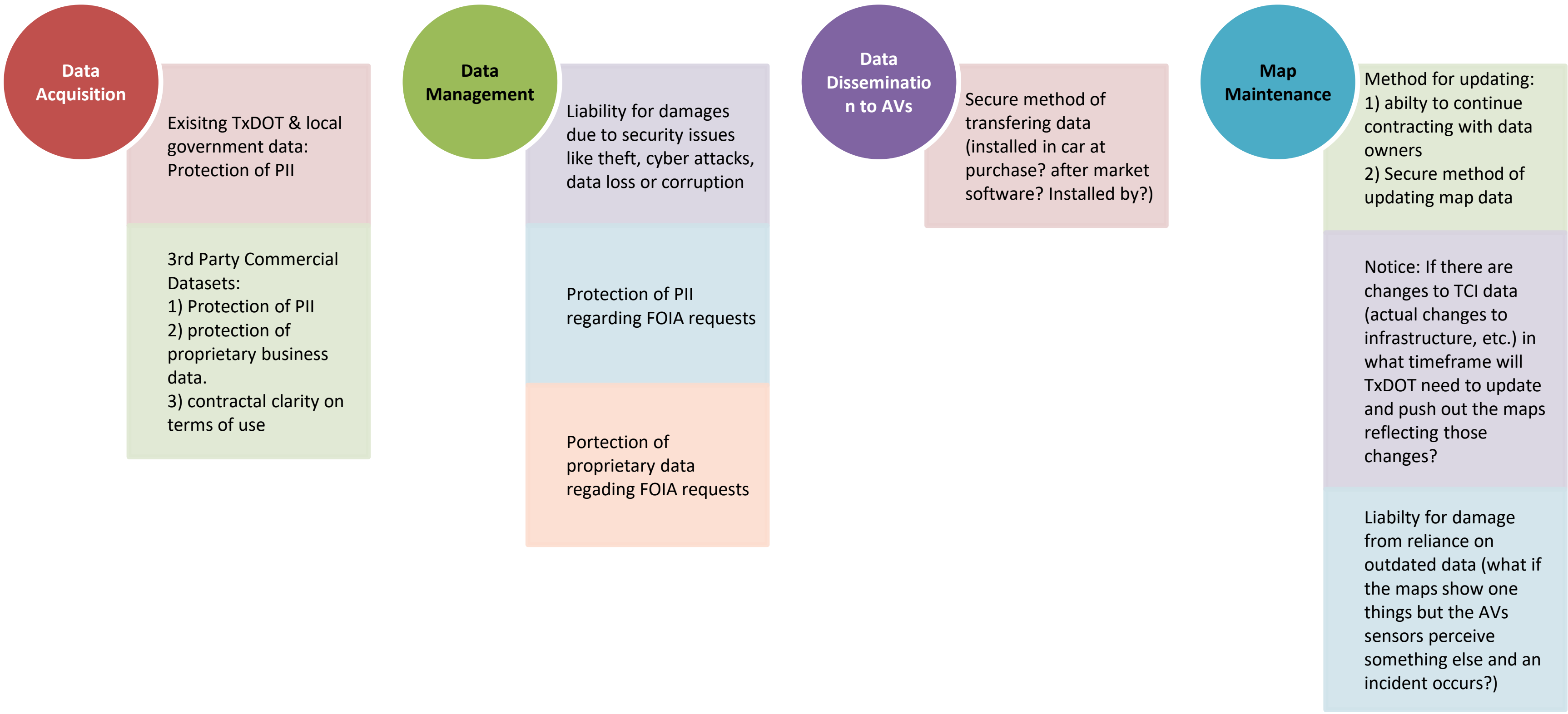


Figure 16. TCI Data Digitization Workflow.

3.4.4 TCI Data Digitization Workflow

3.4.4.1 TCI Data Acquisition

Data used to create the TCI dataset will be collected from three sources:

- **Existing TxDOT or local government data**—Consist of inventory data showing the location of TCI. These are primarily geographic data in comma-separated values (CSV) and shape file formats. These datasets are collected manually or by using vehicular equipment.
- **Third-party commercially available data**—Consist of inventory data similar to those owned by TxDOT, as well as raw photographic data (e.g., street view imagery) and connected vehicle trajectory data. The inventory data are collected differently depending on the company providing the data but can include digitized photographic street view images. Regardless of collection method, the data are delivered to TxDOT in the form of CSV files. Trajectory data from connected vehicles are processed before delivery to TxDOT so that personally identifiable information (PII) about private individuals is protected. Like inventory data, trajectory data will be delivered to TxDOT as CSV files consisting of a series of points that describe vehicle trajectory and accessed through a license, which will be renewed each year. Third-party data will be purchased from a data aggregator who will send the data to a cloud storage service. For the research simulation project, the contractual relationship will be between the Texas A&M Transportation Institute and the aggregator. Should the project move forward to implementation, TxDOT would contract directly with the aggregator.
- **Raw data**—Includes street view imagery and will function as a back-up option to complement existing datasets with imagery. These will be delivered to TxDOT as photographic files. Their content will consist of public and exterior spaces. These will be acquired and paid for on a one-time, rather than subscription, basis.

3.4.4.2 TCI Data Management

Once acquired, TCI data must be stored and managed. As part of this project, the research team developed a data management plan, which will provide guidance about data protection, storage, and access. For the simulation project, only researchers and administrators will have access to the data. Should the project be implemented, the datasets may be publicly accessible. The standards for both the short- and long-term data management plan should acknowledge and accommodate the legal issues arising from data management.

3.4.4.3 TCI Data Dissemination to AVs

The ultimate product of this research is a TCI dataset rather than a digital map installed in and used by AVs. Therefore, this research project will not simulate the dissemination of the data to

AVs. The research will compare how AVs drive with and without the data but will not address how the AV captures the data. However, a discussion of the legal issues that could arise out of disseminating the data is included in the legal analysis below.

3.4.4.4 TCI Digital Map Maintenance

This research produced a TCI dataset rather than a digital map for AVs and, therefore, will not simulate the maintenance of the data. However, a discussion of the legal issues that could arise out of maintaining the data is included in the legal analysis below.

3.4.5 Potential Legal Issues from TCI Data Workflow

The collection, storage, sharing, and maintenance of TCI data raises legal questions about data protection, data privacy, data ownership, and agency liability, among others. Specific legal questions vary depending on the type of data at issue (e.g., location data, photographic data, vehicle telematics), how the data are stored and processed, and how an agency plans to share the data with others (i.e., by either selling or incorporating the data into a free, open, publicly available database for use by road users or other third parties).

3.4.5.1 Data Protection

Kockelman et al. note that, “When vehicles are controlled by computers and connected wirelessly, like other cyber-physical systems, they are vulnerable to attacks, including hacking, and GPS spoofing.” At the same time, “a large amount of data are generated and collected, through onboard sensors. These data contain location information that could be sensitive, e.g., where a car was last parked, and distances traveled as well as time and speed” [3]. Such criminal hacks or negligent breaches to these data can “result in liability for agencies, operators, and OEMs [4].”

3.4.5.2 Data Privacy

Data breaches, however, are not the only legal risk from TCI data. Issues of data privacy may also arise in the process of acquiring, managing, and disseminating TCI data. For example, photographic data might contain recognizable images of a person’s face and could, if not properly scrubbed or protected, result in an invasion of that person’s individual privacy. Location data and vehicle data, even if anonymized, could be layered with other datasets, resulting in the re-identification of someone’s personal data, which is also an invasion of privacy. These invasions of privacy give rise to tort liability from disclosure or widespread dissemination of confidential information objectionable to an average person. In addition, if someone’s privacy is invaded because of a data breach or leak in personal data possessed or distributed by TxDOT to AVs and that person suffers injury or death as a result of that breach or leak, TxDOT could be determined liable for such injury or death.

Personal private data in the possession of a governmental entity can, if not sufficiently anonymized, be inadvertently disclosed in response to a request under a state's Privacy Impact Assessment (PIA). There may also be a risk that personal private data could be used by law enforcement in a way that runs afoul of the Fourth Amendment's protections against unconstitutional searches and seizures [5]. Finally, private data owners may have a proprietary interest in data that comprise trade secrets, the disclosure of which could compromise the private party's effectiveness or decrease their ability to compete in the market.

3.4.5.3 Liability from Digitized TCI Maps

The path that a digitized map takes to reach an AV could introduce legal issues, as would the method and regularity of how and when the TCI maps would be updated. Using TCI data to produce a product for public consumption, packaged for dissemination to AVs, could raise new legal issues of product liability. For example, if TxDOT produces a TCI map for installation in a vehicle, it may effectively make the agency a designer or manufacturer of a product and expose the agency to tort liability for product-related injuries. Under product-related injury laws, TxDOT would be held strictly liable if it is considered a "merchant" that routinely deals in TCI maps of this type, the TCI maps were defective (i.e., differs from its intended design, lacked warnings, and/or could have been designed to be less dangerous), the TCI maps were not altered, and the injured party was an intended user making an intended use of the TCI map.

Even if the TCI maps are installed after-market by customers, TxDOT may be held responsible for providing proper instructions or warnings about the installation. Similarly, legal issues may arise from updates to the TCI maps. If TxDOT implements updates to the TCI maps they produce, implementing the updates remotely using automatic update notifications, it may be responsible for ensuring that it or a third party pushing out those notifications use secure networks for file upload and download.

Legal liabilities may also arise if a TCI map does not properly update or the updated map is not consistent with the roadway conditions on the ground, leading to personal or property damage. If the TCI map updates fail to reach a particular AV or the updates are not consistent with current roadway conditions and a crash occurs because the map data were inconsistent with conditions that the AV perceived, TxDOT could also be found liable. This liability may be similar to those involved with traffic signal data, which some state DOTs provide to third-party automobile manufacturers for a variety of applications, including in-vehicle signal countdowns for drivers. Signal timing is periodically updated, and the signal data, if not properly conveyed to users without appropriate warnings and notifications, leave these agencies legally exposed.

These potential scenarios suggest that the terms of that liability need to be revisited and analyzed for any gaps or conflicts regarding the digitization of TCIs and their dissemination to AVs.

3.4.6 Legal Analysis and Mitigation Measures

Having identified potential legal questions related to the collection, storage, sharing, and maintenance of TCI data, this section provides an overview of the legal principles and theories triggered by issues of data privacy, proprietary data and trade secrets, and Texas state agency tort liability. These legal principles and theories are then used to identify potential mitigation measures to address the liabilities.

3.4.6.1 Privacy

Regardless of whether data are collected directly by transportation departments or purchased from third-party data owners, the processes for acquiring and processing those data must be reviewed and a data privacy model must be put in place [6]. These actions will minimize liability for the transportation agency, protect the privacy interests of those whose PII may be captured in the aggregated data, and protect the proprietary interests of those whose technology and business systems were used to manufacture and generate the TCI and TCI data.

Texas state law has codified privacy protections. In fact, in the most recent legislative session, no fewer than nine privacy bills were introduced [7], of which three were enacted:

- **HB 3746**—This bill increases transparency in reporting data breaches on the attorney general's online data breach portal, requiring the website to provide a listing of notifications received by the Office of the Attorney General of data security breaches and the number of affected residents that have been sent a disclosure of the breach [8].
- **SB 15/HB 3471**—Also known as the Texas Consumer Privacy Act Phase I, this bill "restricts disclosure of personal information to essential government agencies and forbids personal information from redisclosure or resale to private entities such as marketing and technology companies." [9]
- **SB 475**—Designed to further improve cybersecurity standards and data management, this bill "amends current law relating to state agency and local government information management and security, including establishment of the state risk and authorization management program." [10]

TxDOT's management of and conveyance of any data it acquires for TCI digitization, either by resale or by making them available free of charge, will, in part, be governed by these recent laws. TCI digitization will also be influenced by laws governing privacy torts. In general, there are four types of privacy torts:

- **Intrusion**—requires a defendant's conduct (invading the plaintiff's private affairs or concerns) to be intentional or reckless, without lawful justification, and considered as highly offensive and causing distress, humiliation, or anguish by a reasonable person.

- **Appropriation**—requires a defendant to make commercial use of the plaintiff's name or image without the plaintiff's consent.
- **Disclosure**—requires the widespread dissemination of confidential information considered highly offensive by a reasonable person without the plaintiff's consent, the publication of which is not of a legitimate concern to the public.
- **Placing a person in a false light**—widespread dissemination of a major falsehood about the plaintiff, which the defendant had knowledge of or acted in reckless disregard as to its falsity, that would be objectionable to a reasonable person. [11]

Of these privacy torts, TCI digitization would potentially run the risk of intrusion or appropriation, depending on the intent, justification, and use of the data. Affirmative defenses for privacy torts include consent. In Texas, sovereign immunity could also apply. This concept will be discussed in the "Tort Liability for Texas Governmental Units" section below.

3.4.6.2 Proprietary Data and Trade Secrets

A state agency's use of acquired data from a vendor will be governed in part by torts law and contract law, in addition to exemptions to a state's PIA. Laws governing tort liability will be triggered if the agency discloses vehicle data containing proprietary product design data or trade secret information. The disclosing party could be held liable for the tort of appropriation of trade secrets if they used the information for the value associated with it, the information could be identified from the disclosure, and the disclosing party benefited from or was advantaged by the disclosure.

The use and distribution of TCI data could also be affected by contract law and the rights that may be enumerated in agreements with data providers that transfer rights or provide a license to use the data. State DOTs, which have traditionally gathered and used their own data, are relatively new to licensing data for their transportation-related purposes. Agencies interested in sharing such data from third parties with other external stakeholders, regardless of any fee or charge, must refer to purchase or licensing documents. In fact, the National Academies of Sciences, Engineering, and Medicine provides that "the extent to which agencies can share data with external stakeholders is dictated by use permissions outlined in the use agreement." [12]

Protection of third parties' proprietary data and trade secrets can also be determined by a state's PIA, which, in some cases, exempt such information from disclosure. The Texas Government Code exempts trade secrets and certain financial or commercial information "for which it is demonstrated based on specific factual evidence that disclosure would cause substantial competitive harm to the person from whom the information was obtained." [13] Therefore, it may be legally permissible to withhold proprietary data from disclosure as a result of an open records request.

3.4.6.3 Tort Liability for Texas Governmental Units

Tort liability may potentially arise from invasion of privacy, as described above, as well as negligent handling of TCI data. A claim of negligence contains four elements:

- **Duty of care**—The defendant has a duty to conform to a specific standard of conduct for the protection of the plaintiff against an unreasonable risk of injury.
- **Breach**—The defendant's conduct falls short of the level required by the standard of care owed to the plaintiff.
- **Causation**—The breach of duty is the actual and proximate cause of the plaintiff's injury.
- **Damages**—The plaintiff suffered injury.

Despite potential tort liabilities, state agencies benefit from sovereign immunity, and political subdivisions are protected by governmental immunity under Texas state law. Generally, there is no risk of liability to TxDOT for the act, omission, or negligence of a TxDOT employee who, acting within their scope of employment, causes property damage, personal injury, or death.

Sovereign immunity can be waived by a state agency if a constitutional or statutory waiver exists. Current Texas state statute (the Texas Tort Claims Act) provides such a waiver, enumerating the instances and conditions of, and limitations on, a governmental unit's tort liability for property damage, personal injury, and death [14]. Damages for negligence may be recovered for property damage or personal injury or death. Property damages can be recovered from a government agency only where an employee's wrongful act, omission, or negligence involves the operation or use of a motor-driven vehicle or motor-driven equipment. Personal injury or death are recoverable if the wrongful act, omission, or negligence at issue involves either the operation or use of a motor-driven vehicle or motor-driven equipment, or a condition or use of the agency's tangible personal or real property.

Further research is needed to confirm that the law has not yet been applied to cases where a state employee's wrongful act, omission, or negligence involved the operation or use of a vehicle or equipment that used erroneous or otherwise defective TCI data. Similarly, the law is currently silent on whether an agency's TCI data can be regarded as tangible personal or real property, the condition or use of which, if defective, may cause damages.

Today, TxDOT can be held liable for premise defects on its roads and other assets that it owns and maintains. For roads and streets, the standard of care TxDOT is held to is that of a private person to a licensee on private property. Thus, TxDOT has a duty not to injure people on the state's roads "by a wilful or wanton act" or "through gross negligence," which "can be defined as knowing indifference to the rights, welfare, or safety of others." [15] If TxDOT has actual knowledge of a defect on its roadway, and the driving public on that road does not, it has a duty to either warn drivers or make the condition reasonably safe.

In anticipation of future claims of negligence from defective conditions or use of TCI data, TxDOT could acquire, store, maintain, update, and disseminate such data, applying the same standard of care the law applies to premise defects and defects on its roadways. For non-roadway premise defects, government agencies owe the duty of care a private person owes to a licensee on private property, unless the person has paid for use of the premises. Thus, TxDOT could adopt a policy of warning users of defective or otherwise potentially dangerous TCI data, as well as hacking of such data, if it cannot make the data reasonably safe once it has actual knowledge of the defect or threat. If TxDOT accepts payment for use of the TCI data, it would owe a higher duty and need to maintain the data in a reasonably safe condition, inspect so that unreasonable risks of harm can be discovered, and protect against danger by addressing any defects or giving adequate warning of any defects.

Even if TxDOT is determined liable for property damage or personal injury or death, the agency could mitigate damages thanks to state law. As a comparative negligence state, Texas law allows a reduction in a plaintiff's recovery if the plaintiff was partially to blame for their injury. Texas law also caps damages on liability for actions against a governmental entity involving governmental functions to the extent that sovereign immunity has been waived. For state government, liability is limited to $250,000 in money damages for each person, $500,000 for each single occurrence for bodily injury or death, and $100,000 for each single occurrence for injury to or destruction of property.

In addition, TxDOT could protect itself from potential liability through including indemnification clauses in its contractual agreements with third parties using and applying TCI data to AVs. Indemnification is a means of shifting loss between or among parties targeted by plaintiffs injured by use or misuse of TCI data. However, this may give rise to another waiver of sovereign immunity under the joint enterprise theory. TxDOT may engage in a joint enterprise, formed through an agreement for a common purpose, where it and other parties have equal voice in the direction of the enterprise and the other parties serve as TxDOT agents in use and application of TCI data for AVs. In this case, TxDOT could be liable for the partners’ negligence as if it were a private person.

3.4.7 Conclusion

The collection, storage, sharing, and maintenance of TCI data raises legal questions about data protection, data privacy, data ownership, and agency liability, among others. The risks of managing, using, and disseminating this data include risks of infringement of intellectual property, invasion of personal privacy, and products liability defects. These risks and an agency’s exposure to liability for these torts are summarized below.

3.4.7.1 Risks

- Proprietary Data and Trade Secrets: Disclosing vehicle data containing proprietary product design data or trade secret information could cause economic harm to the owner of that data.
- Privacy: Personal private data in the possession of a governmental entity can, if not sufficiently anonymized or protected, be inadvertently disclosed in response to a request under a state's Public Information Act, if shared with and misused by law enforcement, or if disclosed due to a negligent or criminal breach.
- Product Liability: Using TCI data to produce a product for public consumption, packaged for dissemination to AVs, could effectively make the agency a designer or manufacturer of a product and expose the agency to tort liability for product-related injuries.

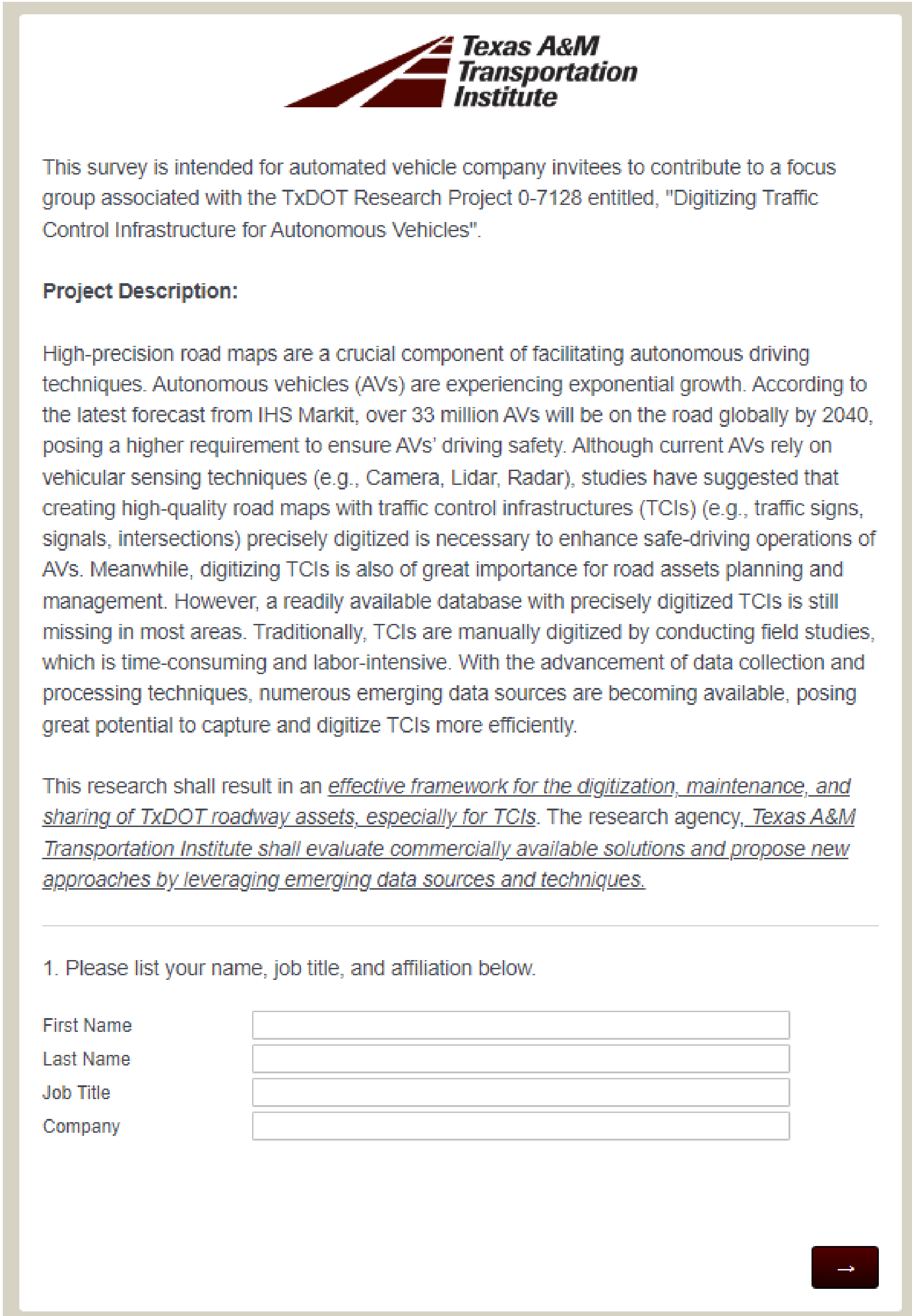
3.4.7.2 Tort Liability

- State agencies in Texas benefit from sovereign immunity from tort liability.
- Sovereign immunity can be waived by a state agency if a constitutional or statutory waiver exists, as it does in Texas.
- A governmental unit can be liable for tort liability in the event of property damage, personal injury, and death if an employee's wrongful act, omission, or negligence involves the operation or use of a motor-driven vehicle or motor-driven equipment.

3.4.7.3 Recommendations

- Processes for acquiring and processing data, however sourced, must be reviewed and a data management plan and data privacy model should be put in place.
- If producing products using TCI data, TxDOT could protect itself from potential liability through including indemnification clauses in its contractual agreements with third parties using and applying TCI data to AVs.

3.5 APPENDIX B. FINDINGS FROM AV COMPANIES SURVEY



Texas A&M Transportation Institute

This survey is intended for automated vehicle company invitees to contribute to a focus group associated with the TxDOT Research Project 0-7128 entitled, "Digitizing Traffic Control Infrastructure for Autonomous Vehicles".

Project Description:

High-precision road maps are a crucial component of facilitating autonomous driving techniques. Autonomous vehicles (AVs) are experiencing exponential growth. According to the latest forecast from IHS Markit, over 33 million AVs will be on the road globally by 2040, posing a higher requirement to ensure AVs' driving safety. Although current AVs rely on vehicular sensing techniques (e.g., Camera, Lidar, Radar), studies have suggested that creating high-quality road maps with traffic control infrastructures (TCIs) (e.g., traffic signs, signals, intersections) precisely digitized is necessary to enhance safe-driving operations of AVs. Meanwhile, digitizing TCIs is also of great importance for road assets planning and management. However, a readily available database with precisely digitized TCIs is still missing in most areas. Traditionally, TCIs are manually digitized by conducting field studies, which is time-consuming and labor-intensive. With the advancement of data collection and processing techniques, numerous emerging data sources are becoming available, posing great potential to capture and digitize TCIs more efficiently.

This research shall result in an *effective framework for the digitization, maintenance, and sharing of TxDOT roadway assets, especially for TCIs*. The research agency, *Texas A&M Transportation Institute shall evaluate commercially available solutions and propose new approaches by leveraging emerging data sources and techniques.*

1. Please list your name, job title, and affiliation below.

First Name

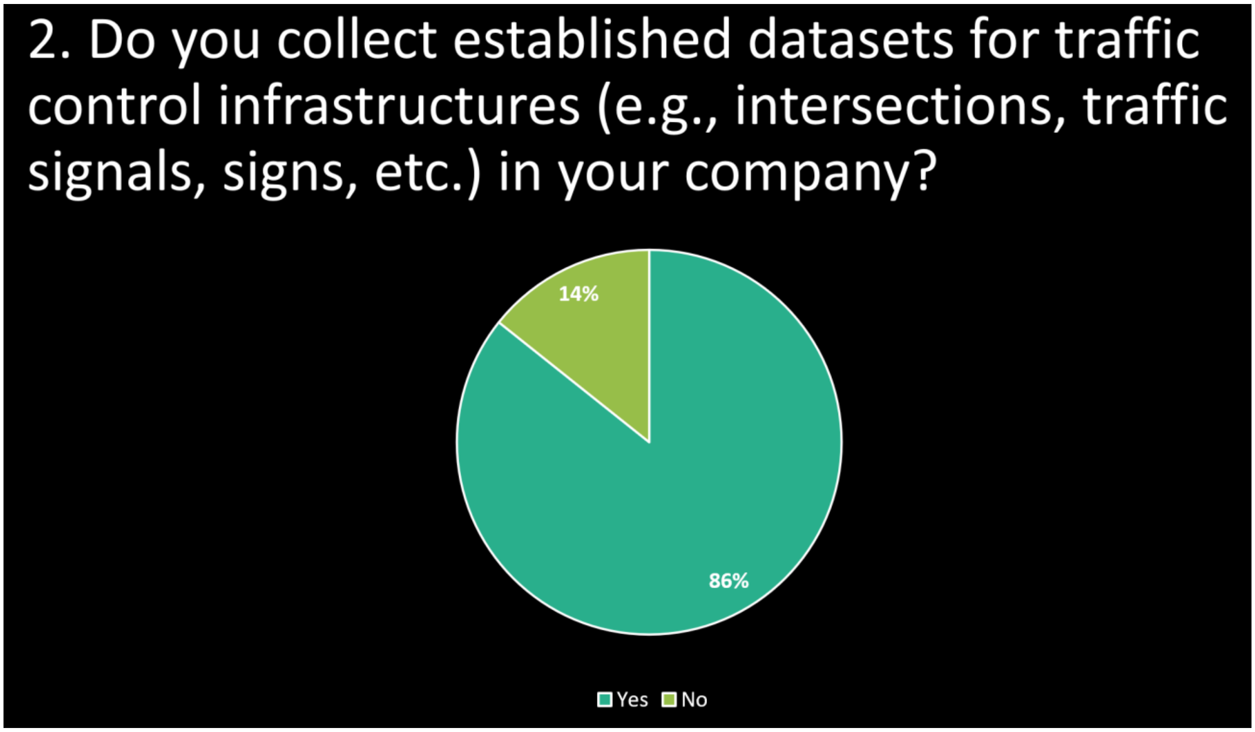
Last Name

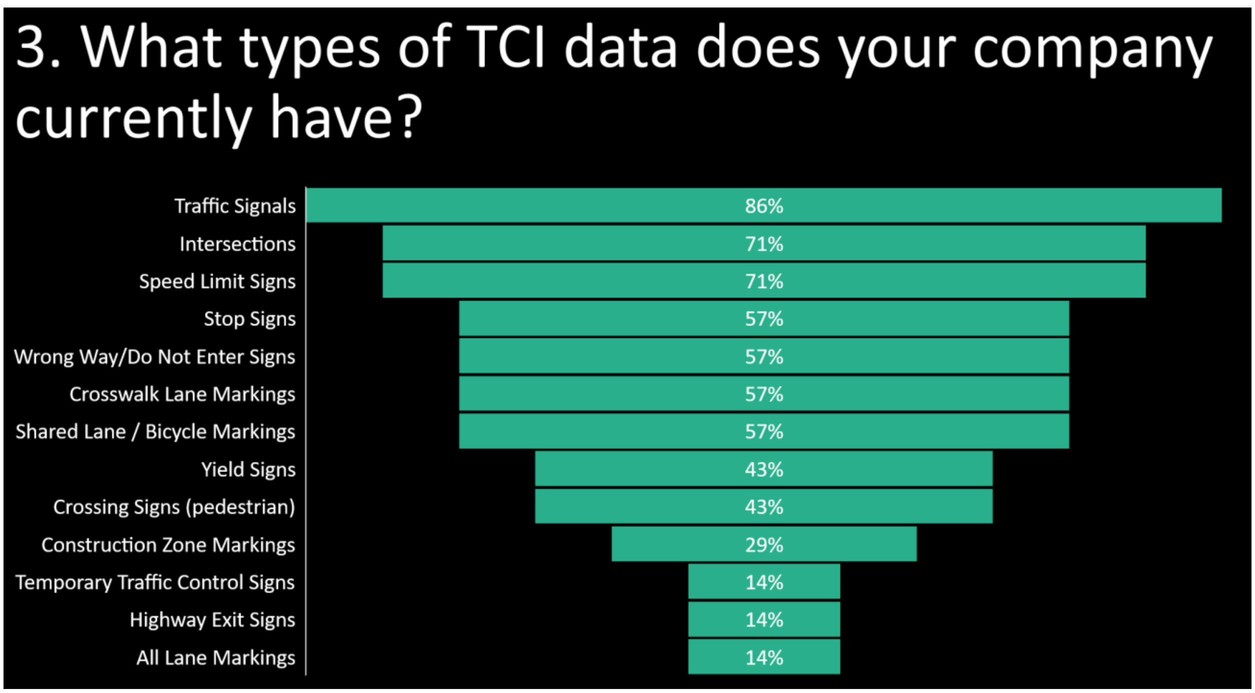
Job Title

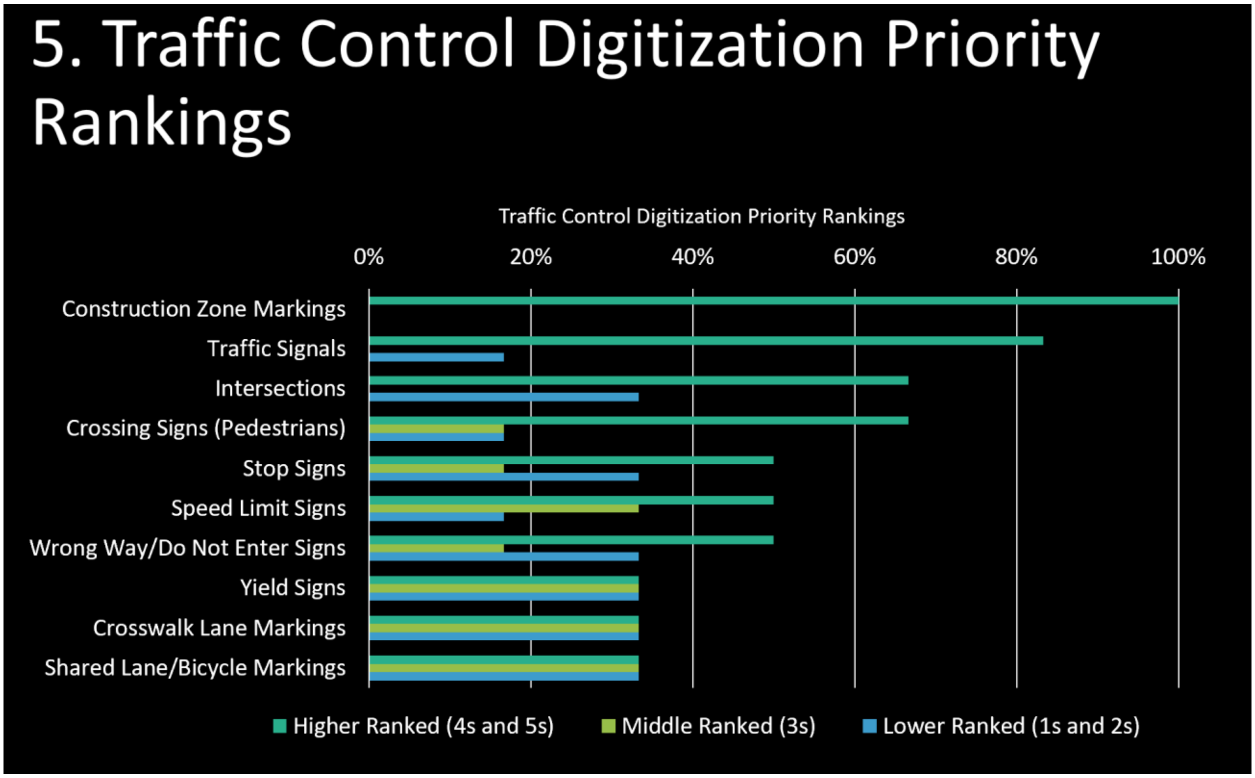
Company

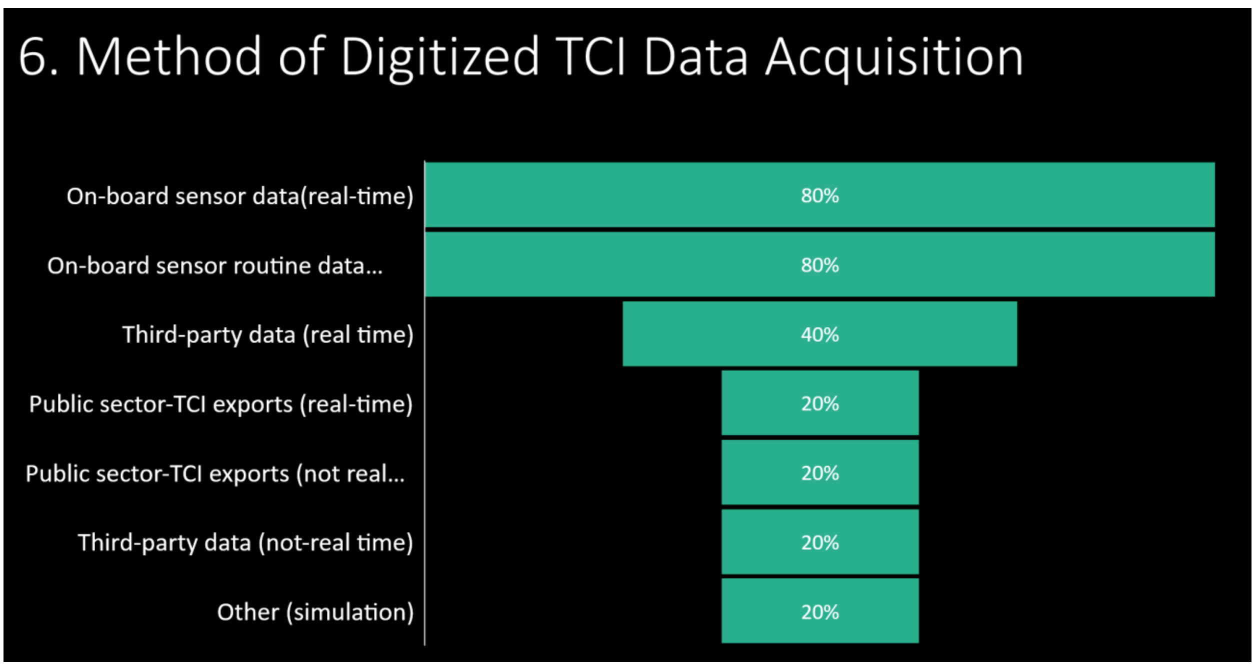
→

Survey Link: https://tti.qualtrics.com/jfe/form/SV_4Pc3N5DMf36xRjg


2. Do you collect established datasets for traffic control infrastructures (e.g., intersections, traffic signals, signs, etc.) in your company?
14%
86%
Yes
No


3. What types of TCI data does your company currently have?
Traffic Signals
86%
Intersections
71%
Speed Limit Signs
71%
Stop Signs
57%
Wrong Way/Do Not Enter Signs
57%
Crosswalk Lane Markings
57%
Shared Lane / Bicycle Markings
57%
Yield Signs
43%
Crossing Signs (pedestrian)
43%
Construction Zone Markings
29%
Temporary Traffic Control Signs
14%
Highway Exit Signs
14%
All Lane Markings
14%


5. Traffic Control Digitization Priority Rankings
Traffic Control Digitization Priority Rankings
0%
20%
40%
60%
80%
100%
Construction Zone Markings
Traffic Signals
Intersections
Crossing Signs (Pedestrians)
Stop Signs
Speed Limit Signs
Wrong Way/Do Not Enter Signs
Yield Signs
Crosswalk Lane Markings
Shared Lane/Bicycle Markings
Higher Ranked (4s and 5s)
Middle Ranked (3s)
Lower Ranked (1s and 2s)


6. Method of Digitized TCI Data Acquisition
On-board sensor data(real-time)
80%
On-board sensor routine data...
80%
Third-party data (real time)
40%
Public sector-TCI exports (real-time)
20%
Public sector-TCI exports (not real...
20%
Third-party data (not-real time)
20%
Other (simulation)
20%

8. TCI Data Providers In Use

- Kimley Horn,
- Iteris,
- Atomic Maps,
- Siradel,
- LYT,
- Vantiq,
- NVIDIA,
- Above,
- WWT,
- AIM,
- Data Gumbo,
- Slingshot,
- Resilienx,
- Aqueti,
- H3,
- Unifly,
- FAA SWIM,
- Truweather,
- Velodyne,
- FlightCorp.
- Trafficware,
- 3rd party global-scale navigation map providers

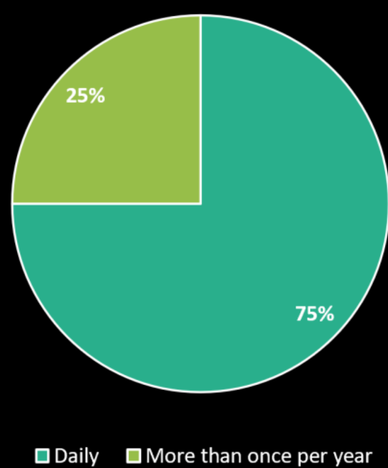
TCI Data in Use (Contd.)

- Sensor data,
- Digital twin stock data,
- Traffic flow,
- Traffic history,
- Incident data,
- RF data,
- City IoT
- 3rd party global-scale navigation map providers

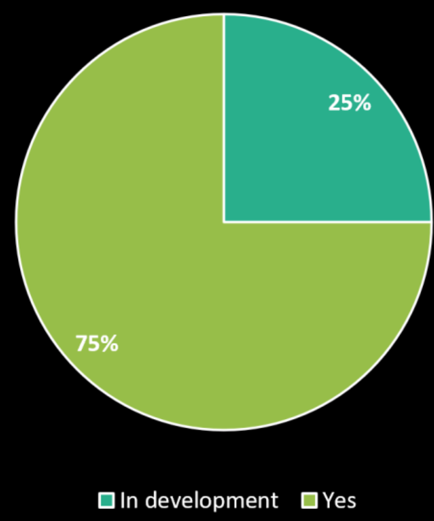
10. How satisfied are you with their data product in terms of quality, cost, or other key consideration?

- 2 respondents responded to question 10.
- Of these two responses, 1 ranked quality as low, and cost as high.
- In ranking 'other' categories, both ranked as low, real-time flow and what was termed “data freshness”, which could be for the same.

11. How Often the AV Trucking Company Updates the TCI Dataset

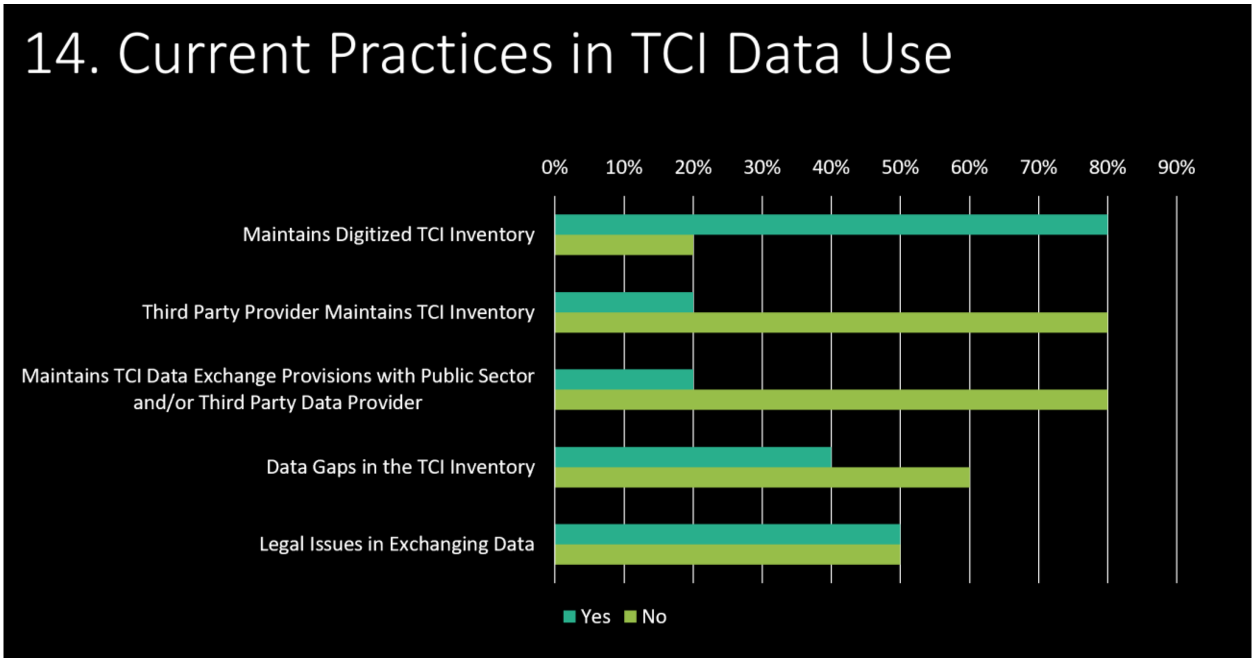


12. Does your company have a data management plan for your TCI assets?



13. What strategies/solutions/standards have you adopted for data quality control, assessment, and enhancement?

- 2 respondents advised they:
  A. a) adopted data exchange standards, and
  B. b) compare maps to the real world as part of their QAQC process.



3.6 REFERENCES

[1] Xiao Li, Jason Wu, Moxuan Li, Gretchen Stoeltje, "Task Report 2, Literature Review, TxDOT Project 0-7128, Digitizing Traffic Control Infrastructure for Autonomous Vehicles (AV)", p. 8.

[2] Ibid.

[3] Kara Kockelman, Lisa Loftus-Otway, Duncan Stewart, Aqshems Nichols, Wendy Wagner, Stephen Boyles, Michael Levin, Jun Liu, Kenneth Perrine, Scott Kilgore, Krishna Murthy Gurumurthy, "BEST PRACTICES FOR MODIFYING TRANSPORTATION DESIGN, PLANNING, AND PROJECT EVALUATION IN TEXAS", pp. 10.

[4] Kara Kockelman, Lisa Loftus-Otway, Duncan Stewart, Aqshems Nichols, Wendy Wagner, Stephen Boyles, Michael Levin, Jun Liu, Kenneth Perrine, Scott Kilgore, Krishna Murthy Gurumurthy, "BEST PRACTICES FOR MODIFYING TRANSPORTATION DESIGN, PLANNING, AND PROJECT EVALUATION IN TEXAS", pp. 11.

[5] Trimble, T.E. et al, "Implications of Connected and Automated Driving Systems, Vol. 3: Legal Modification Prioritization and Harmonization Analysis" pp. xii, 8, 41, 42.

[6] Jones Day, "Autonomous Vehicles: Legal and Regulatory Developments in the United States" 15, 16, 17

[7] Steve Perkins and Ann Baddour, Texas Appleseed, "Texas at the Crossroads – An Update on Privacy Legislation, November 15, 2021, https://www.texasappleseed.org/blog/texas-crossroads-%E2%80%93-update-privacy-legislation

[8] HB 3746, 87(R), Bill text (enrolled) https://capitol.texas.gov/tlodocs/87R/billtext/pdf/HB03746F.pdf#navpanes=0

[9] SB 15, 87(R), Bill Analysis, https://capitol.texas.gov/tlodocs/87R/analysis/pdf/SB00015F.pdf#navpanes=0

[10] SB 475, 87(R), Bill Analysis, https://capitol.texas.gov/tlodocs/87R/analysis/pdf/SB00475F.pdf#navpanes=0

[11] David Canton and Hamad Siddiqui, "Privacy Torts: 4 Types of Invasion of Privacy", August 17, 2021. https://hptechlaw.com/blog/2021/8/16/invasion-of-privacy

[12] National Academies of Sciences, Engineering, and Medicine 2019. Practices on Acquiring Proprietary Data for Transportation Applications. Washington, DC: The National Academies Press. https://doi.org/10.17226/25519, https://nap.nationalacademies.org/cart/download.cgi?record_id=25519, p. 22-23.

[13] Texas Government Code, Sec. 552.110. EXCEPTION: CONFIDENTIALITY OF TRADE SECRETS; CONFIDENTIALITY OF CERTAIN COMMERCIAL OR FINANCIAL INFORMATION. https://statutes.capitol.texas.gov/Docs/GV/htm/GV.552.htm

[14] CIVIL PRACTICE AND REMEDIES CODE, TITLE 5, SUBCHAPTER B. TORT LIABILITY OF GOVERNMENTAL UNITS, https://statutes.capitol.texas.gov/Docs/CP/htm/CP.101.htm

[15] Dick Evans, TMLIRP, Lori Gillespie, TMLIRP, William M. McKamie, McKamie Krueger, LLP, Laura Mueller Assistant General Counsel, Texas Municipal League, Texas Tort Claims Act Basics, p. 6.

[16] N. Kargah-Ostadi, A. Waqar, and A. Hanif, "Automated Real-Time Roadway Asset Inventory using Artificial Intelligence," Transportation Research Record, vol. 2674, no. 11, pp. 220–234, Sep. 2020, doi: 10.1177/0361198120944926.

[17] Allen, B.W. et al., "Handbook for Including Ancillary Assets Transportation Asset Management Programs" Report of FHWA-HIF-19-068. https://www.fhwa.dot.gov/publications/research/infrastructure/19068/19068.pdf

CHAPTER 4. DEVELOP A FRAMEWORK FOR DIGITIZING AND SHARING TCIS

4.1 SUMMARY

In Task 4 of this project, the research team obtained and evaluated sample digitized TCI datasets from multiple third-party data providers. A framework for digitizing and sharing these datasets was built up by the research team. An initial discussion on the simulation platform and AV driving scenarios was provided.

4.1.1 Objective

The objective of this task was four-fold:

1. Investigate the selected data sources for TCI digitization.
2. Evaluate the selected data sources.
3. Build a TCI digitization framework.
4. Discuss the simulation platform and AV driving scenarios.

4.1.2 Overview of Chapter 4

This chapter presents the results of Task 4 and is organized as follows:

- Section 4.2 discusses the TCI digitization data sources that were selected and used in this study.
- Section 4.3 presents the initial evaluation results of the selected TCI digitization datasets.
- Section 4.4 shows the developed framework for TCI digitization and data sharing.
- Section 4.5 discusses the AV simulation platform, driving scenarios, and testing sites.

4.2 SELECTED DATA SOURCES FOR TCI DIGITIZATION

4.2.1 Purpose

This sub-task aims to look further into the potential data sources and solutions for TCI digitization that were identified in Task 3. The selected data sources come from the leading companies on the market: Mobileye, Blyncsy, and Nexar. The selection of these three companies is also dependent on their willingness to collaborate with the research team and provide the team with sample TCI digitized datasets for the evaluation purposes.

4.2.2 Mobileye

Mobileye, an Intel company, is a global leader in the development of computer vision and machine learning, data analysis, localization, and mapping for advanced driver assistance systems and autonomous driving solutions.

Mobileye technology uses a single camera to scan the road ahead and detect potential dangers. Leveraging the same technology, Mobileye has reinvented road mapping for asset management and road maintenance. Mobileye's EyeQ® chip and algorithms have been trained to identify, tag, and classify road assets, detect pavement anomalies, and capture mobility data, all as equipped vehicles travel on their regular routes. Additional road data features are collected through Mobileye 8 Connect™–equipped vehicles. Fleet vehicles retrofitted with Mobileye 8 Connect, a collision avoidance system and data collection enabler, collect information as field workers drive from location to location during their workday.

The road asset and mobility information captured by the vehicles is processed and turned into GIS data layers, updated at a relatively high refresh rate (as shown in Figure 17). The geolocation of each asset or road feature is derived from aggregating detections from multiple drives for higher accuracy. "Locations" are reported as latitude, longitude, and altitude coordinates. The list of features is summarized in Table 8, which is regularly revised to incorporate new releases from Mobileye.

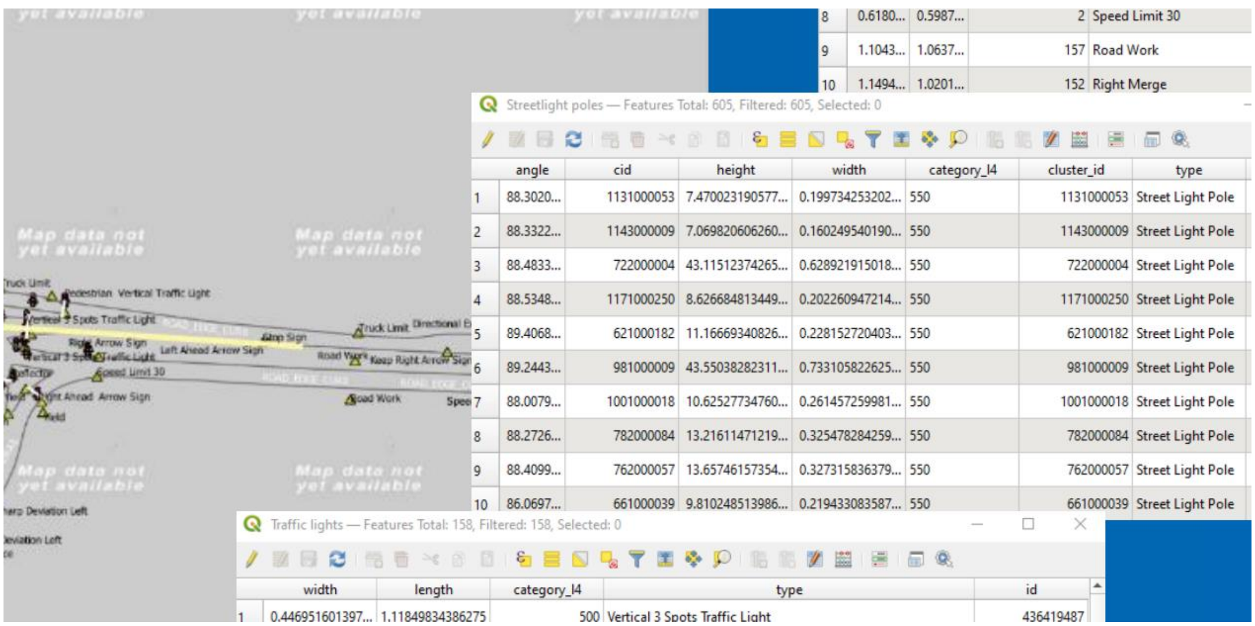


Figure 17. Screenshot of Mobileye's Sample TCI Data.

Table 8. List of Features of Mobileye Data (Released July 2022).

Mobileye Road Asset Data		Mobileye Utility Asset Data	
Features	**Properties**	**Features**	**Properties**
Road edges	Type (guard rails, curb, etc.)	Streetlight poles*	• Type (ME8 and select OEMs) • Tilt Angle**
Road markings	Type (e.g., left-road arrow)	Power poles*	
Lane markings	Type (e.g., solid, dashed, etc.)	**Mobileye Pavement Anomaly Detection**	
Lane width	Lane width	Potholes*, **	• Width • Length • Area • Image (option)**
Traffic signs	Type (overhead, speed limit, stop, etc.) Azimuth	Mobileye Road Image Capture	
Traffic lights	Type (vertical two spot, pedestrian, etc.)	Road Image Capture***	Time of Image Location of Image
Poles	Type (ME8 and select OEMs) Tilt Angle**		
Crosswalks	Type (Zebra Crossing, Solid Crossing, etc.)		

**Only available through Mobileye 8 Connect. **Experimental capability. ***Faces and license plates are obfuscated for privacy.*

4.2.3 Blyncsy

Blyncsy is a movement and data intelligent company founded in 2014 and headquartered in Salt Lake City, Utah. Through the power of big data and location analytics, Blyncsy helps its customers understand how connected devices are moving throughout an environment. Blyncsy's product, Payver, delivers instant insights by using AI to analyze crowd-sourced video data obtained through connected dash cams and smartphones. The AI algorithm extracts road assets information from the crowd-sourced video data, which are then hosted, aggregated, and shared through its cloud-based system, Payver.io.

Payver leverages the power of AI and crowdsourced visual imagery from dash cams to provide automated work zone surveys. This technology collects street-level imagery from over 400,000 weekly active drivers nationwide and assesses the condition of assets in the images in as little as 60 seconds of a vehicle passing. Payver provides agencies with cutting-edge tools to leverage these data in their operations. For real-time work zone reporting, this means that Payver can automatically detect work zones and related equipment, permitted or not, and directly update public feeds and notify relevant authorities so that information is shared with road users in a timely fashion. Figure 18 shows how Blyncsy's Payver "sees" and identifies traffic control assets on the road through a dash camera–captured image.

Figure 18. Screenshot of Blyncsy TCI Sample Data.

4.2.4 Nexar

Nexar is a leading dash cam company that was founded in 2015. Nexar-powered dash cams enable new vision-based applications for better driving. It has developed a portfolio of vision-based data services through its connected dash cam platform for both public and private sector partners to make roadways safer and more efficient. Nexar collects contribute images from its dash cam users to anonymously capture the streetscapes from their drives. Then, the collection of images is processed through its AI-powered algorithms to recognize and locate road assets automatically and accurately. Nexar users cover more than 90 percent of the roads in the top 100 largest U.S. cities, which generate billions of new images monthly.

The road inventory data (as shown in Figure 19) provided by Nexar can detect, monitor, and locate traffic signs and signals, including speed limit signs, yield signs, stop signs, passing restriction signs, no entry signs, one-way signs, pedestrian crossing signs, and school zone signs, among others. It also could monitor the state/condition of road assets.

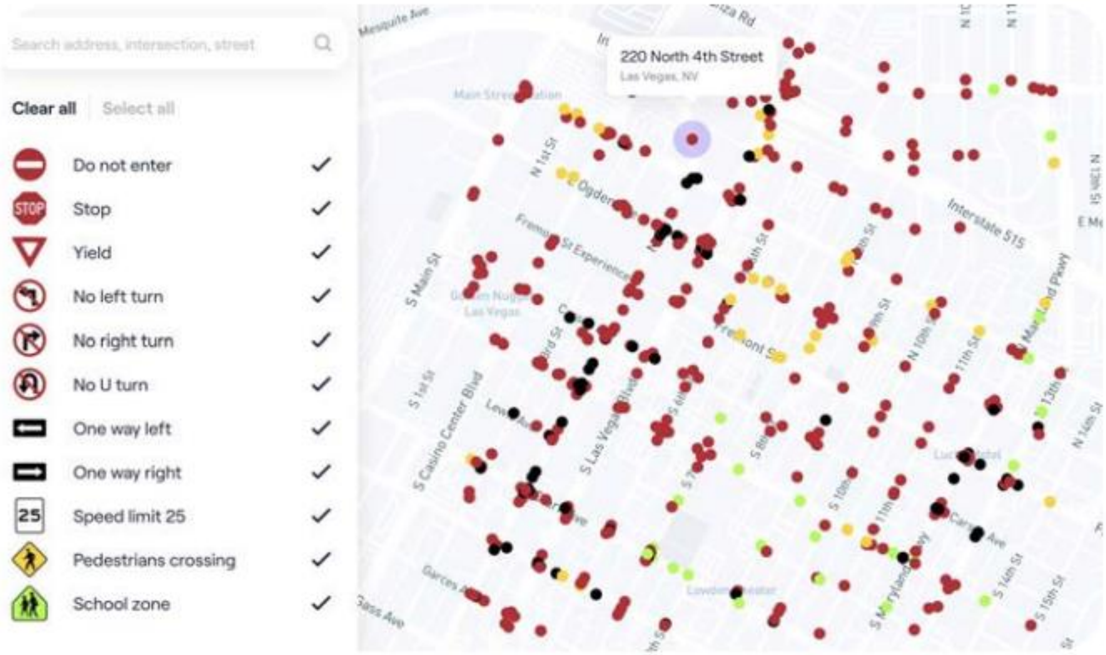


Figure 19. Example of Nexar TCI Data Platform (CityStream).

4.3 EVALUATION OF COMMERICIAL DATA SOURCES

4.3.1 Purpose

This section presents the sample third-party data obtained by the research team and the evaluation results. The research team used the City of Arlington's road asset data as the ground-truth data, which were published by the city. Another ground-truth dataset (Pathway) used by TxDOT was also requested but was delivered late due to the data provider's annual data collection time frame.

4.3.2 Sample Datasets

4.3.2.1 Mobileye

During the task period, the research team actively worked with the Mobileye Data Services team to obtain their sample datasets in several corridors for the evaluation purpose. The corridors include (as shown in Figure 20):

- City of Arlington: I-30 Corridor, from Sandy Ln. to SH360 (Major Freeway).
- City of Arlington: Green Oak Blvd., from I-20 to SH360 (Arterial Corridor).

- City of Dallas: IH-635/LBJ East, from US75 to IH-30 (Major Work Zone Corridor).
- City of Dallas: LP 12/Buckner Blvd., from NW Hwy to US175 (High Transit/Ped/Bike Corridor).

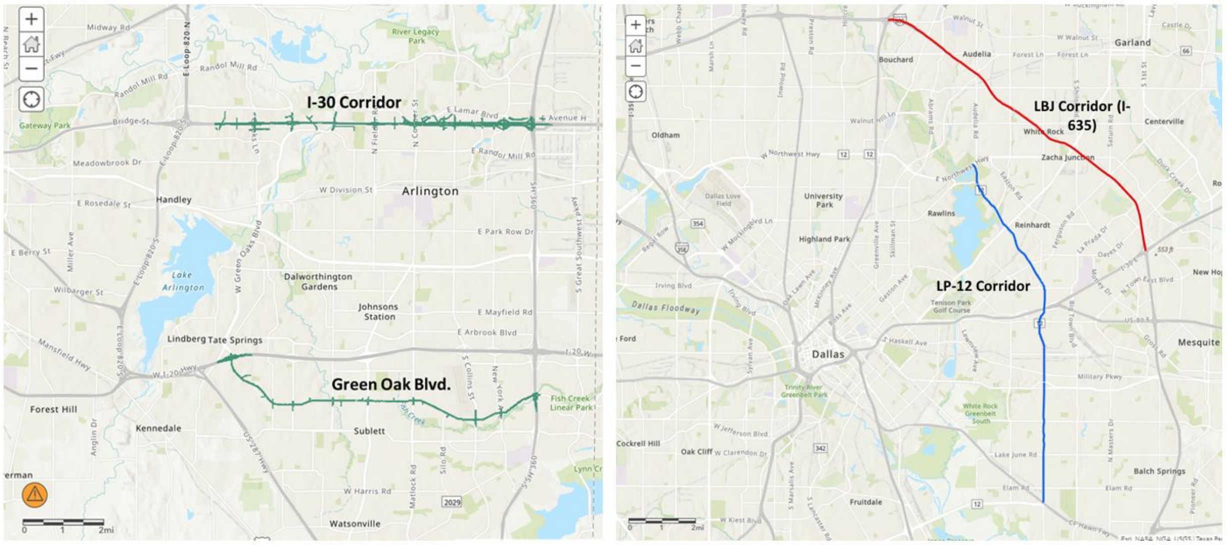


Figure 20. Selected Sample Corridors in the DFW Area.

4.3.2.2 Blyncsy

In 2022, NCTCOG sought firms to supply a situational awareness application testing “sandbox” for staff and local partners to assess the apps. Such apps utilize data from connected vehicles (e.g., dashcam footage) processed using AI techniques to provide roadway operators with tools to improve roadway operations and safety. These tools include: (i) repeated video footage at selected spots on roadways (e.g., high-crash intersections); (ii) automated inventories of street signs, sidewalks, bus stops, and other roadway “furniture”; (iii) effective management of highway assets through frequent identification of new potholes and worn striping; and (iv) identification of work zones and roadway hazards such as roadway debris, inadequate street lighting, and missing signs. Such apps may provide roadway operators greater benefits at lower cost than installing, maintaining, and managing the data from their own roadside equipment and driving around and completing inventory and condition analysis manually.

Blyncsy was selected as the only vendor to test a sample situational awareness pilot that regional partners can use to assess the capabilities, limitations, and potentials of situational awareness apps (Payver in this case). Blyncsy Payver data are mainly road inventory. Elements detected by Payver include Striping, Striping Nighttime Reflectivity/Luminosity, Crosswalk, Cracking, Remote Pavement Surface Evaluation and Rating (PASER), Sign Detection (Day), Sign Reflectivity (Night), Barrels, Street Lights, Impact Attenuators, and Bus Stops. Payver also provides detection images, if needed.

From May to August 2022, Blyncsy collected sample datasets from varied roadway environments over 100 centerline miles on the selected base segments in the DFW area (as shown in Figure 21). As the member of the regional assessment team (across 15 agencies), the Texas A&M Transportation Institute was granted access to the Payver platform and the pilot sample dataset.

4.3.2.3 Nexar

During this task period, the research team also reached out to Nexar and discussed the possibility of exploring its data platform for the evaluation purpose of this project. Following the discussion, a data evaluation agreement was reached in a legal document and the research team was granted a 14-day trial access to Nexar's CityStream platform (as shown in Figure 22). During the trial period, the researchers looked into the platform and compared the data with other data products. At the end of the trial, the research team met with the Nexar team and provided feedback on their product (especially the feasibility for this project).

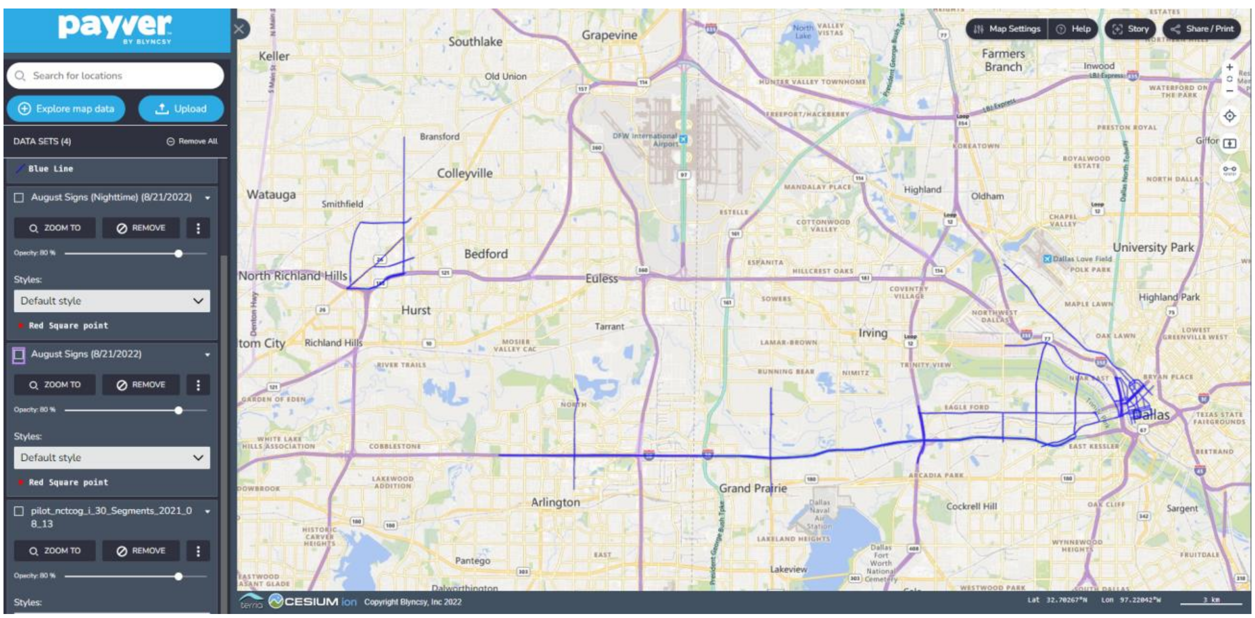


Figure 21. NCTCOG Blyncsy Pilot in the DFW Area.

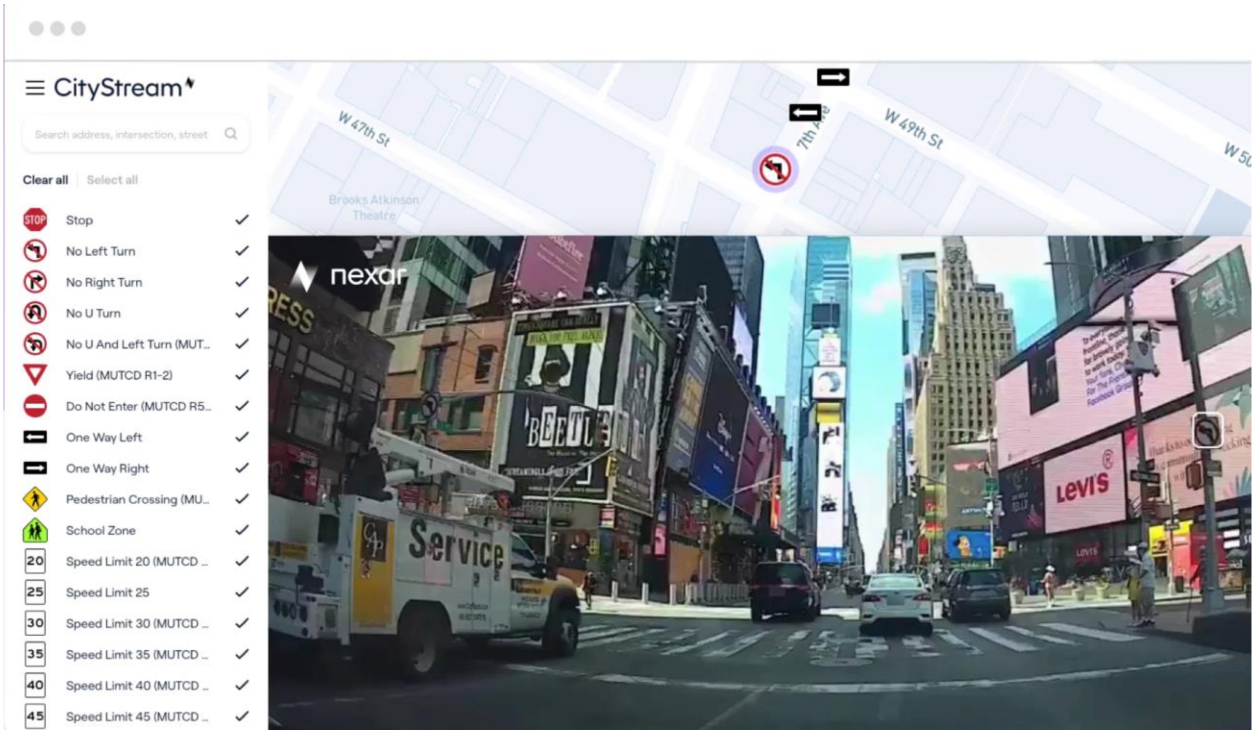


Figure 22. Nexar CityStream Platform.

4.3.3 Evaluation

4.3.3.1 Mobileye

The research team obtained road asset inventory data from the City of Arlington. The researchers used them as the "ground truth" data when compared to the Mobileye sample datasets. For the data type, the Mobileye data contain crosswalks, poles, lane width, road edges, lane markings, traffic signs, traffic lights, road markings, and stop lines information, while the City data only have traffic signs, traffic signals, streetlights, interchange, traffic count, school zones, underground wires, and overall condition index information.

The research team mainly focused on the common information between the two datasets on the sample corridors. For the traffic signs, the City's data are more detailed near the I-30 area (except for the signs on I-30). In the first pair of images in Figure 23, the City's data have more than 22 records near the intersections of Cooks Ln. and I-30, while the Mobileye data have more than 15 records near the intersections. In the second pair of images, the City's data have more than 27 records near the intersections of Coopers St. and I-30, while the Mobileye data have 25 records.

Figure 23. Comparison of Traffic Sign Data on I-30, Arlington, TX.

The findings are the same for the Green Oak Boulevard corridor. The City's data have more enriched records compared to the Mobileye data, as shown in Figure 24.

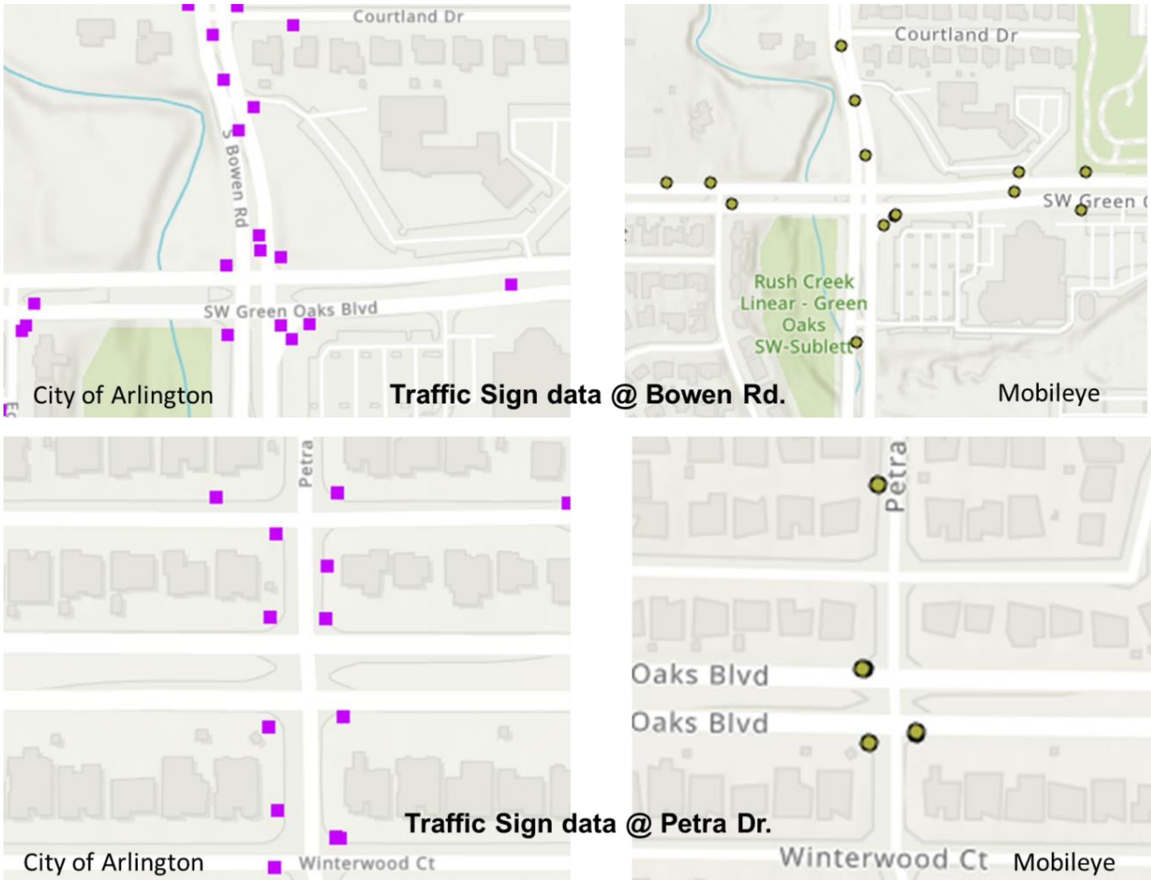


Figure 24. Comparison of Traffic Sign Data on Green Oaks Blvd., Arlington, TX.

Also, for the record attributes, the City's data contain the *Manual on Uniform Traffic Control Devices* (MUTCD) code and the street name. The Mobileye sign data cover the azimuth field, sign type, and category I4 (the rough classification for the traffic detected signs). For example, as shown in Figure 25, for a traffic sign that was recorded in both of the datasets, the sign type is coded as "R14-1" in the City's database, which represents a "Truck Route" sign type. However, in the Mobileye dataset, the type is only shown as a directional sign with a category I4 code 43.

Figure 25. Differences of Traffic Sign Data Attributes.

For the traffic light data, Mobileye has significantly more detailed data than the City's records. In the first pair of images in Figure 26, the City only has two records, while Mobileye has 21 records. For the second pair of images, the City's data also have two records, while the Mobileye data have 23 records.

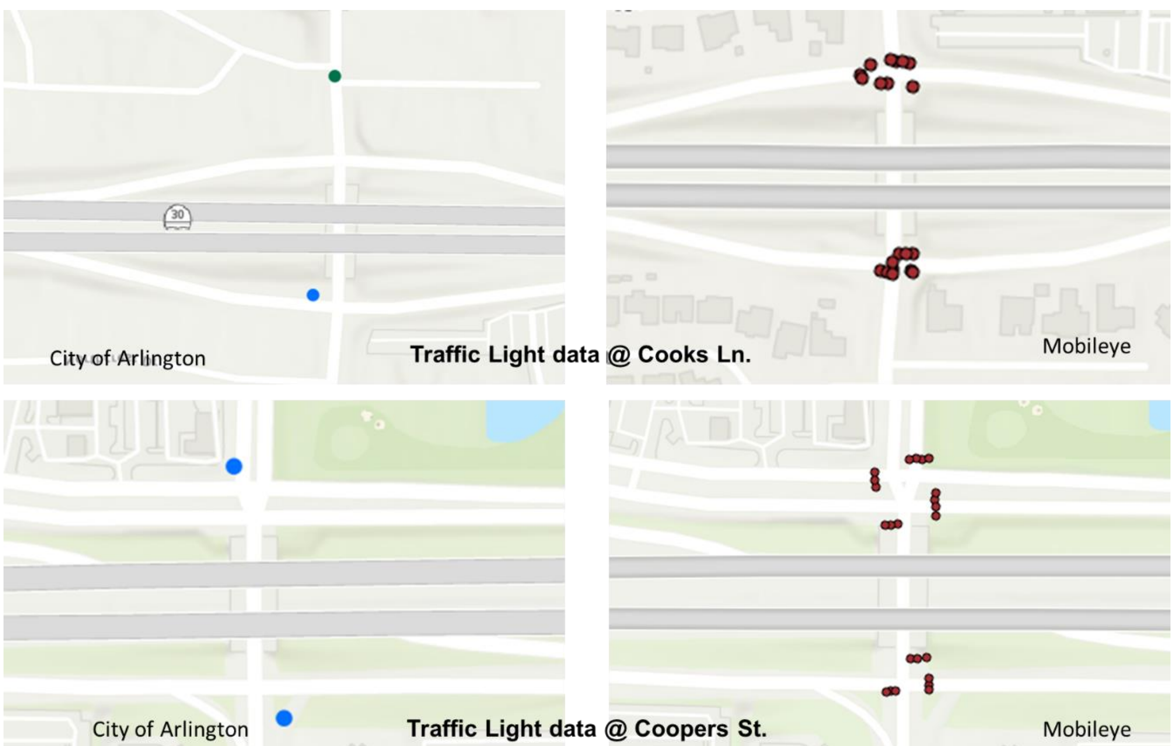


Figure 26. Comparison of Traffic Light Data Attributes on I-30.

The situation is the same for Green Oak Boulevard. While the City's data only have a few records, the Mobileye data have a more detailed description of the traffic lights along this corridor, as shown in Figure 27.

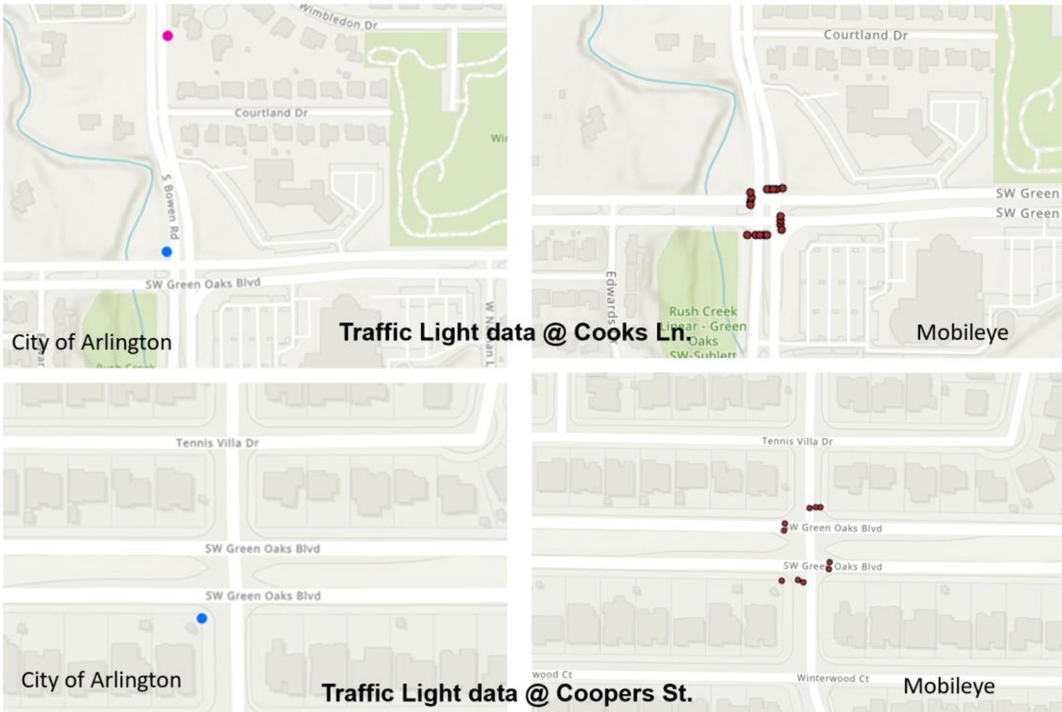


Figure 27. Comparison of Traffic Light Data Attributes on Coopers Street

The last common information type is lane width (Figure 28). The City's data record the lane in segment and calculates the width of the entire segment. However, the Mobileye data record the lane in point so that they calculate the part of the segment and record the lane width. As a result, the Mobileye data are at a much more detailed level.

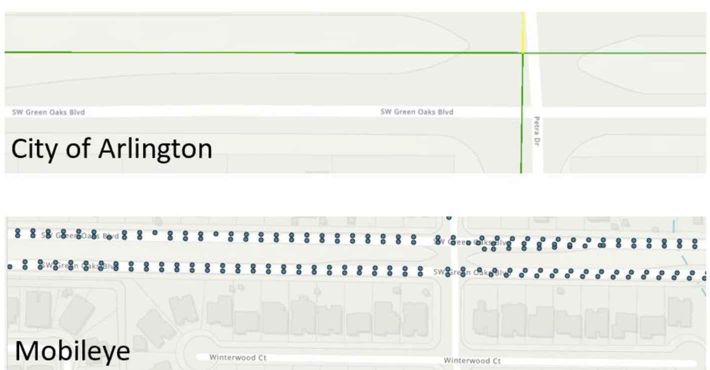
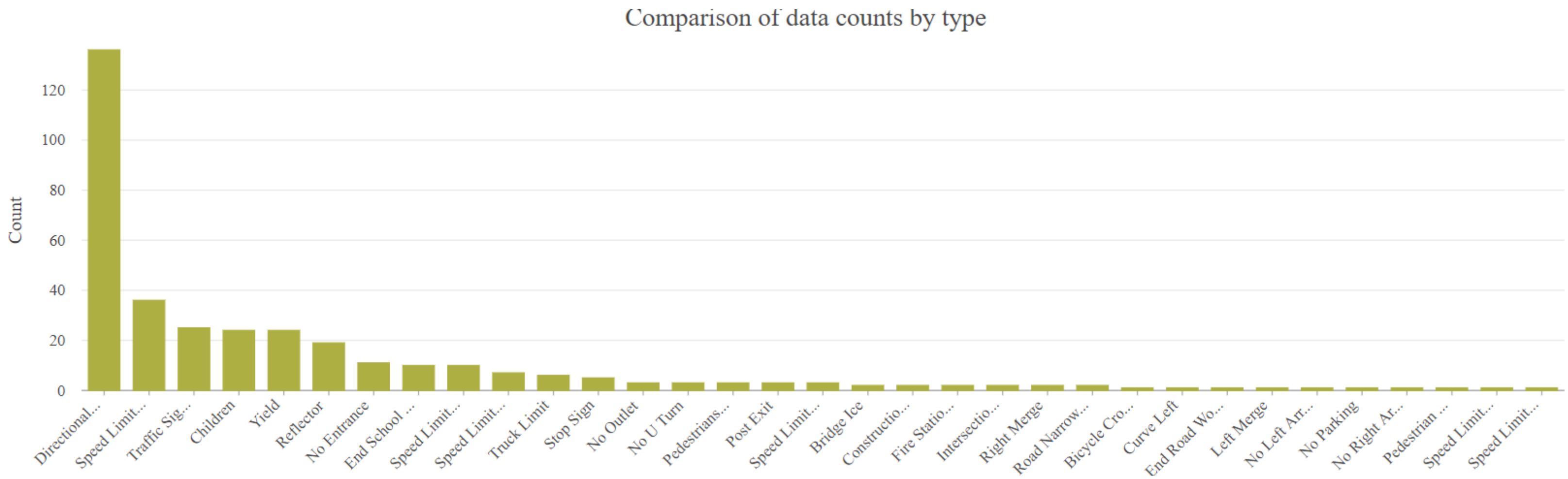
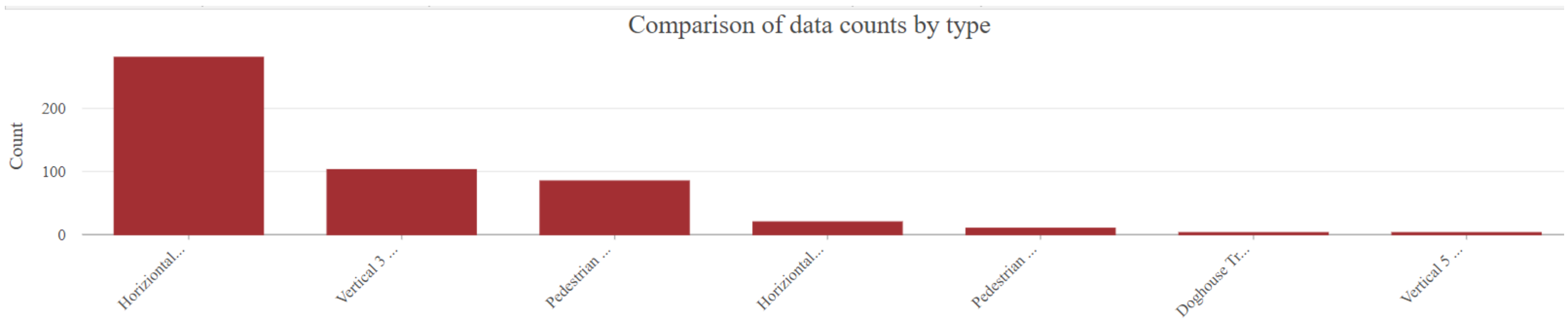


Figure 28. Comparison of Lane Width Data.

The City records three types of traffic lights: signals, school flashers, and warning flashers. Mobileye has more detailed types, as shown in Figure 29.



(a) Mobileye traffic sign data



(b) Mobileye traffic light data

Figure 29. Comparison of Mobileye Data Types for Traffic Signs and Lights.

4.3.3.2 Blyncsy

In the NCTCOG Pilot project, Blyncsy collected about 10 types of road asset data on the selected corridors in the DFW area. The red squares represent the data collected by the probe-based vehicles (as shown in Figure 30). Elements detected by Payver include Striping, Striping Nighttime Reflectivity/Luminosity, Crosswalk, Cracking, PASER, Sign Detection (Day), Sign Reflectivity (Night), Barrels, Street Lights, Impact Attenuators, and Bus Stops.

The Blyncsy Pilot also collected data along the I-30 corridor, which have some overlap with the Mobileye datasets. This gave the researchers a great opportunity to look at the data products of the two leading companies in the market.

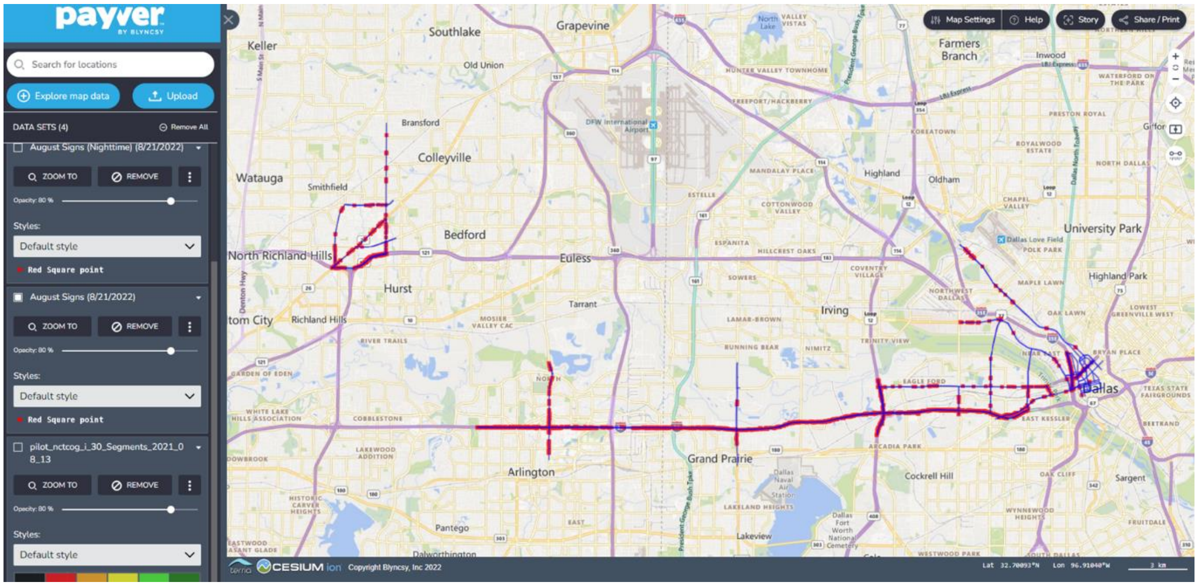


Figure 30. Collected Traffic Sign Data from NCTCOG Blyncsy Pilot Project.

After comparing the two datasets, the research team found some issues with the Blyncsy data. First, the signs were not located accurately on the map (the location appeared to be the vehicle's location where the sign image was), and the same sign may be detected and shown multiple times on the map. For example, as shown in Figure 31, the two square dots should be the same exit directional sign at the I-30 exit ramp, but the location is not as accurate when compared to Mobileye's data, as shown in Figure 32. The Blyncsy data also did not identify the sign type, and it is difficult to download the data of multiple features from the platform. The research team provided this feedback to NCTCOG and Blyncsy at follow-up meetings.

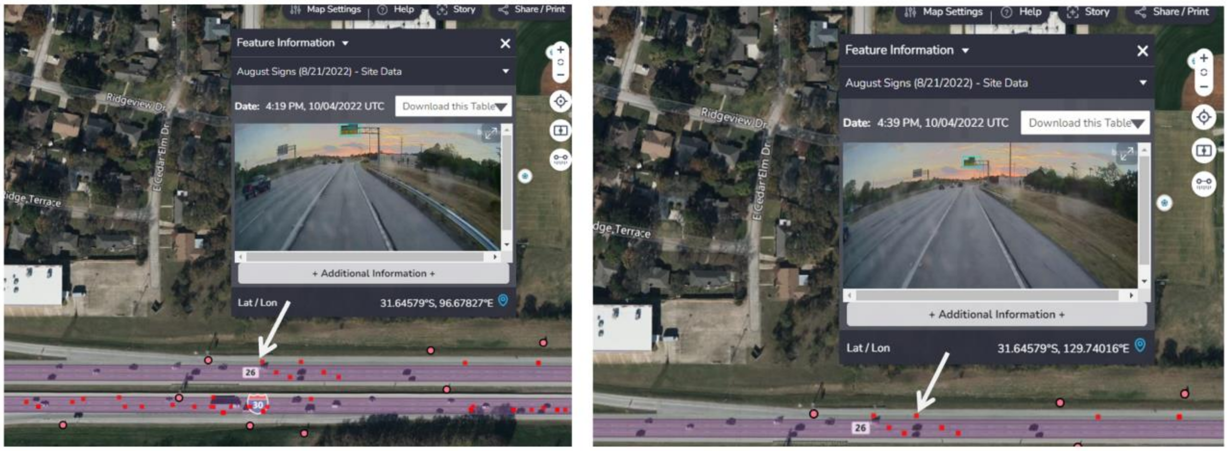
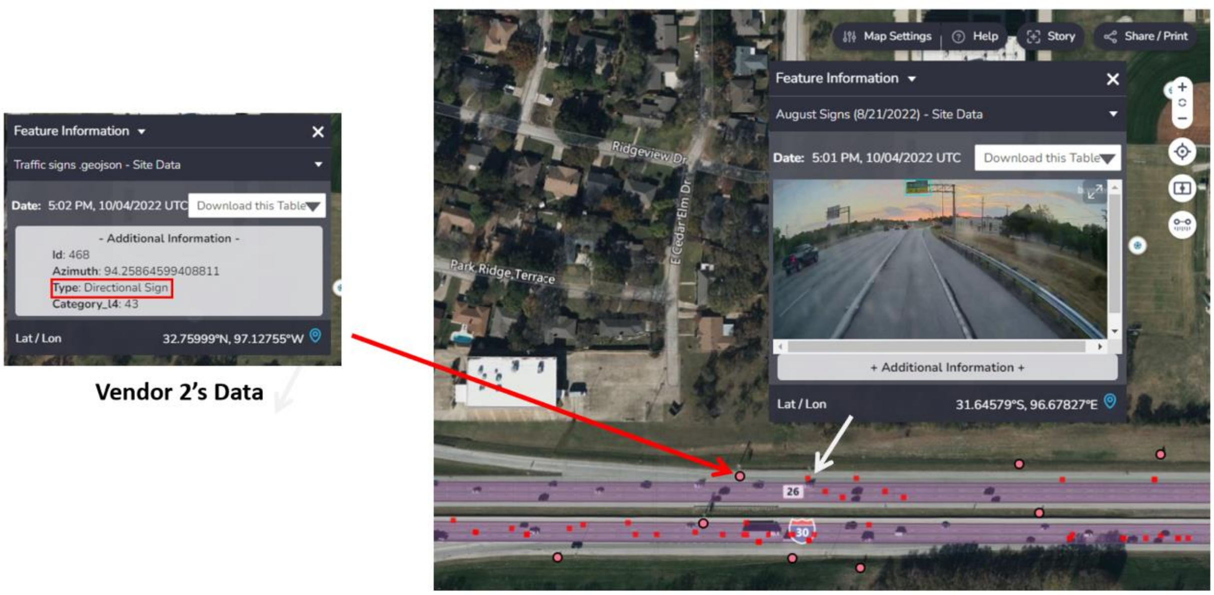


Figure 31. Duplicate Detection Results in the Blyncsy NCTCOG Pilot Data.



(Note: Vendor 2 is Mobileye)

Figure 32. Location Accuracy Issue of the Blyncsy NCTCOG Pilot Data.

4.3.3.3 Nexar

The research team also compared the Blyncsy Payver Platform (NCTCOG Pilot project) with the Nexar CityStream platform (trial of the City of Los Angeles, CA). These two datasets did not cover the same geospatial area.

The CityStream platform provides images of the traffic elements detected. The CityStream data consist of:

- Road work zones: As shown in Figure 33, CityStream identifies work zone segments along with individual traffic elements such as barricades or cones.

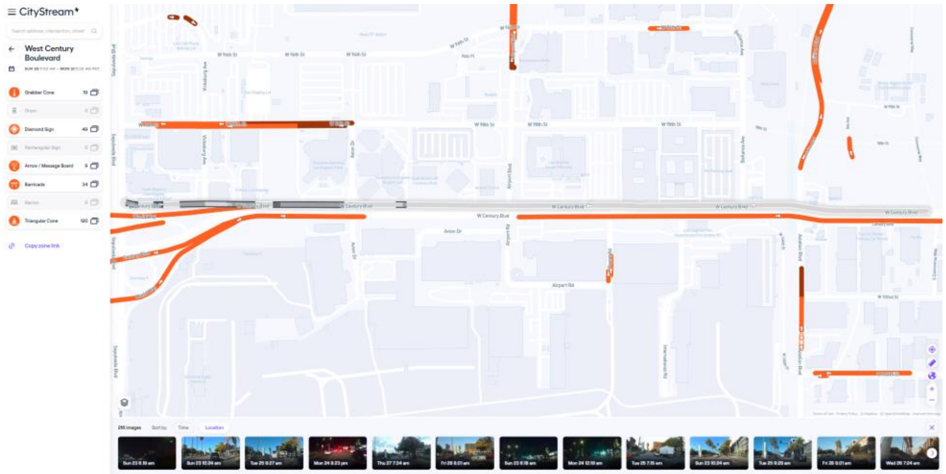


Figure 33. CityStream Work Zone Detection.

- Road inventory: CityStream detects the following road signs (Figure 34):

Figure 34. CityStream Road Sign Detection.

- Virtual camera: CityStream allows the user to see images of roadway sections where Nexar dash cam–equipped vehicles are traveling (as shown in Figure 35). It also provides historical images.

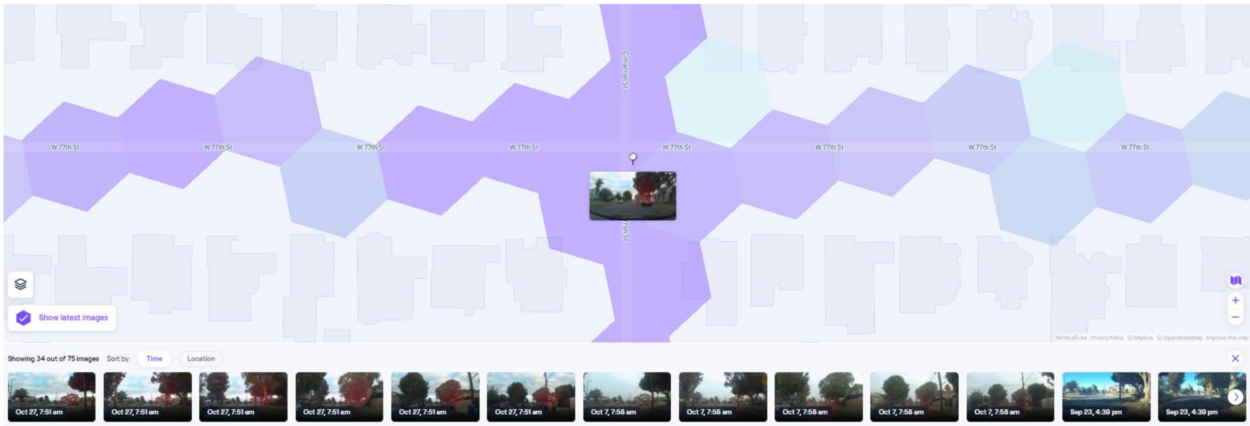


Figure 35. CityStream Virtual Camera.

- Nexar streets: This aspect of the data is similar to Google Street View but, in theory, up to date because of the high amount of Nexar dash cam–equipped vehicles.

Blyncsy Payver data are mainly road inventory. Elements detected by Payver include Striping, Striping Nighttime Reflectivity/Luminosity, Crosswalk, Cracking, PASER, Sign Detection (Day), Sign Reflectivity (Night), Barrels, Street Lights, Impact Attenuators, and Bus Stops. Payver provides detection images.

CityStream is superior when it comes to work zone detection. Payver detects work zone elements but does not identify work zones. Regarding roadway inventory, CityStream only detects vertical signs. Payver detects vertical signs and pavement and striping condition. Payver does not provide virtual camera nor street view.

Nexar CityStream is a platform ready for mass consumption, while Blyncsy Payver is a customized platform where transportation agencies can detect their desired roadway assets/elements.

After the trial period, the research team decided not to proceed with the Nexar data for this study. The main reason was that the Nexar CityStream data are suitable for rolling road asset inventory management at a large scale but are not suitable for improving AV driving since they cannot provide a high-definition map. After discussing with the Nexar team, the researchers stopped the trial of the platform.

4.3.3.4 Takeaways

- Among the three evaluated data products, Mobileye is superior to the other two providers in accuracy and special coverage for the needs of this study.
- Blyncsy's dataset is not mature enough to collect the TCI data needed to meet the needs of this study (when tested). Nevertheless, the platform might continue to develop, and the

data could become more refined over time. The evaluation results merely capture the outcomes as of the researchers' testing period.

- Nexar's platform is more suitable for rolling history road asset management, not for improving AV driving and safety.
- The collected data still have room to be improved (e.g., the traffic sign with the correct MUTCD code). This applies to all three data products.
- Spatial coverage will vary across the country. It is best to contact the data provider before defining the study area to ensure there is coverage.
- Penetration rate and frequency of Mobileye is low in general, but it is sufficient to accurately collect road asset data. Mobileye can also escalate its detection capabilities in any specific area to meet special needs.
- Drive frequency values (over 10 per hour) in Urban Interstate, Freeway, or Expressway could be sufficient to detect work zones, stall vehicles, and road hazards in real time.

4.4 DEVELOPED FRAMEWORK FOR TCI DIGITIZATION AND DATA MANAGEMENT

4.4.1 The Developed Framework

Based on an inventory workflow raised by Nima Kargah-Ostadi's team [16], the TAM guide [17], and stakeholders' responses obtained from the focus group discussion, the research team proposed a conceptual framework for TCI inventory establishment and management. This conceptual framework contains seven modules, including project description and data need, data acquisition, data management, quality check and legal issue assessment, TCI inventory creation, TCI data update, and data sharing and visualization (as shown in Figure 36).

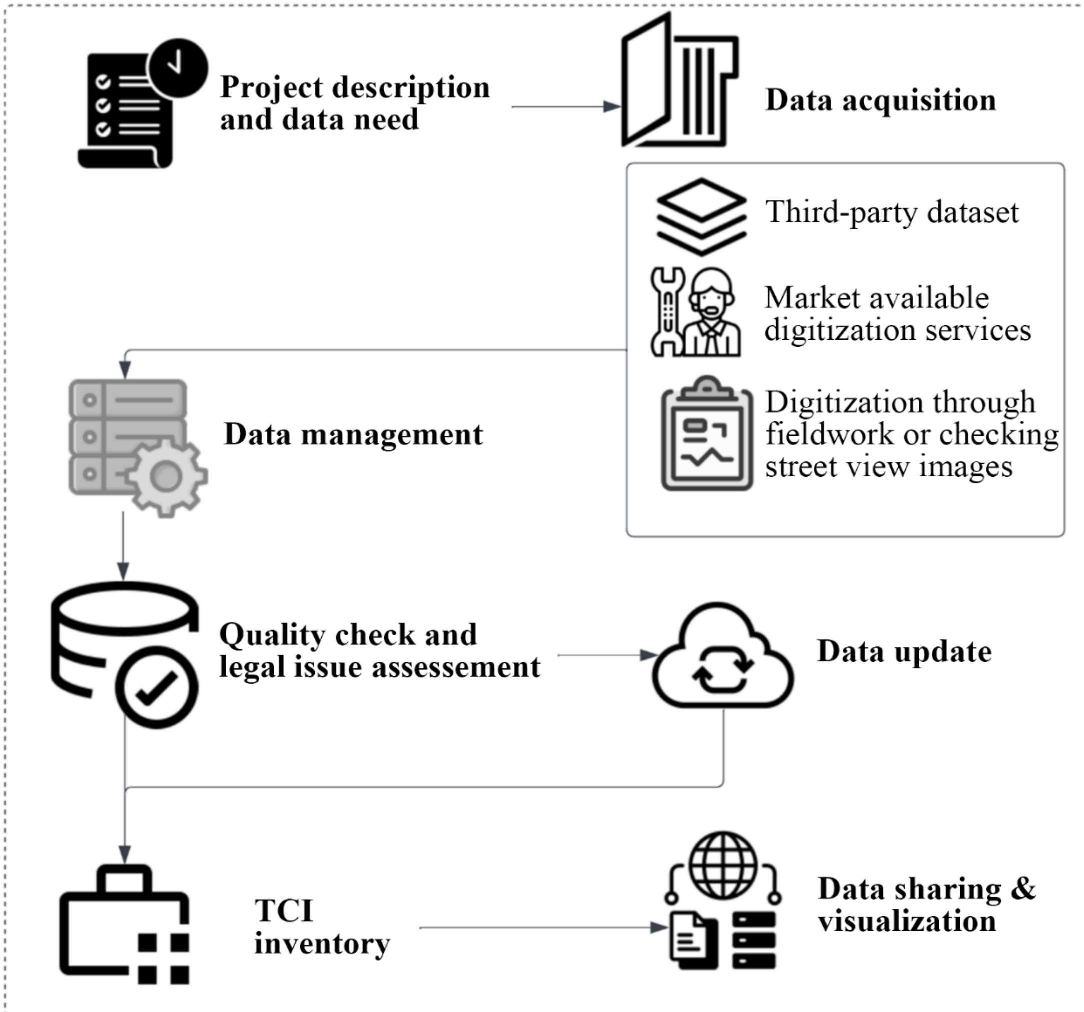


Figure 36. Conceptual Framework for TCI Inventory Establishment and Management.

4.4.2 Framework Modules

4.4.2.1 Project Description and Data Need

Define goals and objectives: "Transportation agencies have traditionally focused their asset management implementations on pavement and bridge assets. There is less information/data available about the ancillary traffic control infrastructures (e.g., signs and signals, guardrail, culverts, pavement markings, etc.) that an agency maintains." [17] Therefore, when an agency considers expanding its data collection practices to include TCIs, it may need to conduct a strategic review that accounts for agency policies, budgets, resources, and performance, as well as any legislated mandates and available technology related to the management of infrastructure or supporting assets. The agency should define its goals and objectives for the TCI digitization task so resources can be aligned with agency priorities.

Prioritizing data need: With limited resources and budget constraints, agencies need to decide how extensively available resources are dedicated to collecting, managing, and using asset information versus delivering work that preserves and improves asset conditions and system performance. Prioritizing the data needs for TCIs is an essential step to ensure the effective and efficient use of limited resources. The primary product of this effort is a ranked list of TCIs for inclusion in the transportation asset management program, with the TCIs grouped into prioritized tiers.

4.4.2.2 Data Acquisition

Data collection approaches: In general, there are three types of approaches that agencies can take for the TCI data collection: third-party TCI datasets, commercially available TCI digitization services, and building on your own.

- Third-party TCI datasets: Data are purchased and licensed from third-party data providers. It is the easiest solution to obtain ready-to-use TCI data. No expertise is required for data collection and processing. However, the digitized TCI types, data quality, data coverage, and data format are pre-defined by the data provider, which has the least flexibility for agencies to customize the dataset based on their data needs. The use of commercial datasets needs to follow the data provider's terms of use, which are rules, specifications, and requirements for the use of a product or service. Agencies need to carefully review and follow the terms of use when using and disseminating the data.
- Commercially available TCI digitization services: Data are collected through external companies' digitization solutions. Compared to purchasing market-available data sets, agencies have more control over defining the data collection scope, area, and format through negotiation with the digitization services providers. The dataset should be completed and delivered within an agreed period. Using the commercial digitization services may raise less license issues than using the commercial datasets; however, agencies are also suggested to negotiate and document the terms of use for the digitized data through the commercial services.
- Build on your own: Data are collected by the agencies' employees. There are two sub-types of methods that can be adopted to perform the data collection: manual data collection and automated data collection.

    - Manual data collection: Data are collected by personnel walking in the field, recording information while looking out the windshield of a vehicle, or virtually collecting data from online street view images (e.g., Google Street View). It requires little to no expertise for data collection and processing. This solution usually takes longer time and more labor to perform the data collection task. Generated data can be very detailed if necessary.

- Automated data collection: Data are collected through specially equipped vehicles at near-traffic speeds to collect asset data. Two popular techniques for establishing inventories are photogrammetry and mobile LiDAR. Typically, the sensors are mounted on the vehicle to capture various road features. Asset inventory is created by extracting data from a data processing interface usually licensed by the equipment manufacturer.

<u>Selection of data collection approaches</u>: Prior to an approach for data collection, an agency needs to comprehensively understand the characteristics and the intended function of the assets on which data are to be collected, identify fiscal or other constraints that need to be considered, and determine the maintenance approach to be used for the asset.

4.4.2.3 Data Management

<u>Data management plan</u>: “Managing the agency’s data is a shared responsibility, making collaboration across business units critical for determining not only what data is needed to manage assets but also how the data will be organized and managed. [17]” Agencies should develop a data management plan (DMP) to effectively maintain the data. The main functions or purpose of the DMP should be summarized as follows:

- To ensure data conform to standard classifications.
- To ensure validity of the data.
- To ensure data integrity and internal consistency.
- To secure and maintain primary data.
- To allow easy access to primary data.
- To process the data efficiently as required.
- To allow different data sets to be integrated, thereby increasing their overall utility.

4.4.2.4 Quality Check and Legal Issue Assessment

<u>Quality check</u>: Agencies should conduct field audits or “virtual” audits of traffic infrastructures through high-quality panoramic imagery captured at street level by Google Street View (GSV) (serving as the ground truth of TCIs) to verify the effectiveness of the adopted digitization solutions. Agencies should use specific performance metrics to describe the capacity of their solution/system for digitizing real-world traffic control infrastructures in terms of Completeness, Accuracy, Reliability, and Sensitivity (CARS). Essentially, the evaluation could be an iterative step that tests and verifies system functions and data flows for each system requirement as designed.

<u>Legal issue review</u>: Legal issues may arise from the acquisition of data for use in TCI digitization, as well as their dissemination to road users when TCI systems are implemented. These legal issues will vary depending on the type of data at use (e.g., location data,

photographic data, vehicle telematics), how the data are stored and processed, and how an agency plans to incorporate them into a publicly available database (e.g., data map, privacy impact analysis) for use by third parties and road users. In all these cases, the collection and sharing of TCI data pose legal questions about data protection, data privacy, data ownership, and agency liability, among others. Thus, whether collected directly by transportation agencies or purchased from third parties, the processes for acquiring, processing, and distributing TCI data may need to be reviewed to mitigate potential liabilities for transportation agencies, protect proprietary interests of private data owners that comprise trade secrets, and protect the privacy interests of those whose personally identifying data may be captured to create TCI.

4.4.2.5 TCI Inventory Creation

Inventory attributes: "There are some basic data attributes that are essential to adequately execute maintenance activities. Basic essential data attributes include the type of asset, location of the asset, and a unique identifier (asset ID). In a condition-based maintenance approach, there is at least one additional piece of data that is essential: asset condition. This can be collected and displayed in numerous ways. It can be one value—number, letter, word, and so on—or can be numerous values representing the different components or functionalities within the asset. Often it is simplest to designate one value representing the overall health index of the asset, which may be a culmination of smaller asset element values. [17]"

Data dictionary: Using data to manage assets requires decision makers to understand and trust the data they are using. A data dictionary defines each of the asset attributes needed. Development of a data dictionary prior to selecting a data collection approach will ensure consistency in the data collection process and can also help in agency-wide data integration.

4.4.2.6 Data Update

Inventory update: The condition of TCIs keeps changing over time. Meanwhile, the road network can also be enhanced with a lot of new TCIs installed. Therefore, there is a practical need to regularly update the existing TCI data. The general data update process can be summarized as follows. First, the current version of TCI inventory is used as a baseline inventory, which contains the old asset records in a certain study area. Meanwhile, the performing team should periodically perform data collection and generate a new inventory of TCI data. When running this method, if the targeted assets in the new inventory have the same type as the corresponding ones in the baseline inventory, they are verified to be correctly recorded in the inventory. On the other hand, for those assets that are classified as different types in the two inventories, or are classified as empty class, they would be identified as questionable assets and manual inspection would be applied in subsequent work.

4.4.2.7 Data Sharing and Visualization

Data sharing and visualization: Agencies should consider developing data visualization and sharing platforms (e.g., GIS, dashboard, data analytics) that can help with the data dissemination, planning, and investment decision.

The framework is built using ArcGIS Online, which is a cloud-based GIS platform developed by Esri (as shown in Figure 37). It allows users to create, store, share, and manage maps, data, and geospatial information and services. The platform provides a variety of tools and functionality for data analysis, mapping, and visualization, including the ability to create custom maps, perform spatial analysis, and share data and maps with others. ArcGIS Online also provides access to a large repository of geospatial data and pre-built maps and apps. The platform can be used by individuals, organizations, and government agencies to support a wide range of applications, including the TCI digitized inventory of this study.

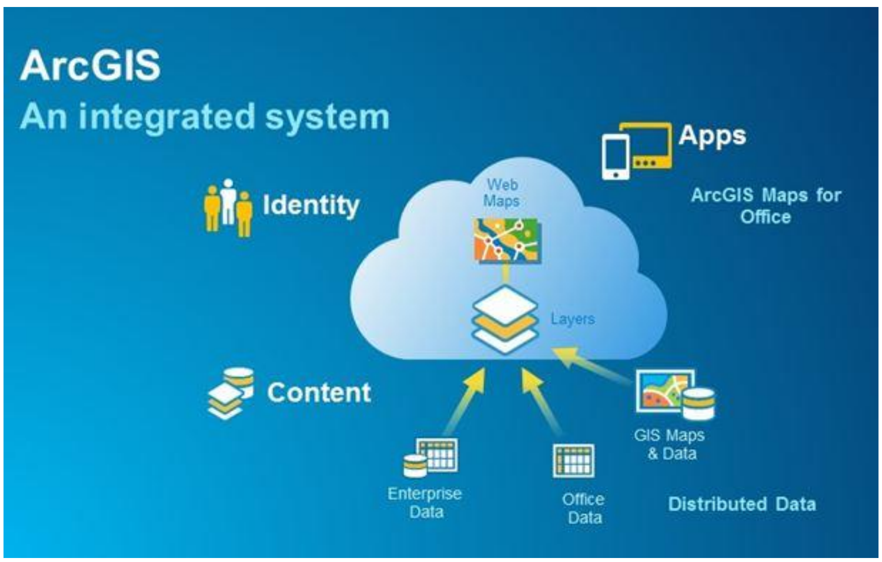


Figure 37. ArcGIS Online Platform.

Data are purchased and licensed from third-party data providers (Mobileye in this case) and then are imported into the ArcGIS Online platform as TCI digitization maps. The data are visualized on an ArcGIS Online dashboard that provides a visual representation of data and information in a web-based interface. The dashboard allows users to display, analyze, and interact with maps, charts, graphs, and other data-driven elements in a single view. Dashboards in ArcGIS Online are customizable and can be tailored to meet the specific needs of different users and organizations. The platform provides a variety of data visualization tools and widgets, such as heat maps, bar charts, pie charts, and tables, that can be used to create interactive and engaging dashboards. Additionally, users can access and display data from a variety of sources, including ArcGIS Online maps and layers, external data sources, and data stored in cloud-based platforms. The dashboards can be shared and embedded in websites and other platforms, making it easy to share data and insights with others. The agency can use this platform to create, update, maintain and manage their TCI data inventory.

4.5 DEVELOPING AV SIMULATION AND DRIVING SCENARIOS

4.5.1 Purpose

To evaluate how TCI digitization would help improve AV operations and safety performance, in Task 4 the research team simulated several driving scenarios with TCIs digitized and accessed by AVs through the Multi-Agent Based Traffic Safety Simulation System (MATISSE)—an agent-based intelligent traffic simulation system developed by the research team. This section summarizes the research team's efforts on the simulation work, including but not limited to:

- Selecting two testbed corridors.
- Developing three driving scenarios.
- Building the simulation platform and the customized AV models.
- Documenting the preliminary outputs.

4.5.2 Two Testbed Corridors

The research team built a high-level map with 3D models of the City of Dallas in the MATISSE simulation platform (as shown in Figure 38).

Figure 38. 3D Model of the City of Dallas in MATISSE.

The research team also built accurate models of AV sensing devices (see Figure 39).

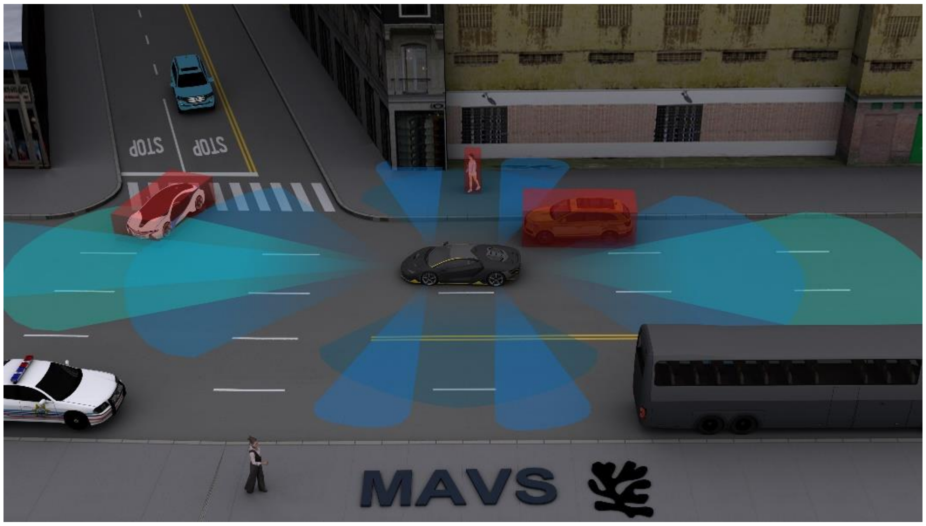


Figure 39. Simulated AV Sensors in MATISSE.

To better address the needs of this research project, the research team selected two testbed corridors in the City of Dallas:

- IH-635/LBJ East, from US75 to IH-30 (Major Work Zone Corridor).
- LP 12/Buckner Blvd., from NW Hwy to US175 (High Transit/Ped/Bike Corridor).

The selected corridors are either major commuting corridors with heavy construction activities or corridors with high-volume transit riders, pedestrians, and bicyclists. AVs have great potential to improve the operation and safety issues on these corridors. The selected corridors are shown in an ArcGIS Online map in Figure 40.

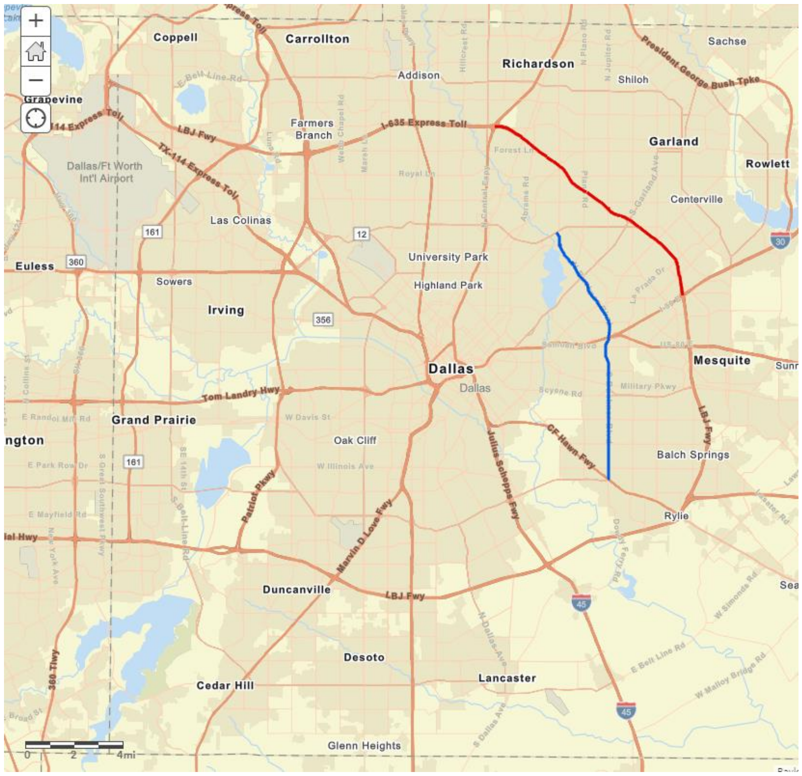


Figure 40. Two Selected Testbed Corridors.

The research team successfully developed an Open Street Map (OSM) conversion algorithm within its MATISSE system[1], which directly generates editable data layers for TCIs. Entire road networks can be imported from OSM. Several advanced algorithms have been developed to reliably convert OSM graphs to MATISSE graphs and automatically generate missing information (e.g., number of road lanes, traffic light locations, and allowable traffic movements). Figure 41a shows an OSM graph for one of the complex signalized intersections in Paris, France. In the OSM, the single signalized intersection is represented using six nodes. Micro-simulation models such as SUMO or VISSIM convert each of the OSM nodes into one signalized intersection (see Figure 41b), which results in an incorrect representation of the real network topology. As shown in Figure 41c, MATISSE's network structure and conversion algorithms allow an accurate conversion of the information.

[1] MATISSE https://www.utdmavs.org/matisse/

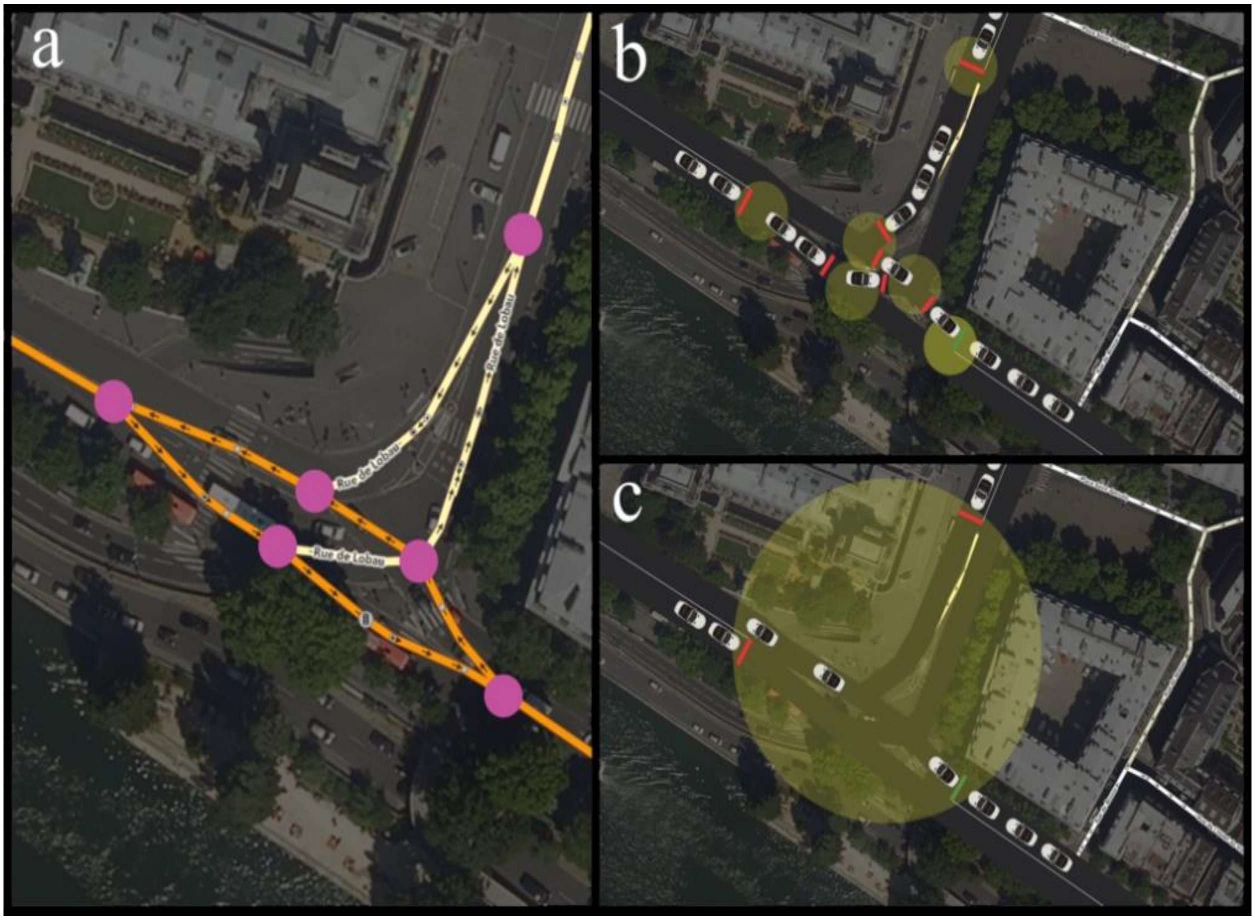


Figure 41. An Intersection and Its Representation in Open Street Map and MATISSE.

4.5.3 Three Driving Scenarios

To better demonstrate how AVs can benefit from digitized TCIs, the researchers developed three driving scenarios:

- Scenario 1: An AV uses its own perception system and its own high-definition maps to drive. In this scenario, AVs should self-drive as expected.
- Scenario 2: An AV is traveling, but its own perception system and its own high-definition map are not working. In this scenario, accidents may occur due to the malfunction of the perception system and incomplete map information.
- Scenario 3: An AV's own perception system is not working, but it can use the external map with digitized traffic control infrastructure. In this scenario, the AVs can self-drive with the help of TCI digitization maps.

4.5.4 Simulation Platform and Processes

4.5.4.1 MATISSE Platform

To evaluate the influence of digitized TCIs for AV driving, the researchers simulated the driving scenarios with TCIs digitized and accessed by AVs through MATISSE.

MATISSE[2] is a large-scale microscopic traffic simulator built from the ground up as a multi-agent system (see Figure 42). In MATISSE, vehicle drivers, AVs, and intersection controllers are modeled as virtual agents. A variety of features facilitate the simulation of TCIs and include simulated perception for virtual human drivers (e.g., vision scope, hearing); simulated sensors for virtual AVs (e.g., LiDAR, radar); simulated infrastructure-to-infrastructure (I2I), vehicle-to-vehicle (V2V), and vehicle-to-infrastructure (V2I) communications; modification of virtual vehicle and driver properties at run-time without interrupting the simulation; emergence of unanticipated (non-scripted) events (e.g., accidents) at run-time; and the automatic conversion of OSM road networks.

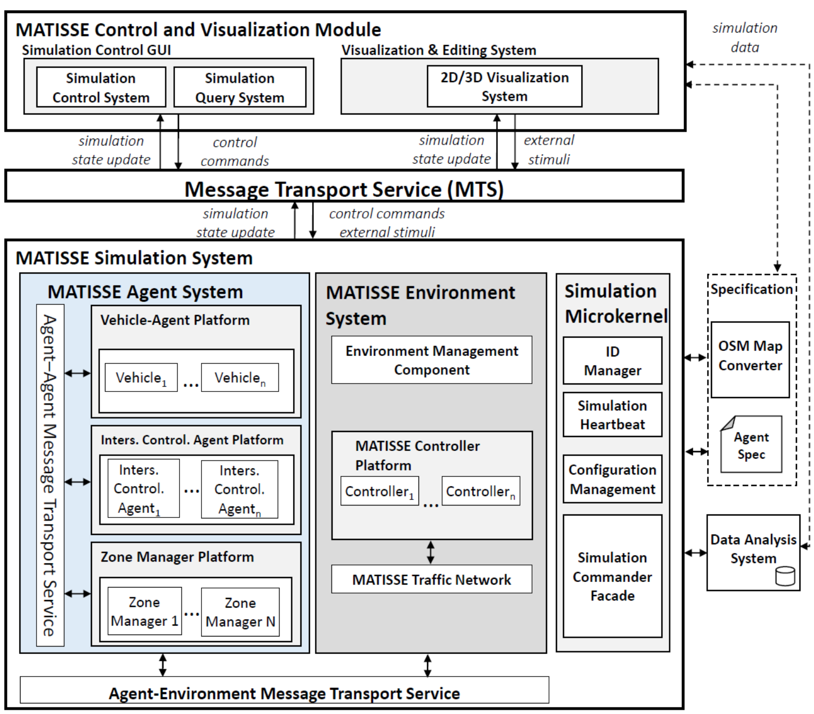


Figure 42. Conceptual Framework of MATISSE.

[2] M. Al-Zinati and R. Wenkstern, "Agent-environment interactions in large-scale multi-agent-based simulation systems," in Proceedings of the 18th International Conference on Autonomous Agents and Multi-Agent Systems, Montreal, Canada, 2019.

In addition, MATISSE allows the execution of hybrid simulations where data collected in the field (e.g., sensor, controller data) are fed into the simulator in real time. MATISSE allows for the detailed representation of complex road networks (see Figure 43) and was used to validate agent-based traffic control systems and agent-based connected and autonomous vehicle (CAV) algorithms, and is currently being used to assess the environmental impacts of CAVs.

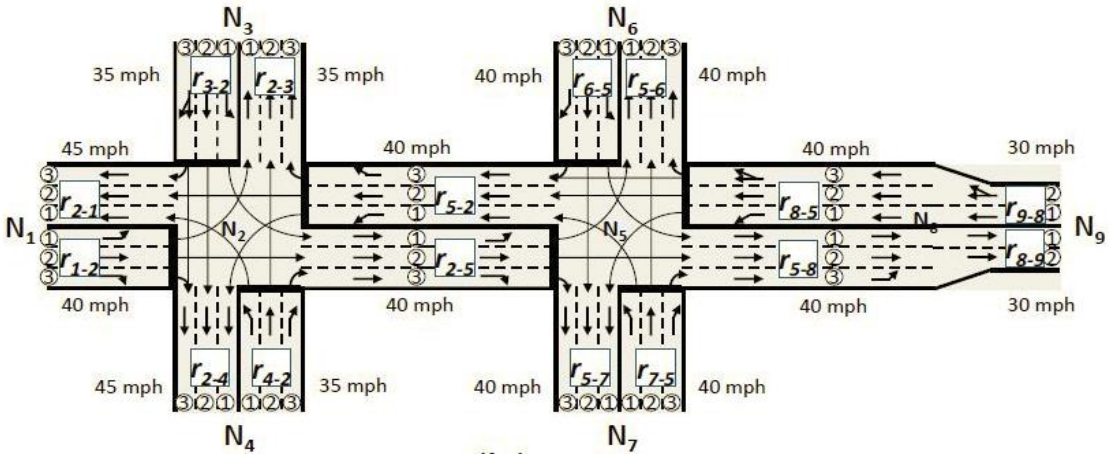


Figure 43. Traffic Network Definition in MATISSE.

An intelligent traffic management system involves a large number of interactions and collaborative decision-making among various agents. As such, supporting communication capabilities is fundamental to the implementation and validation of a simulated intelligent transportation system. MATISSE provides mechanisms for the simulation of V2V, V2I, and I2I communications (as shown in Figure 44). A simulated vehicle agent can communicate with other vehicles and intersection control agents located within its circle of influence. Similarly, an intersection controller agent can communicate with vehicle agents within its circle of influence. In addition, vehicle agents and intersection controller agents can communicate with their service managers to exchange information about the traffic environment. Figure 44 also shows these communication mechanisms.

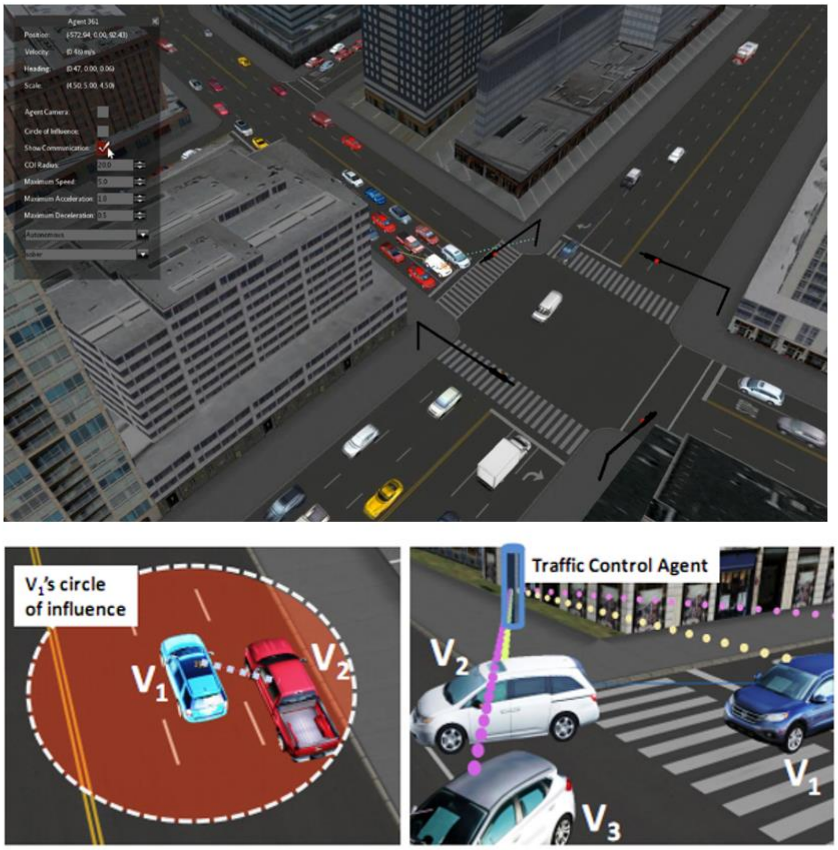


Figure 44. Simulation of V2X Communications in MATISSE.

CHAPTER 5. SYSTEM EVALUATION AND MAINTENANCE

5.1 SUMMARY

In Task 5 of this project, the research team assessed the precision of the TCI datasets by comparing them with the ground truth data, constructing a web GIS-based platform specifically designed for the digitization of TCIs, simulating multiple real-world scenarios, and analyzing the advantages of integrating TCI digitized data to improve the safety measures of AVs. The team is also creating a comprehensive DMP that ensures efficient data maintenance throughout the project.

5.1.1 Objective

The objective of this task is four-fold:

1. Evaluate the accuracy of TCI dataset from multiple sources.
2. Develop a web GIS-based TCI digitization and validation platform.
3. Simulate the impact of digitized TCI data on AVs' operations.
4. Draft a DMP to guarantee optimal data management during the project.

5.1.2 Overview of Chapter 5

This chapter presents the results of Task 5 and is organized as follows:

- Section 5.2 presents comprehensive evaluation results of the TCI datasets in the four selected corridors.
- Section 5.3 illustrates the developed web GIS-based TCI digitization/validation platform.
- Section 5.4 shows simulations on various real-world scenarios and the benefits of incorporating TCI digitized data in enhancing the operation efficiency and safety for AVs.
- Section 5.5 discusses the developed DMP that will effectively oversee data maintenance throughout the project.

5.2 TCI DATA EVALUATION AND FUSION

The research team embarked on a project to collect, assess, and digitize TCIs data from a variety of third-party sources. TCIs, which include elements such as traffic signals, road signs, and road markings, are crucial for managing and optimizing traffic flow [1]. The aim of the subtask was to evaluate the accuracy of these TCI datasets in comparison with ground truth data and establish a methodical framework for future digitization and assessment of TCIs.

Data were collected from four different sources: Mobileye, the City of Arlington, Mapillary, and OSM. The data corresponded to four distinct corridors located within the Dallas-Fort Worth metroplex area. Following the collection phase, a thorough evaluation process was conducted. The research team created a buffer zone consisting of manually identified ground truth data, which was then used as a benchmark to assess the accuracy of the collected TCI datasets.

The evaluation revealed that the data sourced from Mobileye was the most comprehensive and accurate compared to the other three datasets. By evaluating these diverse datasets, the research team gained a comprehensive view of the different types of TCIs present in various environments and understood their spatial distribution and status. The overall workflow is shown in Figure 45.

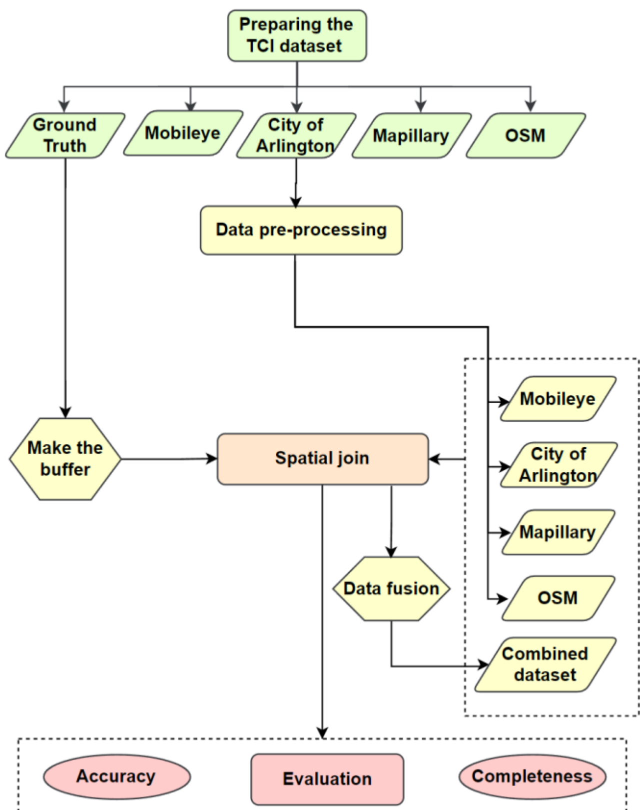


Figure 45. Workflow of TCI Data Evaluation and Data Fusion.

5.2.1 Introduction

TCIs are critical components of any transportation system. They include physical devices and systems designed to guide, warn, and regulate the flow of traffic. Examples of TCIs encompass traffic signals, traffic signs, road markings, speed bumps, traffic circles, crosswalks, and more. TCIs serve as crucial tools for managing traffic flow, enforcing traffic laws, and ensuring the safety of motorists and pedestrians. They convey important information to drivers about road conditions, speed limits, and other potential hazards, all of which help to reduce the risk of accidents. Additionally, by controlling the timing and flow of traffic, TCIs can help to improve road capacity, reduce congestion, and enhance overall traffic efficiency.

As urban populations grow and transportation needs evolve, there is an ever-increasing demand for accurate and comprehensive data on TCIs. To effectively manage and optimize urban traffic, city planners, traffic engineers, and policy makers need precise data on the location, type, and condition of TCIs. This data can help identify infrastructure needs, plan road improvements, develop intelligent transportation systems, and evaluate traffic policies [2].

The task of this study aims to obtain digitized sample datasets of TCIs from multiple third-party data providers. These providers can range from commercial companies and government agencies to open-source platforms. By obtaining these diverse datasets, the research team hopes to gain a comprehensive view of the different types of TCIs present in various environments and understand their spatial distribution and status.

Following data acquisition, the team conducted a rigorous evaluation of these datasets. The team assessed their completeness, accuracy, and relevance to the traffic management objectives. A critical part of this evaluation involves comparing the digitized data with ground truth data to determine its reliability and accuracy.

Finally, the research team built a robust framework for digitizing TCIs and evaluating the resulting datasets. This framework provides a guidance for the data collection and evaluation efforts, ensuring the team produces reliable, accurate, and up-to-date TCI data that can inform the traffic management strategies. The research team's ultimate goal is to leverage this data to create safer, more efficient, and more sustainable urban transportation systems (especially improving the operation of AVs).

5.2.2 Selected Data Sources for TCI Digitization

5.2.2.1 Purpose

This sub-task aims to look further into the potential data sources and solutions for TCI digitization that were identified in Task 4. The selected data sources come from:

- Leading commercial company: Mobileye.
- Government agency: City of Arlington.
- Open source: Mapillary and OSM.
- Ground truth data: manually identified by researchers.

The selection of these four sources is also dependent on their willingness to collaborate with the research team and provide the team with sample TCI digitized datasets for the evaluation purposes.

5.2.2.2 Mobileye

Mobileye, an Intel company, is a global leader in the development of computer vision and machine learning, data analysis, localization, and mapping for advanced driver assistance systems and autonomous driving solutions [3].

Mobileye technology uses a single camera to scan the road ahead and detect potential dangers. Leveraging the same technology, Mobileye has reinvented road mapping for asset management and road maintenance. Mobileye's EyeQ® chip and algorithms have been trained to identify, tag, and classify road assets, detect pavement anomalies, and capture mobility data, all as equipped vehicles travel on their regular routes. Additional road data features are collected through Mobileye 8 Connect™–equipped vehicles. Fleet vehicles retrofitted with Mobileye 8 Connect, a collision avoidance system and data collection enabler, collect information as field workers drive from location to location during their workday.

The road asset and mobility information captured by the vehicles is processed and turned into GIS data layers, updated at a relatively high refresh rate (as shown in Figure 46). The geolocation of each asset or road feature is derived from aggregating detections from multiple drives for higher accuracy. "Locations" are reported as latitude, longitude, and altitude coordinates. The following list of features is summarized in Table 1, which is regularly revised to incorporate new releases from Mobileye.

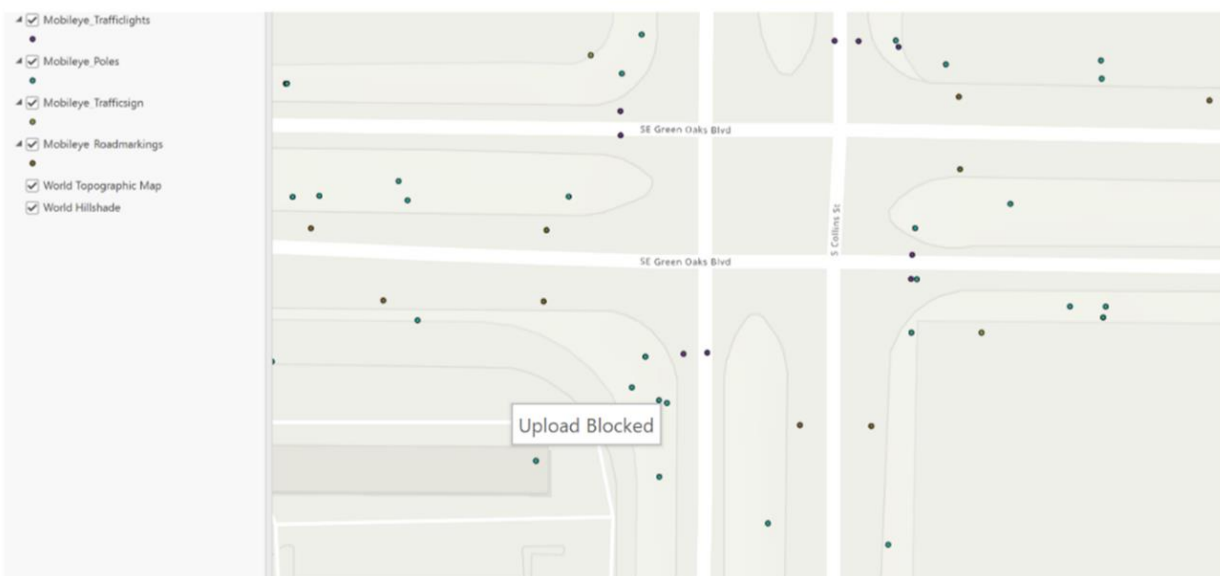


Figure 46. Screenshot of Mobileye's Sample TCI Data.

5.2.2.3 City of Arlington

Arlington is a city in the state of Texas, located in Tarrant County. It forms part of the Mid-Cities region of the Dallas–Fort Worth–Arlington metropolitan statistical area and is a principal city of the metropolis and region. The city has a population of 394,266 in 2020, making it the second-largest city in the county, after Fort Worth, and the third-largest city in the metropolitan area, after Dallas and Fort Worth. Arlington is the 50th-most populous city in the United States, the seventh-most populous city in the state of Texas, and the largest city in the state that is not a county seat [4].

The Arlington Open Data Portal serves as the city's central clearinghouse for accessing, visualizing, and interacting with Arlington's open data sets. Users can download City data in geospatial and table formats, explore data through interactive maps, and connect to data to better understand the community (Figure 47).

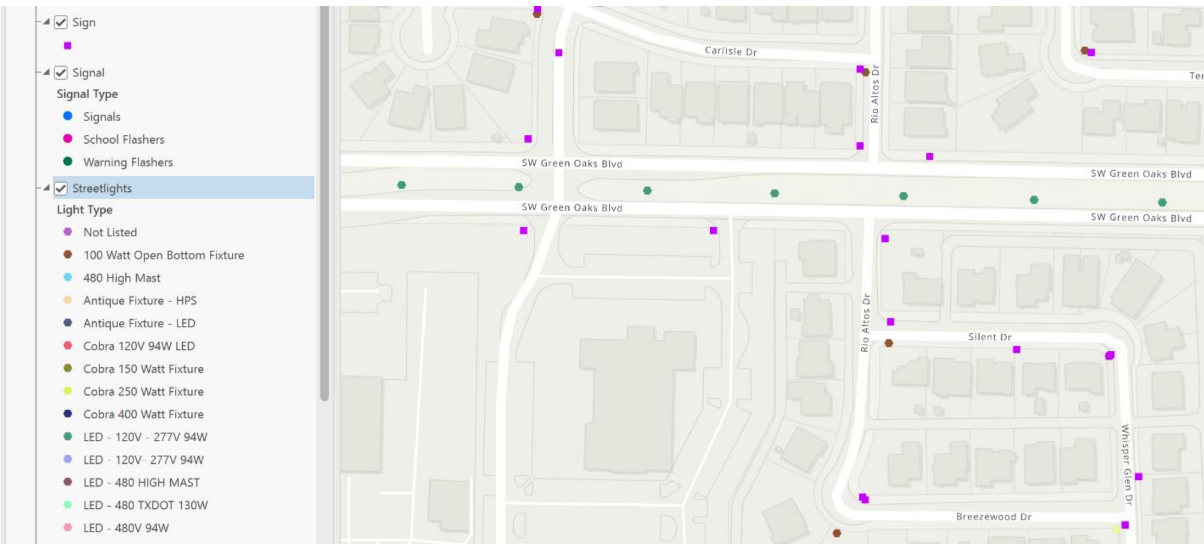


Figure 47. Screenshot of City of Arlington Data.

5.2.2.4 Mapillary

Mapillary is a service for sharing crowdsourced geotagged photos. The company was founded in 2013 and later acquired by Meta in 2020. It allows individuals and companies to share and use street-level images and map data from all around the world [5].

Users capture and upload geotagged photos or images to the Mapillary platform from a variety of sources such as smartphones, action cameras, and 360-degree cameras. This can be achieved by taking photos manually or by using the Mapillary mobile app, which has a mode for automatic capture of images. Dashcams and other automated devices can also be used. The uploaded photos are then placed on a map using their geotagging data. Once the images are on the platform, they undergo a computer vision analysis. This process identifies and categorizes objects within the images, such as roads, traffic signs, buildings, etc., enhancing the map data (Figure 48).

Mapillary's platform and services have been particularly useful for improving and supplementing maps in areas where detailed mapping data may not be available or are out of date. They have also been used in combination with machine learning to better understand the physical world's layout and to aid in the development of AVs.

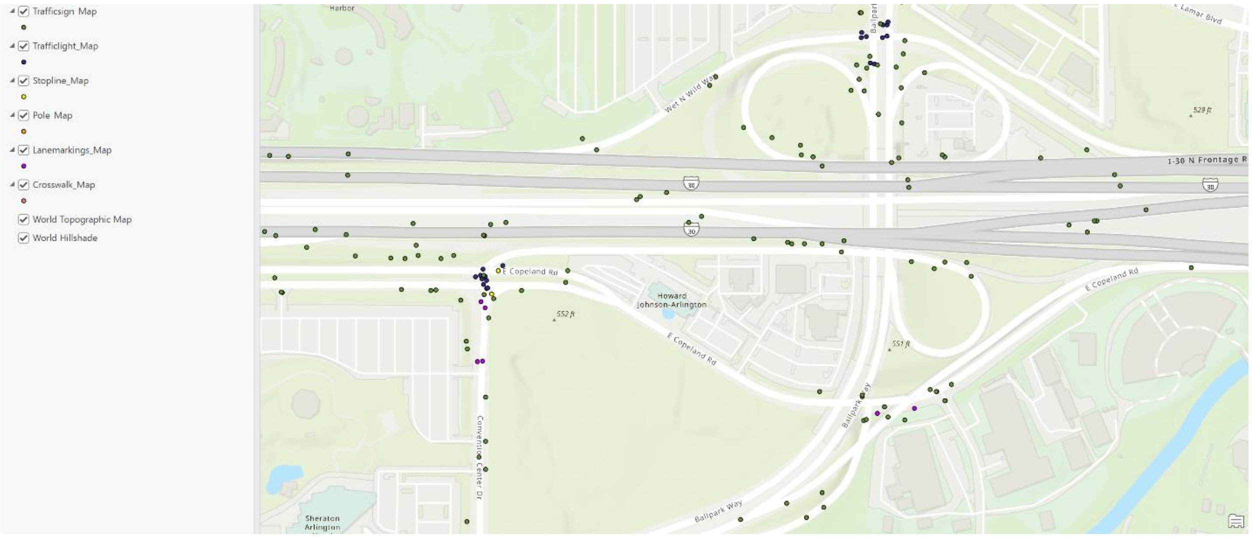


Figure 48. Screenshot of Mapillary Data.

5.2.2.5 OpenStreetMap

OpenStreetMap is a collaborative project to create a free editable map of the world. It was founded in 2004 in the United Kingdom by Steve Coast, and it is often compared to Wikipedia but for maps, due to its crowdsourced nature [6].

OSM data is collected from a variety of sources (as shown in Figure 49):

- Volunteer Surveyors: Much of the data in OSM are collected by volunteers who walk, bike, drive, or hike with a GPS device or a smartphone to record geographic data. They then upload these data to the OSM database.
- Importing Existing Data: Where it is available and compatible with OSM's license, existing geographic data (from government databases, for example) can be imported into the OpenStreetMap database.
- Aerial Imagery: Contributors use aerial imagery (like from satellites or airplanes) as a base for drawing maps. Companies like Microsoft and Mapbox, as some government agencies, have provided such imagery for OSM use.
- User-Generated Content: Through editing software such as iD Editor (an in-browser map editor developed by the OpenStreetMap community) or Java OpenStreetMap Editor (JOSM), contributors can manually add or modify features such as roads, buildings, parks, and much more based on their local knowledge or using other referenced data sources.

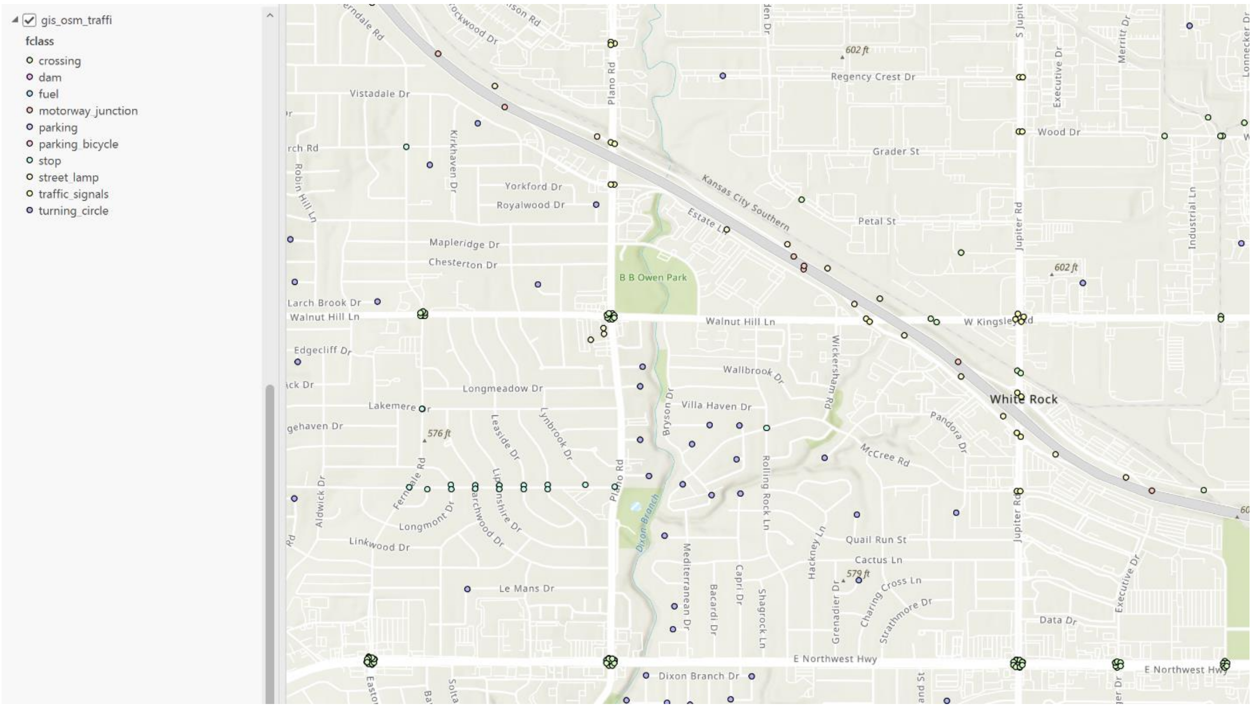


Figure 49. Screenshot of OSM Data.

5.2.3 Data Exploration Analysis

5.2.3.1 Purpose

This section presents the sample third-party data obtained by the research team and pre-processing results. The research team extracted and cleaned data on the four testing corridors: I-30 Corridor (Arlington), Green Oak Blvd (Arlington), IH-635/LBJ East (Dallas), and LP 12/Buckner Blvd (Dallas). Also, the research team identified the ground truth data manually as the benchmark dataset for further analysis.

5.2.3.2 Selected Study Area

The research team focused on four corridors for the evaluation purpose. The corridors include (as shown in Figure 50):

- City of Arlington: I-30 Corridor, from Sandy Ln. to SH360 (Major Freeway).
- City of Arlington: Green Oak Blvd., from I-20 to SH360 (Arterial Corridor).
- City of Dallas: IH-635/LBJ East, from US75 to IH-30 (Major Work Zone Corridor).
- City of Dallas: LP 12/Buckner Blvd., from NW Hwy to US175 (High Transit/Ped/Bike Corridor).

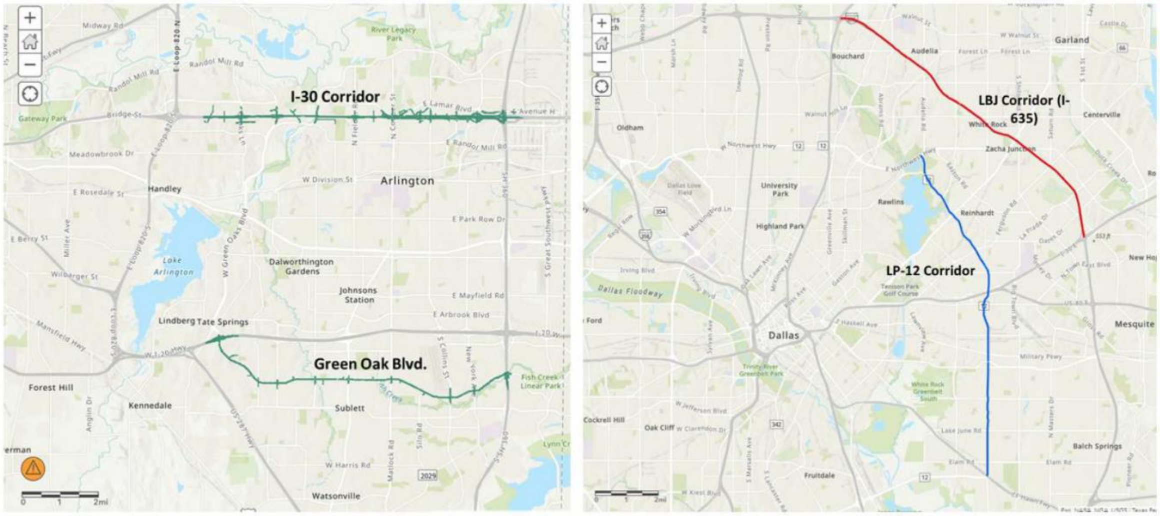


Figure 50. Selected Sample Corridors in the DFW Area.

5.2.3.3 Mobileye Dataset

Mobileye's technology is designed to provide vehicles with the ability to detect and understand the world around them, and their systems generate a variety of traffic-related data (Table 9):

- Roadway Markings and Geometry: Mobileye's technology can detect and interpret road features like lane markings, road edges, barriers, and the road's curvature. This can provide important context for understanding a vehicle's current location and planning its path.
- Traffic Signs: Mobileye systems are capable of recognizing a wide range of traffic signs, such as stop signs, speed limit signs, yield signs, and more. Understanding traffic signs is crucial for complying with traffic rules.
- Traffic Signals: Mobileye can also detect traffic signals and determine their current state (e.g., red, yellow, green). This is another critical aspect of understanding and complying with traffic rules.
- Vehicles and Pedestrians: Mobileye's technology is designed to detect other vehicles on the road, as well as pedestrians and cyclists. This is essential for avoiding collisions and driving safely.
- Obstacles and Hazards: Mobileye systems can identify potential obstacles and hazards on the road, including debris, potholes, and other dangers that might require the vehicle to take evasive action.
- Free Space and Path Planning: The systems can identify the drivable area or free space around the vehicle, which is important for path planning and navigation, especially in complex environments.
- 3D Mapping Data: Mobileye's Road Experience Management (REM) technology generates detailed and up-to-date 3D maps using data collected from vehicles equipped

with its technology. These maps can provide valuable context for AVs and driver assistance systems.

Table 9. Selected Dataset—Mobileye.

Category	Count	Category	Count
Bridge Ice	32	Reverse Curve Right 2 Lanes	3
Children	27	Reverse Curve Right 3 Lanes	11
Construction Area Ahead	26	Right Merge	33
Construction Work for X Miles	1	Road May Flood	2
Curve Left	4	Road Narrows Left	9
Dead-End	1	Road Narrows Right	9
Directional Sign	630	Slippery When wet	1
End Road Works	9	Speed Limit 20	13
End School Zone	10	Speed Limit 25	7
Fire Station	2	Speed Limit 30	7
Intersection	2	Speed Limit 35	2
Left Merge	7	Speed Limit 40	60
Low Clearance	18	Speed Limit 45	47
No Entrance	23	Speed Limit 50	6
No Left Arrow Sign	6	Speed Limit 60	21
No Outlet	3	Speed Limit 70	18
No Parking	19	Speed Limit 75	8
No Passing Start	2	Speed Limit 80	2
No Right Arrow Sign	4	Stop Ahead	7
No Turn on Red	1	Stop Sign	30
No U Turn	6	Thru Traffic Merge Left	1
Pedestrian Limit	15	Traffic Signals Ahead	27
Pedestrians Crossing	13	Truck Limit	4
Post Exit	39	Trucks Crossing	8
Reflector	101	Two Way Traffic	1
Reverse Curve Left 2 Lanes	2	Winding Left	1
Reverse Curve Left 3 Lanes	12	Yield	32
Reverse Curve Right	6	**Grand Total**	**1391**

5.2.3.4 City of Arlington

The traffic data provided by the city of Arlington include the following data (Table 10):

- Traffic Sign: Traffic signs are crucial for road safety, providing important information to drivers and pedestrians. They include stop signs, yield signs, speed limit signs, directional signs, and many others. The data associated with traffic signs can include their type, location, and condition.

- Traffic Lights: The dataset includes information about the locations and types of traffic lights.
- Pole: poles can refer to various vertical structures such as light poles, utility poles, traffic signal poles, or sign poles. Information on poles can include their location, the type of pole, and what equipment it is carrying.
- Road Marking: Road markings provide guidelines and information to drivers and pedestrians. They include lane dividers, stop lines, crosswalk markings, directional arrows, and more.

Table 10. Selected Dataset—City of Arlington.

Category	**Count**	**Category**	**Count**
Poles	**1727**	Construction Area Ahead	1
Electricity Pole	252	Curve Left	1
General Pole	399	Directional Sign	90
Street Light Pole	409	End School Zone	9
Traffic Light Pole	29	Fire Station	2
Traffic Sign Pole	364	Intersection	2
Tree Trunk	274	No Entrance	10
Road marking	**150**	No Outlet	3
Bicycle Lane Road Mark	1	No Parking	1
Left Right Road Arrow	1	No Right Arrow Sign	1
Left Road Arrow	86	No U Turn	1
Right Road Arrow	12	Pedestrian Limit	1
Road Text Only	34	Pedestrians Crossing	2
Straight Right Road Arrow	2	Post Exit	1
Straight Road Arrow	14	Reflector	14
Traffic lights	**218**	Speed Limit 20	9
Horizontal 3 Spots Traffic Light	146	Speed Limit 30	1
Horizontal 4 Spots Traffic Light	12	Speed Limit 40	4
Pedestrian Horizontal Traffic Light	32	Speed Limit 45	31
Pedestrian Vertical Traffic Light	7	Stop Sign	4
Vertical 3 Spots Traffic Light	20	Traffic Signals Ahead	19
Vertical 5 Spots Traffic Light	1	Truck Limit	4
Traffic signs	**246**	Yield	14
Children	21	**Grand Total**	**2341**

5.2.3.5 Mapillary

Mapillary offers a variety of traffic-related data that have been extracted from the image data. The service provides insights on physical attributes of the road environment that could be crucial for traffic management, infrastructure planning, and AV training algorithms. Below are some types of traffic-related data in Mapillary (Table 11):

- Traffic Signs: Mapillary uses machine learning algorithms to detect and classify different types of road signs from their street-level images. This includes stop signs, yield signs, speed limit signs, and other regulatory signs.
- Lane Markings: Information about the number of lanes on a road, lane directions, and other road surface markings can also be extracted from the images.
- Traffic Lights: The dataset includes information about the locations and types of traffic lights.
- Crosswalks: Crosswalks, their locations, and characteristics are identified and classified in the dataset.
- Stop Lines: Stop line is a white line painted on the road indicating where vehicles should stop in response to a stop sign or traffic signal. Stop lines are critical for understanding the layout of intersections and determining where vehicles are expected to stop.
- Poles: The dataset includes information about poles; it could include light poles, utility poles, traffic signal poles, or sign poles. Each type of pole has different implications for traffic patterns and infrastructure planning.

Table 11. Selected Dataset—Mapillary.

Category	Count	Category	Count	Category	Count	Category	Count
Crosswalk	**36**	Maximum Speed Limit 10	3	Maximum Speed Limit 60	1	Turn Left Ahead	6
Crosswalk Plain	36	Maximum Speed Limit 15	33	Obstacle Delineator	3	Turn Left	16
Lanemarking	**132**	Maximum Speed Limit 20	3	One Direction Right	1	Turn Left Or Right	3
Discrete Arrow Left	27	Maximum Speed Limit 30	2	Time Restrictions	3	Turn Right	14
Discrete Arrow Right	5	Maximum Speed Limit 35	3	Turn Left	1	U Turn	7
Discrete Arrow Split Left Or Straight	6	Maximum Speed Limit 40	49	Weight Limit	10	Vehicles Carrying Hazardous Goods Permitted	21
Discrete Arrow Split Right Or Straight	3	Maximum Speed Limit 45	8	Width Limit	2	Wrong Way	2
Discrete Arrow Straight	14	Maximum Speed Limit 50	5	Airport	2	Yield	17
Discrete Other Marking	40	Maximum Speed Limit 55	1	Dead End Except Bicycles	2	Added Lane Left	8
Discrete Symbol Bicycle	2	Maximum Speed Limit 60	7	End Of Advisory Maximum Speed Limit 60	6	Added Lane Right	43
Discrete Text	35	Maximum Speed Limit 70	49	End Of Built Up Area	4	Bicycles Crossing	1
Pole	**248**	Maximum Speed Limit 75	12	Highway Exit	62	Crossroads	1
Support Pole	248	Maximum Speed Limit 110	2	Highway Interchange	818	Curve Left	17
Stopline	**40**	No Entry	34	Highway Reference Location	13	Curve Right	7
Discrete Stop Line	40	No Left Turn	16	Hospital	7	Double Curve First Left	2
Trafficlight	**311**	No Motor Vehicles	1	Interstate Route	52	Double Curve First Right	5
Traffic Light General Horizontal	187	No Parking	15	Minimum Speed 30	15	Hairpin Curve Right	6
Traffic Light General Single	1	No Pedestrians	19	Minimum Speed 60	6	Height Restriction	69
Traffic Light General Upright	84	No Right Turn	4	Motorway	1	Pedestrians Crossing	18
Traffic Light Other	4	No Turn On Red	1	Tourism	9	Railroad Crossing	1
Traffic Light Pedestrians	35	No Turns	1	Train Or Light Rail Station	11	Road Narrows Left	3
Trafficsign	**2266**	No U Turn	6	Turn Left	2	Road Narrows Right	1
Advisory Exit Or Ramp Speed	1	No Vehicles Carrying Dangerous Goods	2	Do Not Block Intersection	2	School Zone	6
Chevron Left	12	One Way Left	37	Do Not Pass	4	Slippery Road Surface	1
Chevron Right	13	One Way Right	49	Dual Lanes All Directions On Left	1	Soft Shoulder	2
Dead End	40	One Way Straight	1	Dual Lanes All Directions On Right	1	Stop Ahead	18
Detour	1	Pedestrians Push Button	5	Dual Lanes Go Left Or Right	3	Texts	149
Distance	4	Priority Road	15	Dual Lanes Go Straight On Left	5	Traffic Merges Left	15
Go Left	2	Road Closed	4	Dual Lanes Go Straight On Right	13	Traffic Merges Right	31
Go Right	5	Sidewalk Closed	1	Dual Lanes Turn Left	4	Traffic Signals	6
Keep Left	2	Speeding Fines Increased	5	Dual Lanes Turn Left No U Turn	2	Triple Curve Left	1
Maximum Speed Limit 15	4	Stop	54	Dual Lanes Turn Left Or Straight	19	Triple Curve Right	7
Maximum Speed Limit 20	6	Stop Here On Red Or Flashing Light	1	Go Straight	3	T Roads	18
Maximum Speed Limit 30	11	Traffic Signal Photo Enforced	5	Go Straight Or Turn Left	3	Trucks Rollover	3

Maximum Speed Limit 35	1	Trams And Buses Only	1	Go Straight Or Turn Right	2	Two Way Traffic	1
Maximum Speed Limit 40	3	Triple Lanes Go Straight Center Lane	6	Keep Right	13	Winding Road First Left	3
Maximum Speed Limit 45	21	Triple Lanes Turn Left Center Lane	3	Lane Control	10	Yield Ahead	1
Maximum Speed Limit 50	9	Triple Lanes Turn Right Center Lane	2	Left Turn Yield On Green	20	**Grand Total**	**3033**

5.2.3.6 OSM Dataset

OSM has related traffic data including crossing, dam, fuel, motorway junction, parking, bicycle parking, stop sign, streetlamp, traffic signal, and turning circle (Table 12). These are described in the following:

- Crossing: OSM contains crossing information. This could be an uncontrolled crossing (without traffic signals) or a controlled crossing (with traffic signals, pedestrian lights, etc.).
- Motorway Junction: A motorway junction is a location where traffic can move on or off a motorway.
- Stop Sign: In OSM, stop signs are usually tagged on the node where the sign is located and often on the way (the road) that it applies to as well.
- Streetlamp: OSM usually contains streetlamp information; they often are placed along the roadside.
- Traffic Signal: Traffic signals are usually represented as nodes in OSM at the locations where the traffic lights are installed.

Table 12. Selected Dataset—OSM.

Category	**Count**
Crossing	223
Motorway Junction	51
Stop	3
Streetlamp	30
Traffic Signals	155
Grand Total	**462**

5.2.3.7 Ground Truth Data

The research team checked the TCI data manually to get the ground truth dataset. Since Mobileye data were collected recently with relatively higher accuracy than the other three datasets, the research team used it as the basis for manual identification. In the current study phase, the research team exclusively examined speed limit signs within their designated area, and further investigations are underway to explore additional types of TCI.

The research team utilized Google Earth Pro to import the Mobileye speed limit data and cross-reference them with ground truth speed limit signs, which were identified through the examination of Google Earth Street View (see Figure 51 and Figure 52). Out of the 191 Mobileye speed limit signs, a manual review confirmed that 161 of them were accurate representations, resulting in an overall accuracy rate of 84.29 percent for the Mobileye data. The team identified a common reason for inaccuracies: Mobileye often classifies road signs with numbers as speed limit signs. By excluding this particular type of inaccuracy, the accuracy of the Mobileye data would rise to 91.62 percent for the tested corridors.

Figure 51. Extracting Ground Truth Data Manually from Street View Image on Google Earth.

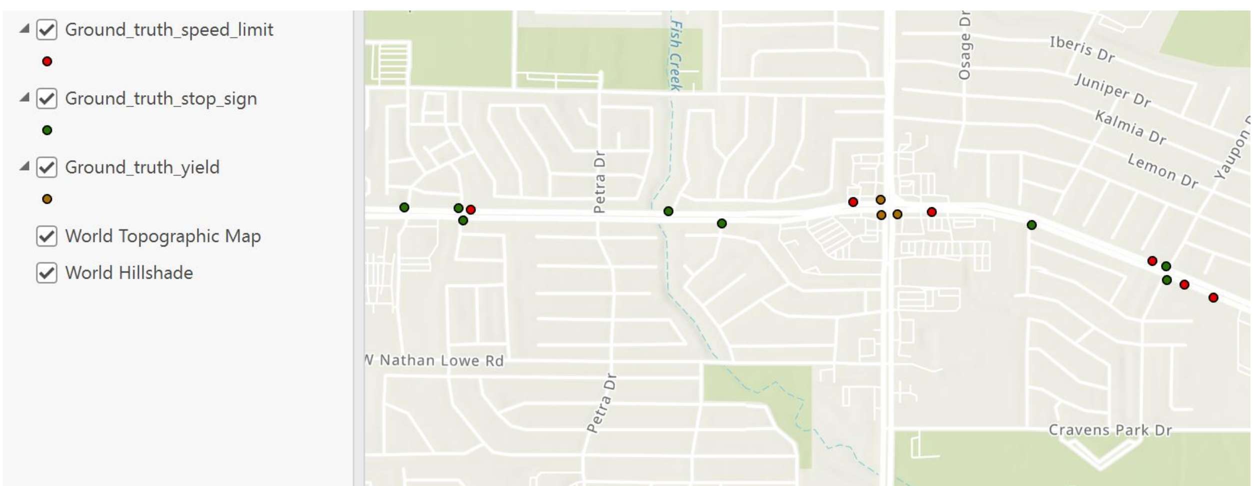


Figure 52. Screenshot of Ground Truth Data.

5.2.3.8 Combined Dataset after Data Fusion

The research team hoped to improve the data reliability by combing multiple datasets. In order to reduce redundant data from multiple datasets after combing, Density-Based Spatial Clustering of

Applications with Noise (DBSCAN) was applied to cluster these datasets. DBSCAN is a popular clustering algorithm used as an alternative to K-means. It does not require the user to set the number of clusters a priori, and it can capture clusters of complex shapes compared to K-means, which usually captures spherical clusters [7]. A combined dataset was created for each kind of traffic dataset. DBSCAN can be applied to cluster objects over both space and time when the time component can be considered as an additional dimension. For the combine dataset of speed limit sign, because it also had different speed limit values, the research team considered the number of speed limit as time attribute in the study and clustered speed limit data to reduce redundant and repeated data.

5.2.4 Final TCI Datasets

5.2.4.1 Speed Limit Sign Datasets

The research team extracted and cleaned speed limit data from four sources for analysis. For the Mapillary Dataset, there were some speed limit signs remarkably close to each other, which looked redundant and looked like one speed limit sign instead of multiple ones. In order to improve the accuracy for data fusion, the research team applied DBSCAN to the cluster Mapillary dataset before processing it with other datasets.

For speed limit sign, the research team considered the number of speed limit as time attribute in the study and cluster Mapillary data to reduce redundant and repeated data points.

After data extraction and cleaning, the research team obtained the following datasets (Figure 53 and Figure 54):

- Mobileye: 191 (in all four corridors).
- Arlington: 37 (in the Green Oaks Blvd).
- Mapillary: 130 (in I-30, the Lyndon B Johnson Fwy, and the Buckner Blvd).
- OSM: 5 (in in I-30, the Lyndon B Johnson Fwy).
- Ground truth data: 161 (in all four corridors).

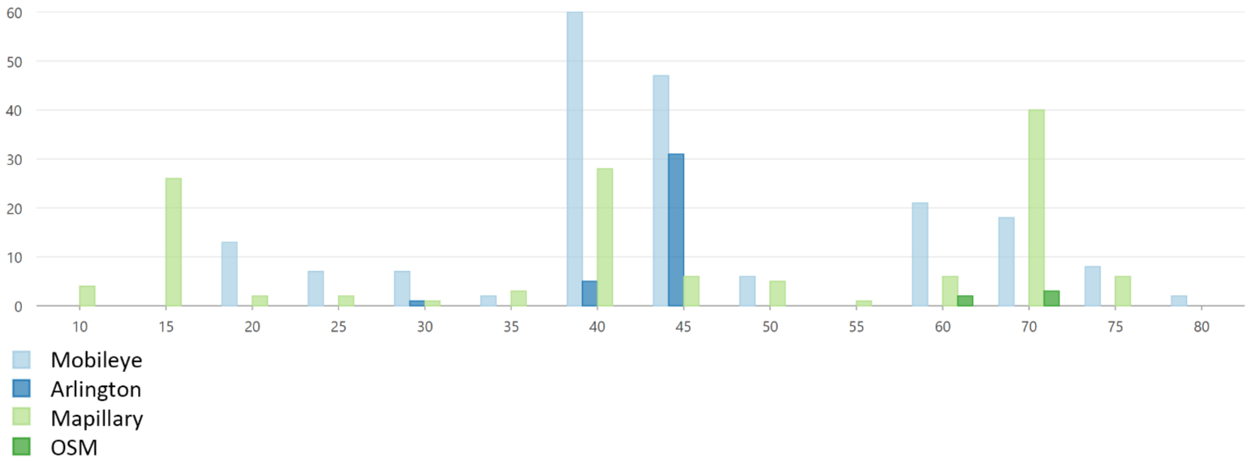
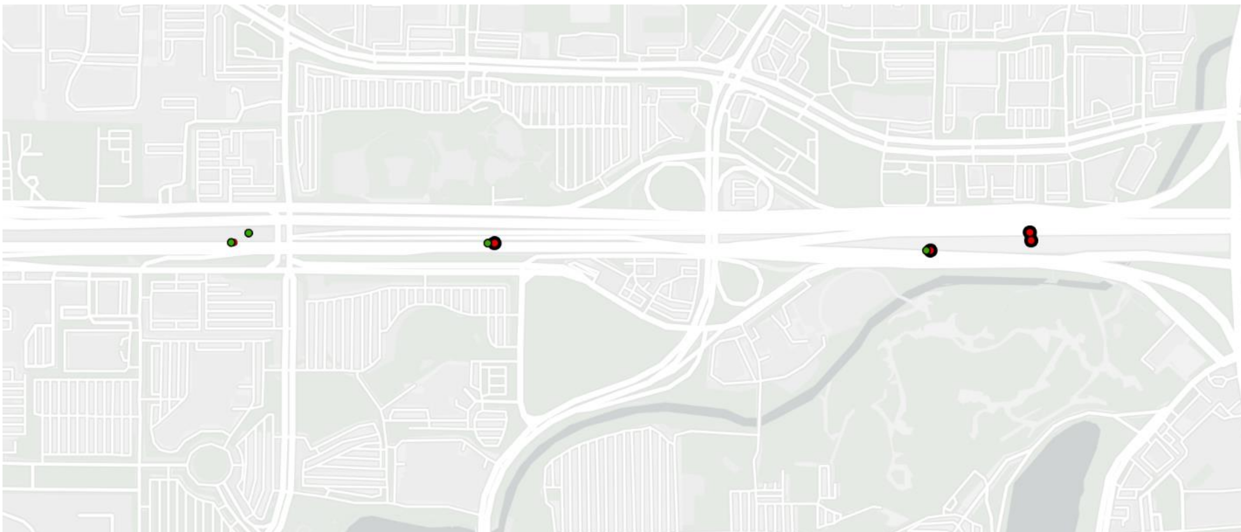
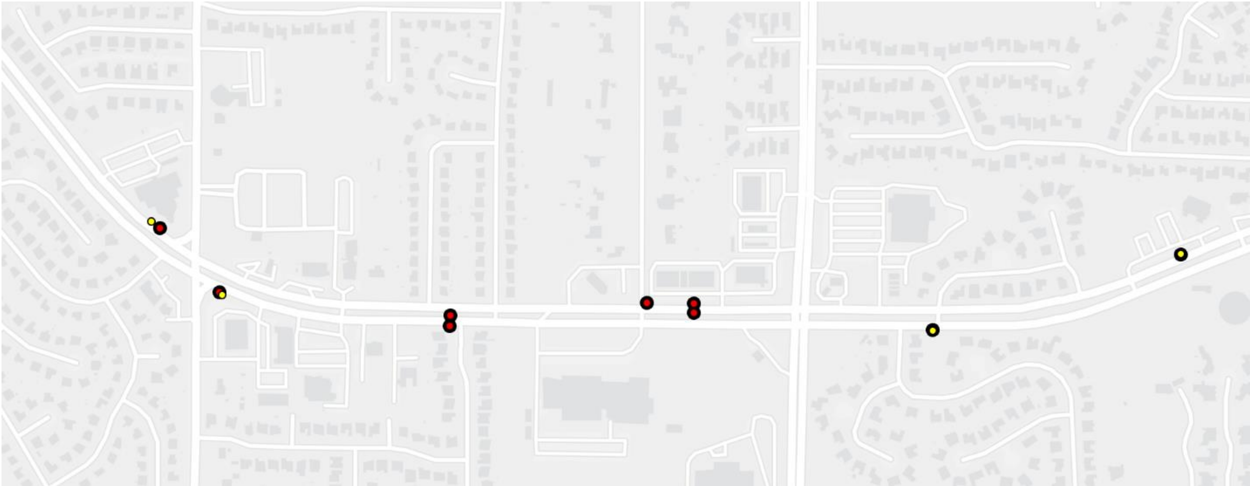


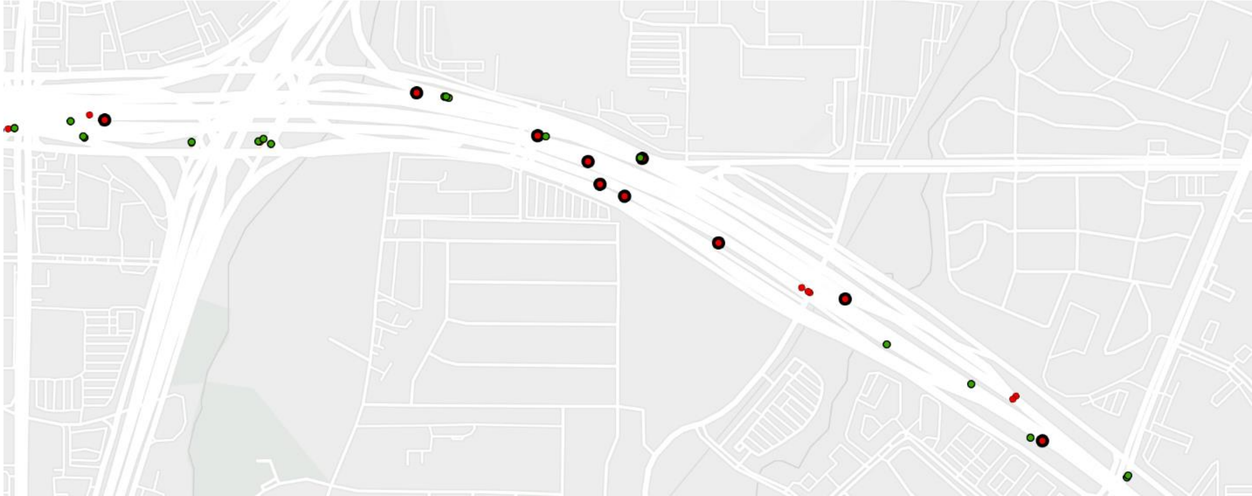
Figure 53. Number of Speed Limited Signs of Mobileye, City of Arlington, Mapillary, and OSM.



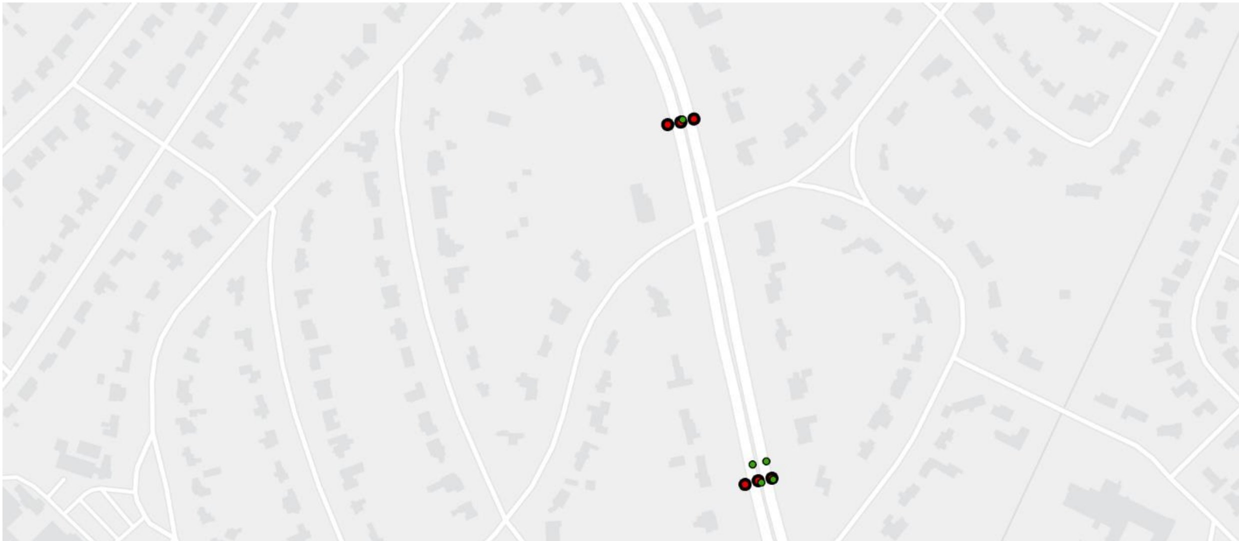
a. Comparison of speed limit sign data on I-30, Arlington, TX.



b. Comparison of speed limit sign data on Green Oaks Blvd., Arlington, TX.



c. Comparison of speed limit sign data on LP 12/Buckner Blvd., Dallas, TX.



d. Comparison of speed limit sign data on IH-635/LBJ East, Dallas, TX.

Speed limit sign legend

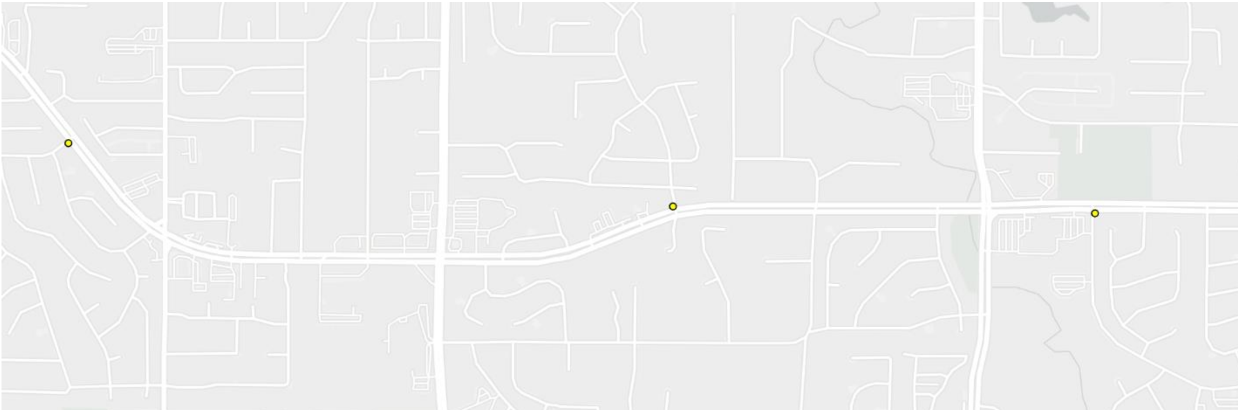
- Mobileye
- Arlington
- Mapillary
- OSM
- Ground Truth

Figure 54. Comparison of Speed Limit Sign Data after Cleaning and Pre-processing.

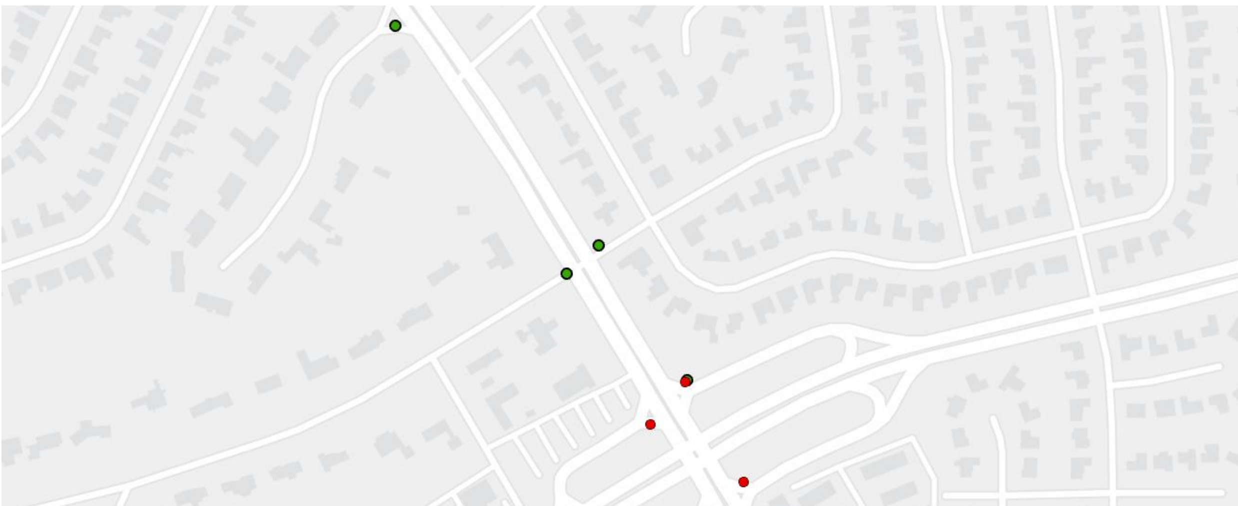
5.2.4.2 Stop Sign Datasets

The research team extracted and cleaned yield data. No stop signs were founded in all datasets in I-30, so the team only evaluated yield signs in Green Oaks Blvd, Lyndon B Johnson Fwy, and Buckner Blvd. The research team had the following datasets (Figure 55):

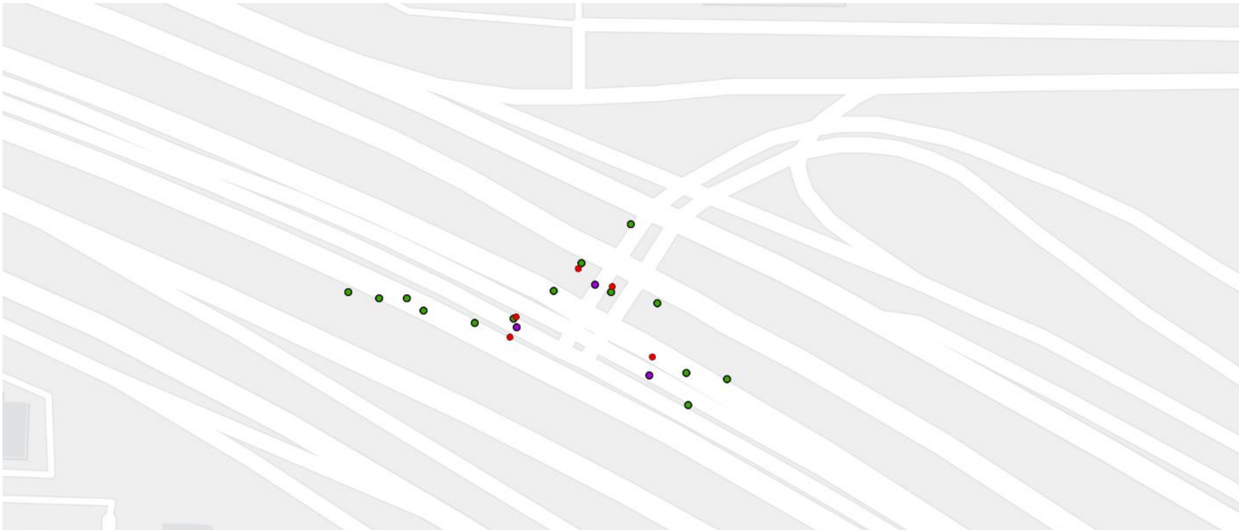
- Mobileye: 30 (in Green Oaks Blvd, Lyndon B Johnson Fwy and Buckner Blvd).
- Arlington: 4 (in Green Oaks Blvd).
- Mapillary: 56 (in Lyndon B Johnson Fwy and Buckner Blvd).
- OSM: 3 (in Lyndon B Johnson Fwy).
- Combined dataset after data fusion: 88 (in Green Oaks Blvd, Lyndon B Johnson Fwy and Buckner Blvd).
- Ground truth data: 139 (in Green Oaks Blvd, Lyndon B Johnson Fwy and Buckner Blvd).



a. Comparison of stop sign data on Green Oaks Blvd., Arlington, TX.



b. Comparison of stop sign data on LP 12/Buckner Blvd., Dallas, TX.



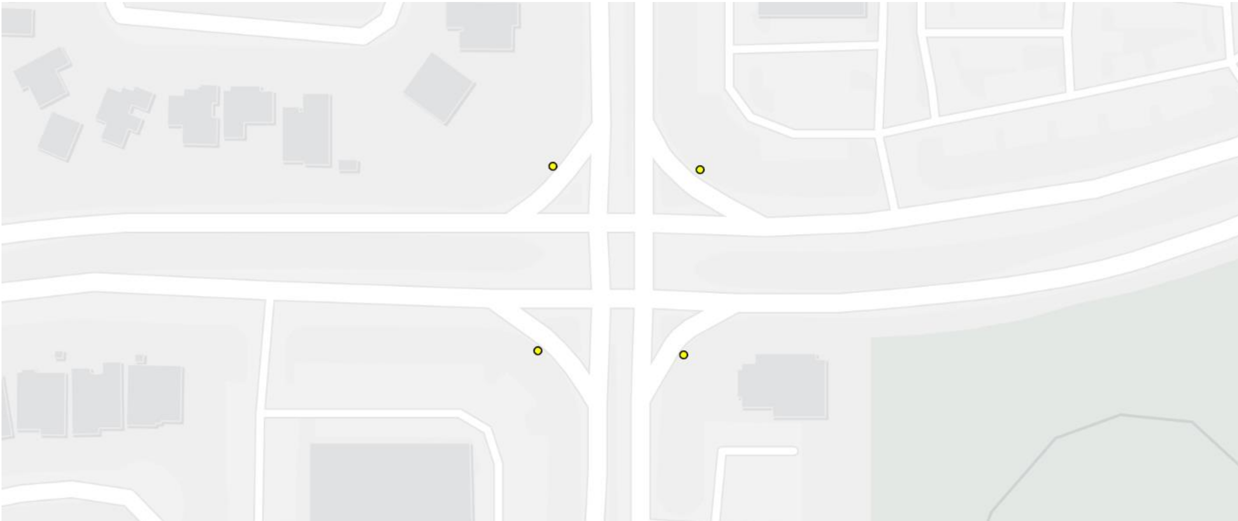
c. Comparison of stop sign data on IH-635/LBJ East, Dallas, TX.

Figure 55. Comparison of Stop Sign Data.

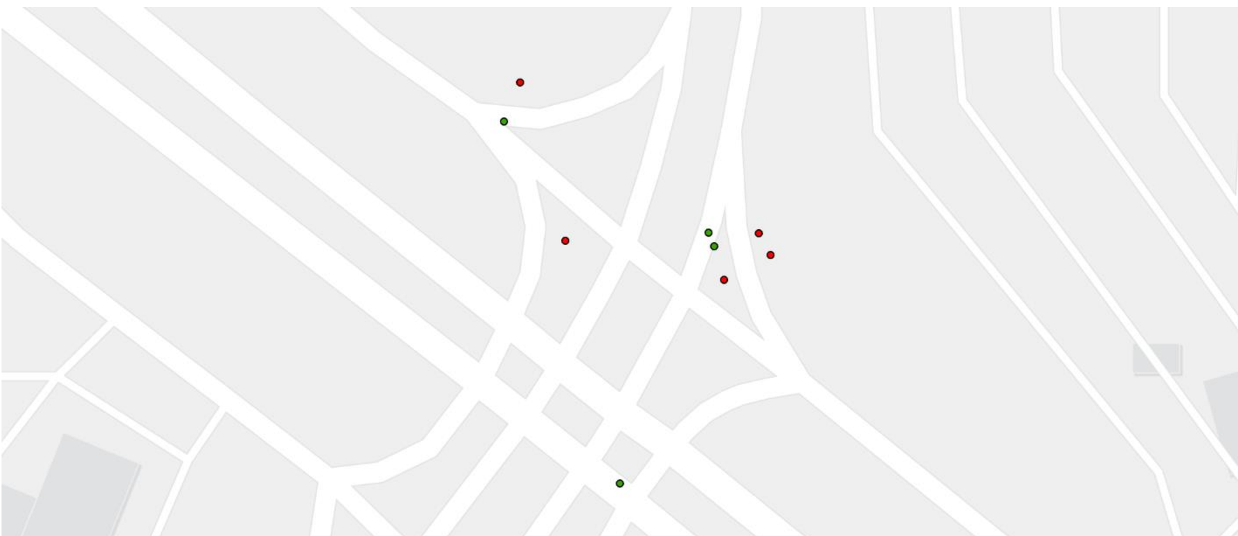
5.2.4.3 Yield Sign Datasets

The research team extracted and cleaned yield data. Because OSM data did not contain yield sign, data from the other three sources were collected for analysis. Also, no yield signs were founded in all datasets in I-30, so the research team only evaluated yield signs in Green Oaks Blvd, Lyndon B Johnson Fwy, and Buckner Blvd. The research team had the following datasets (Figure 56):

- Mobileye: 32 (in Green Oaks Blvd, Lyndon B Johnson Fwy and Buckner Blvd).
- Arlington: 14 (in Green Oaks Blvd).
- Mapillary: 17 (in Lyndon B Johnson Fwy and Buckner Blvd).
- Combined dataset after data fusion: 48 (in Green Oaks Blvd, Lyndon B Johnson Fwy and Buckner Blvd).
- Ground truth data: 29 (in Green Oaks Blvd, Lyndon B Johnson Fwy and Buckner Blvd).



a. Comparison of yield sign data on Green Oaks Blvd., Arlington, TX.



b. Comparison of yield sign data on LP 12/Buckner Blvd., Dallas, TX.

c. Comparison of yield sign data on IH-635/LBJ East, Dallas, TX.

Figure 56. Comparison of Yield Sign Data.

5.2.5 Data Fusion and Evaluation

5.2.5.1 Purpose

This section focused on the evaluation of selected TCI datasets based on their accuracy and completeness and comparing them to ground truth data. The research team finally merged several datasets to create a fused dataset, then applied a clustering algorithm known as DBSCAN to reduce redundancy and ensure the reliability of the combined data. The evaluation was then carried out on three types of TCI data: speed limit signs, stop signs, and yield signs. Two primary indicators measured the quality of these datasets:

1. Accuracy: evaluating the correctness of the datasets.
2. Completeness: examining how many ground truth data can be matched with or covered by the selected datasets.

The research team assumed that both accuracy and completeness are equally important when measuring data quality and reliability. This study offers an in-depth look at the performance of various datasets when digitizing traffic sign data, thereby providing valuable insights for improving future data collection and validation methodologies.

5.2.5.2 Combined Dataset after Fusion

The research team hoped to improve the data reliability by combing multiple datasets. In order to reduce redundant data from multiple datasets after combing, DBSCAN was applied to cluster these datasets. A combined dataset was created for each kind of traffic dataset (Figure 57). DBSCAN can be applied to cluster objects over both space and time when the time component can be considered as an additional dimension. For the combine dataset of speed limit sign, because it also had different speed limit values, the research team considered the number of speed limit as time attribute in the study and cluster speed limit data to reduce redundant and repeated data.

- Combined Speed Limit Sign Dataset: 358 (in all four corridors).
- Combined Stop Sign Dataset: 88 (in Green Oaks Blvd, Lyndon B Johnson Fwy and Buckner Blvd).
- Combined Yield Sign Dataset: 48 (in Green Oaks Blvd, Lyndon B Johnson Fwy and Buckner Blvd.

a. Comparison of combined data on I-30, Arlington, TX

b. Comparison of combined data on Green Oaks Blvd., Arlington, TX.

c. Comparison of combined data on LP 12/Buckner Blvd.

d. Comparison of combined data on IH-635/LBJ East, Dallas, TX.

Figure 57. Comparison of Combined Datasets.

5.2.5.3 Evaluation Methods

In this study, the research team evaluated the quality of datasets by reliability and considered two indicators to measure datasets: accuracy and completeness.

- Accuracy
    - Accuracy aimed to evaluate the correctness of selected datasets.

$$A = \frac{p}{P}$$

    - A = Accuracy of the selected dataset, p = the amount of data of the selected dataset matched with ground truth data, P = the amount of selected dataset.
- Completeness
    - Completeness aimed to evaluate the number of ground truth data that can be matched with selected datasets. The research team assumed ground truth data within a certain distance to traffic signs from selected datasets were matched ground truth data.

$$C = \frac{g}{G}$$

    - C = Completeness of the selected dataset, g = the amount of ground truth data matched with the selected dataset, G = the amount of ground truth data.
- Reliability
    - Consider accuracy and completeness to evaluate the quality of the dataset.

$$R = w_1 A + w_2 C$$

    - R = Reliability of the selected dataset, w_1 = the weight of accuracy, w_2 = the weight of completeness.

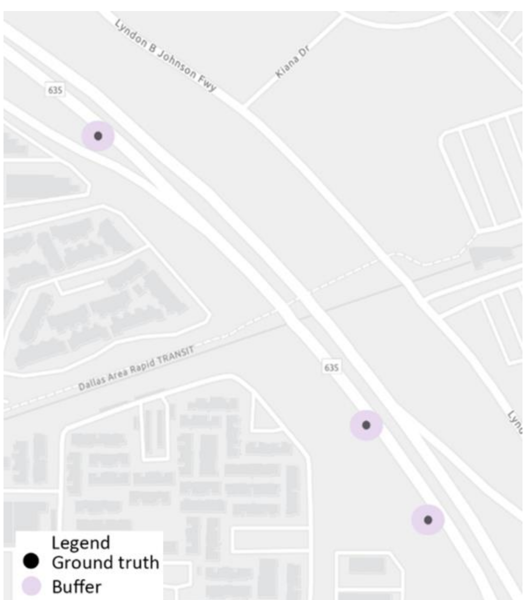
The weight can be changed depending on the purposes or situation. In this current research, the research team assumed that two indicators are equally important. This method can be applied to evaluate the stop sign and yield sign, but it was not enough for speed limit signs. Because the value of the speed limit of selected datasets needs to be matched with ground truth, the research team also evaluated the speed limit value on the sign to make sure to get the correct results.

The procedures to evaluate the reliability is listed below:

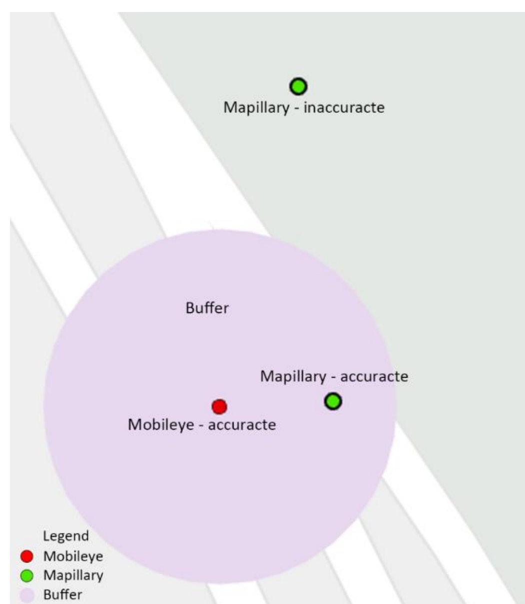
1. Make a buffer around the ground truth data (50 ft in current research).
2. Join spatially the selected datasets with the buffer.
3. Evaluate the accuracy: the matched data of selected datasets divided by the total amount of the selected dataset.
4. Evaluate the completeness: if a buffer of the ground truth data has traffic signs from selected datasets located in it, it indicated that the buffer is covered by traffic signs in the selected dataset.
5. Evaluate the reliability: average the results of accuracy and completeness.

When evaluating speed limits signs, the value of speed limit also needs to be matched in selected and ground truth datasets.

An example was shown to evaluate the speed limit sign (Figure 58). Two speed limit signs, one from Mobileye and another from Mapillary, were located in the buffer. This indicated they could be considered at the same location as the ground truth data. Then, after checking the number of speed limit, the two signs had the same value as the ground truth data, so both can be considered as correct speed limit signs. Also, the buffer of the ground truth data had two speed limit signs located within it; this ground truth speed limit sign can be considered covered by Mapillary and Mobileye datasets. In this way, the accuracy and completeness can be evaluated. The Mapillary speed limit sign at the top was out of the buffer, which indicated the distance between it and the ground truth data was too far, so it could not be considered a correct speed limit sign.



a. Example buffer for ground truth data.



b. Example of spatial joining the buffer with data.

Figure 58. Demonstration of Evaluation Methods.

5.3 EVALUATION RESULTS

The results of reliability for each data source were listed below (Table 13).

Table 13. Evaluation of Traffic Signs in Selected Datasets.

TCI Data Types	Traffic sign	Accuracy	Completeness	Average reliability
Speed limit sign	Mobileye	83.77%	99.38%	91.57%
	Arlington	94.59%	21.74%	58.17%
	Mapillary	26.92%	29.19%	28.06%
	OSM	20.00%	0.62%	10.31%
	Combined datasets	**100.00%**	**99.38%**	**99.69%**
Stop sign	Mobileye	100.00%	93.75%	96.88%
	Arlington	100.00%	11.43%	55.72%
	Mapillary	35.18%	59.38%	47.28%
	OSM	100.00%	4.32%	52.16%
	Combined datasets	**100.00%**	**96.87%**	**98.44%**
Yield sign	Mobileye	96.88%	41.38%	69.13%
	Arlington	92.86%	41.38%	67.12%
	Mapillary	64.71%	34.48%	49.59%
	Combined datasets	**100.00%**	**100.00%**	**100.00%**

The detailed observations from the results are summarized as follows:

- Speed limit signs exhibited varying degrees of reliability among the datasets. Mobileye and the combined dataset demonstrated notably higher levels of reliability at 91.57 percent and 99.69 percent, respectively, signifying superior accuracy and comprehensiveness. Conversely, Arlington and Mapillary displayed lower reliability scores of 58.17 percent and 28.06 percent, respectively, while OSM obtained the lowest score of 10.31 percent.
- When looking at stop signs, the combined dataset has the highest at 98.44 percent, followed by Mobileye at 96.88 percent. Arlington, OSM and Mapillary have relatively close reliability, 55.72 percent, 52.16 percent, and 47.28 percent, respectively.
- For yield sign, combined dataset has a perfect performance to reach 100 percent reliability. Mobileye also performs well with a reliability of 69.13 percent. Arlington comes next at 67.12 percent. Mapillary has the lowest score of 49.59 percent. After data fusion, the reliability score had been increased from the highest value of a single dataset (69.13 percent) to 100 percent, about a 30.87 percent improvement.

Based on the data analysis, it becomes evident that the combined dataset stands out as the most dependable source of information. This finding underscores the significance of employing combined datasets to ensure optimal reliability when digitizing traffic sign data. Among the four datasets considered, Mobileye demonstrated superior overall performance. In the case of Arlington, the data exhibited commendable performance for the yield sign and speed limit sign. However, individual datasets displayed significant variations in reliability. This discrepancy indicates a pressing need for enhancing data collection methods for specific providers, particularly for Mapillary and OSM, which are both open-source datasets. Unlike commercial providers, which have dedicated teams, advanced technology, and rigorous validation processes ensuring accuracy and reliability, open-source datasets rely on crowd-sourced contributions, leading to variable data quality and lacking uniformity in their datasets.

5.3.1 Future Work

The insights derived from this study provide a clear path for future work, with the ultimate objective of enhancing the quality, comprehensiveness, and reliability of TCI data. Based on the current findings, the research team's future efforts will be geared towards the following areas:

- Expanding Dataset Coverage: The team will collaborate with the City of Arlington, Mobileye, and other potential data providers to extend the coverage of TCI data beyond the currently studied corridors.
- Enhancing Open-Source Data Quality: Efforts will be made to improve the data quality from open-source platforms like Mapillary and OSM through initiatives such as community drives for data collection and additional data validation measures.
- Examining Additional TCI Elements: The current study primarily focused on certain types of TCI data. Moving forward, the team will evaluate other critical TCI elements such as traffic lights, traffic islands, and more. This will provide a more comprehensive understanding of traffic infrastructure.
- Developing a Comprehensive Evaluation Framework: The evaluation framework will be refined to be more robust and inclusive, incorporating additional performance metrics and considering factors like data freshness and update frequency.
- Expanding the Study Area: Given the local variability in TCI data quality and coverage, the research team aims to expand their study to encompass broader areas, possibly at a city-wide scale. This will provide a more representative understanding of TCI data across diverse environments.
- Investing in Proprietary Data Collection: To ensure data reliability and control over the data collection areas, the team will explore investing in proprietary data collection efforts.
- Leveraging Advanced Technologies: The use of advanced technologies like AI and machine learning will be explored for automating the TCI data collection and evaluation process.

- Fostering Collaborations: The research team will seek partnerships with other stakeholders such as academic institutions, government bodies, and private companies, pooling resources and sharing expertise to enhance TCI data quality and coverage.

5.3.2 Reference

[1] A. H. Ashara, S. Saleh, U. Hassan, and M. J. Kaura, “Traffic Management System in Abuja City Center, using Geographic Information Systems (GIS) and Global Positioning System (GPS). A case study of FCT, Nigeria,” *J. Phys.: Conf. Ser.*, vol. 1529, no. 5, p. 052068, May 2020, doi: 10.1088/1742-6596/1529/5/052068.

[2] Y. Wen *et al.*, “Mapping dynamic road emissions for a megacity by using open-access traffic congestion index data,” *Applied Energy*, vol. 260, p. 114357, Feb. 2020, doi: 10.1016/j.apenergy.2019.114357.

[3] “Learn About The Mobileye Mission and Vision | Mobileye AU.” https://ims.mobileye.com/fleets/au/about/ (accessed Jun. 30, 2023).

[4] “Arlington Arlington » PlanUrStay,” May 29, 2023. https://planurstay.com/st_location/united-states-of-america/arlington/ (accessed Jun. 30, 2023).

[5] S. O’Hear, “Mapillary, the crowdsourced database of street-level imagery, has been acquired by Facebook,” *TechCrunch*, Jun. 18, 2020. https://techcrunch.com/2020/06/18/mapillary-facebook/ (accessed Jun. 30, 2023).

[6] “OpenStreetMap.” https://www.openstreetmap.org/#map=5/38.007/-95.844 (accessed Jun. 30, 2023).

[7] J. Sander, M. Ester, H.-P. Kriegel, and X. Xu, “Density-Based Clustering in Spatial Databases: The Algorithm GDBSCAN and Its Applications,” *Data Mining and Knowledge Discovery*, vol. 2, no. 2, pp. 169–194, Jun. 1998, doi: 10.1023/A:1009745219419.

5.4 WEB GIS-BASED TCI DIGITIZATION PLATFORM

5.4.1 Summary

The research team developed a web GIS-based application based on the developed TCI data digitization framework. The platform was developed by ArcGIS App Builder and then was held on ArcGIS Online (an Esri-developed cloud-based GIS platform) for web-based access (as shown in Figure 58).

ArcGIS Online enables users to create, store, share, and manage maps, data, and geospatial information and services. It also offers a wide range of tools and functionalities for data analysis, mapping, and visualization. It allows users to generate customized maps, conduct spatial analysis, and collaborate by sharing data and maps. ArcGIS Online provides access to a vast repository of geospatial data, pre-built maps, and applications. It caters to the needs of individuals, organizations, and government agencies, supporting various applications, including the digitized inventory of TCIs in this study.

ArcGIS App Builder is a powerful tool developed by Esri that enables users to create customized web and mobile applications without the need for extensive coding. This intuitive app-building platform offers a user-friendly interface and a range of configurable templates, widgets, and tools to develop highly functional and tailored applications.

One of the customized widgets developed using ArcGIS App Builder is specifically designed for TCI data fusion and evaluation. This widget enhances the capabilities of the application by providing advanced functionality for data integration and analysis. The data fusion aspect of the widget allows for the combination of multiple datasets from various sources into a unified and comprehensive TCI dataset. By leveraging this widget, users can merge data from different providers, such as government agencies, commercial sources, and crowd-sourced platforms. This data fusion process results in an enriched dataset that encompasses a wider range of TCI information.

Another customized widget facilitates data evaluation by implementing advanced analytical algorithms and methodologies. It enables users to assess the accuracy, completeness, and consistency of the TCI data. This evaluation process involves comparing the imported TCI data with ground truth references or authoritative sources to identify any discrepancies or inconsistencies. Users can visualize and analyze the differences between datasets, enabling them to identify areas of improvement and take corrective actions.

The customized widgets integrate seamlessly into the ArcGIS App Builder application, allowing users to access its functionality within the app’s interface. Its inclusion enhances the application’s capability to effectively manage and evaluate TCI data, providing users with valuable insights and enabling informed decision-making in traffic infrastructure management. The widget’s configuration options allow users to tailor its behavior to specific project requirements. Users can customize the data fusion and evaluation parameters, adjust visualization settings, and define thresholds for data quality assessment. This flexibility ensures that the widget aligns with the specific needs of different users and organizations.

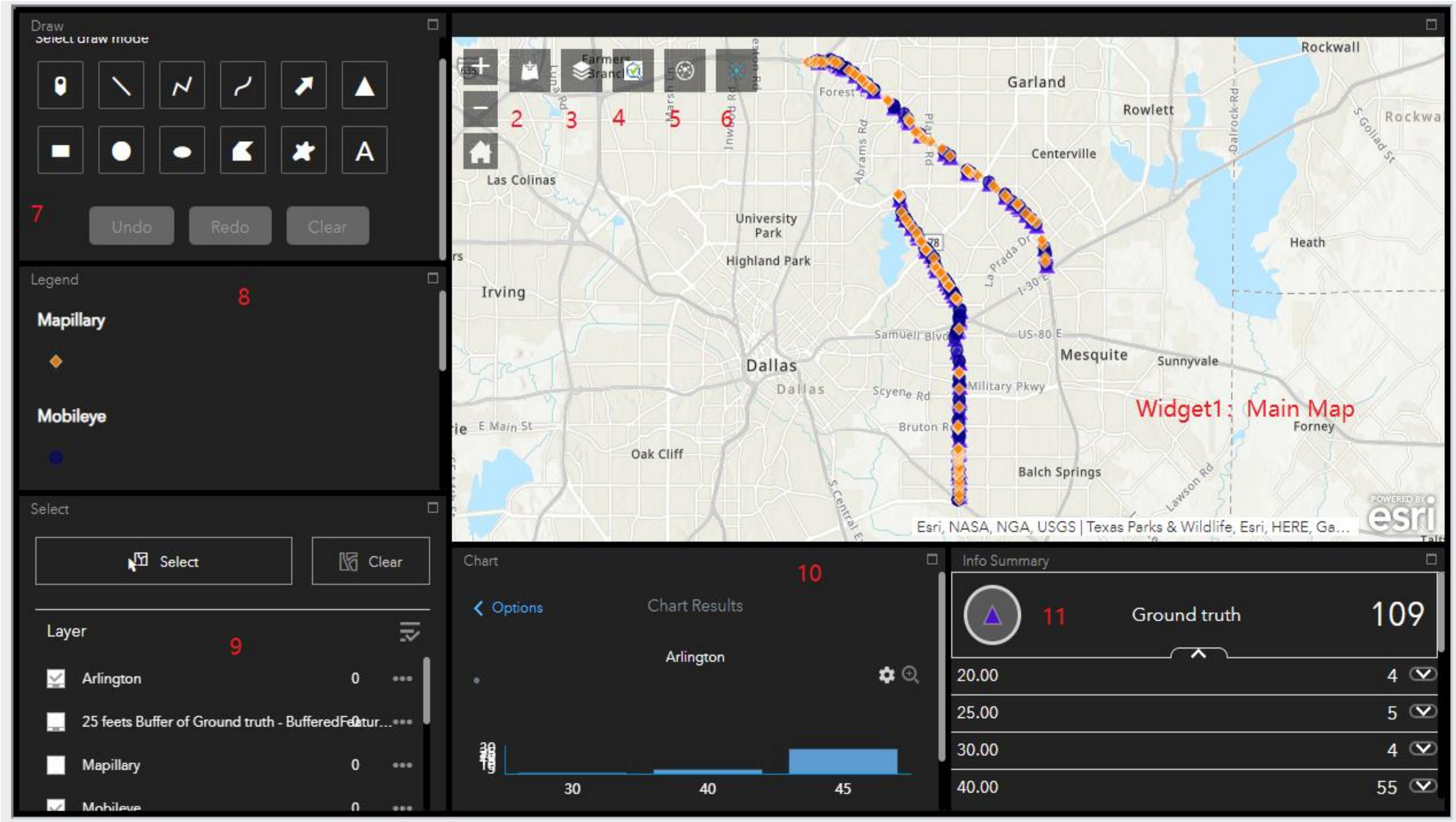
5.4.2 Design of the Application

This web application consists of several key components (as shown in Figure 59):

- Main Map Panel (Widget 1): This central component displays the map from ArcGIS Online, providing users with the flexibility to customize the map’s layout and elements. Users can interact with the map through various functionalities such as data visualization, map roaming, zooming, layer selection, and the ability to use custom widgets.

- In-Panel Widgets (Widgets 2–6): These widgets are located within the main map panel and enable users to engage in interactive activities with the map. They enhance the user experience by offering features like drawing tools, measurement tools, spatial analysis, and other map-related interactions.
- Off-Panel Widgets (Widgets 7–11): Positioned around the map, these widgets display specific information about the data on the map to the user. They provide valuable insights and may include data filters based on map element attributes, allowing users to refine and focus their analysis.

Overall, this web application offers a dynamic and user-friendly platform for exploring and visualizing geographical data, empowering users with comprehensive tools for efficient data analysis and decision-making. The architecture of the application is shown in Figure 60.



(Powered by Esri ArcGIS Online Cloud Platform and Web App Builder)

Figure 59. Developed Web GIS-based TCI Digitization Application.

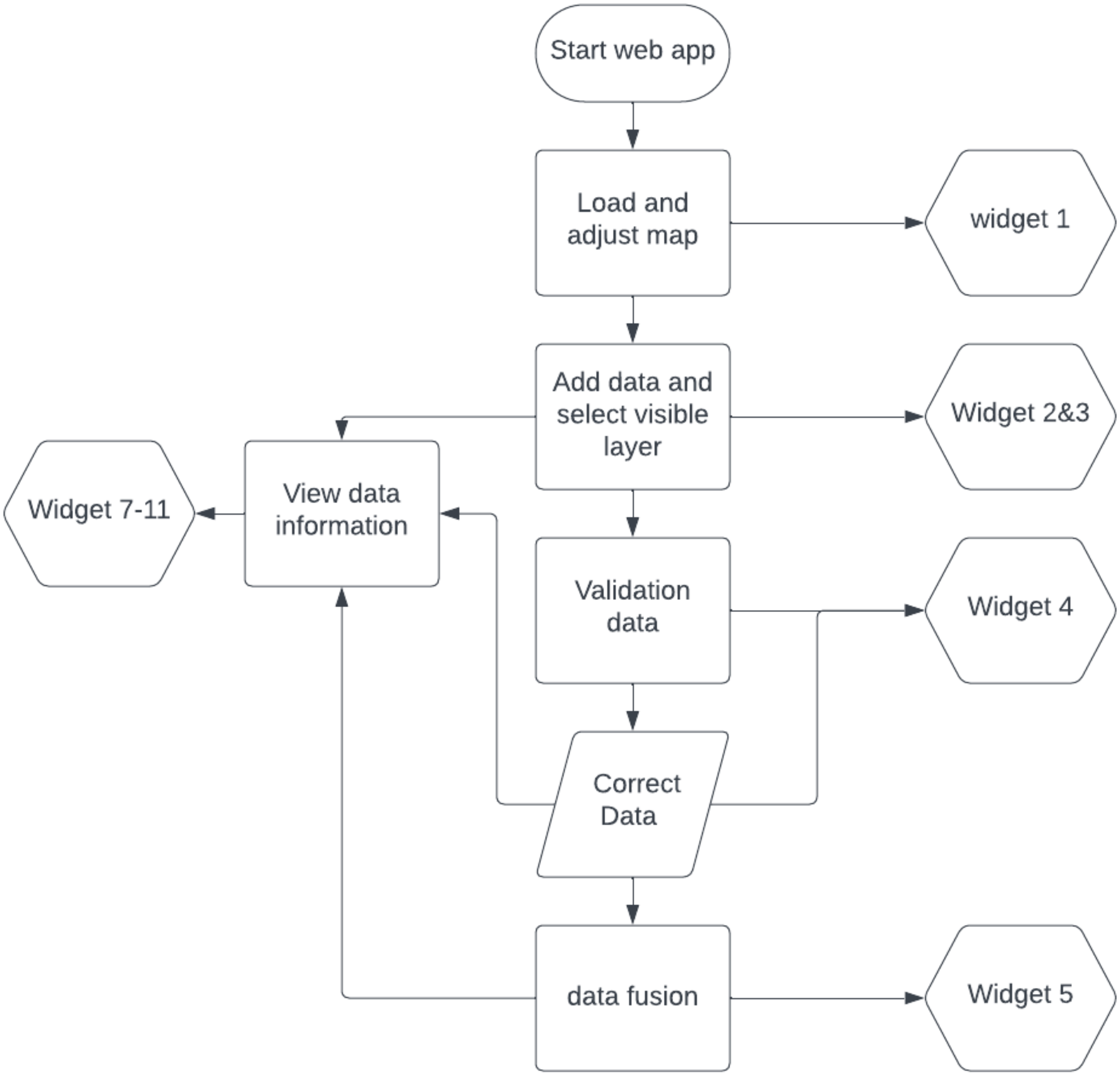


Figure 60. Architecture of the Application.

5.4.3 Main Widgets

- **Main map widget**: When the user uses this widget, the user can view, zoom, or move the map like other map applications (Figure 61).

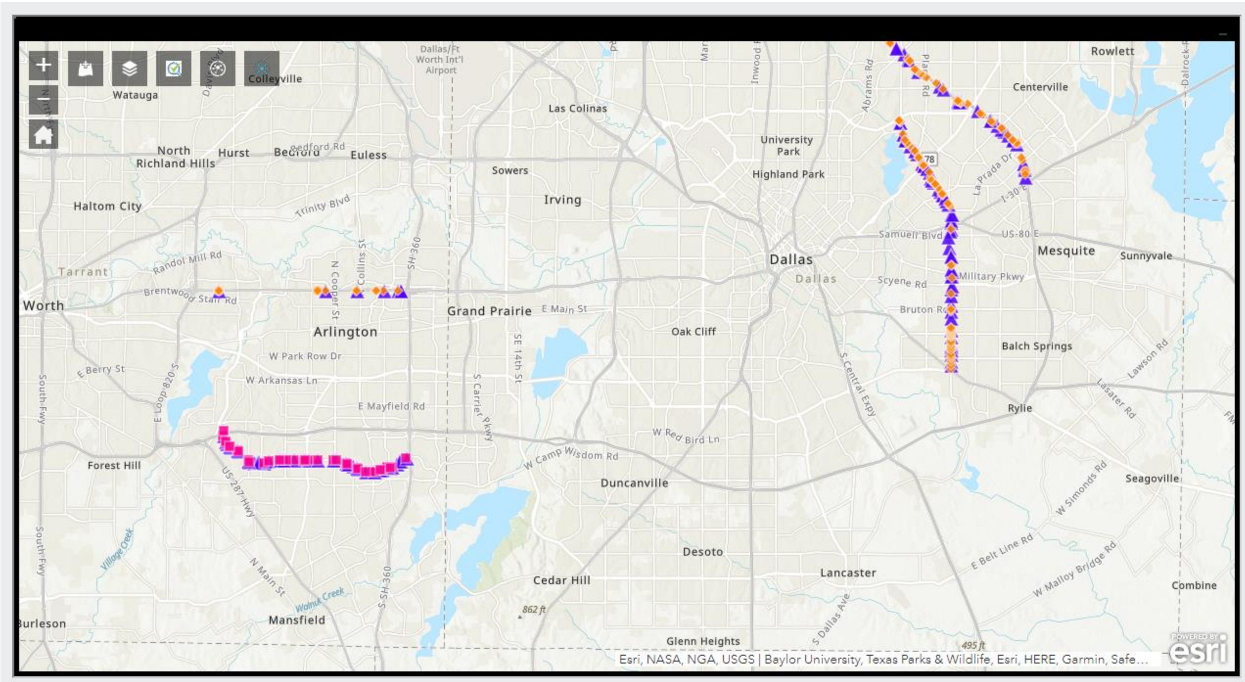


Figure 61. Interface of "Main Map" Widget.

- **Add Data widget**: This widget allows users to add data to web applications. The source of data can be ArcGIS online services, external links, and local data. The supporting data formats include Shapefile, CSV, KML, and other diverse geographic data formats (Figure 62).

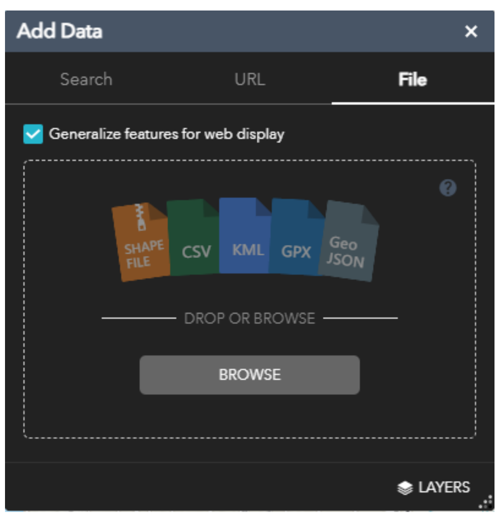


Figure 62. Interface of "Add Data" Widget.

- **Layer List widget**: The Layer List widget provides a list of operational layers and their symbols, and allows users to turn individual layers on and off. Each layer in the list has a checkbox that allows the user to control its visibility. Some layers contain sublayers or subtypes (Figure 63).

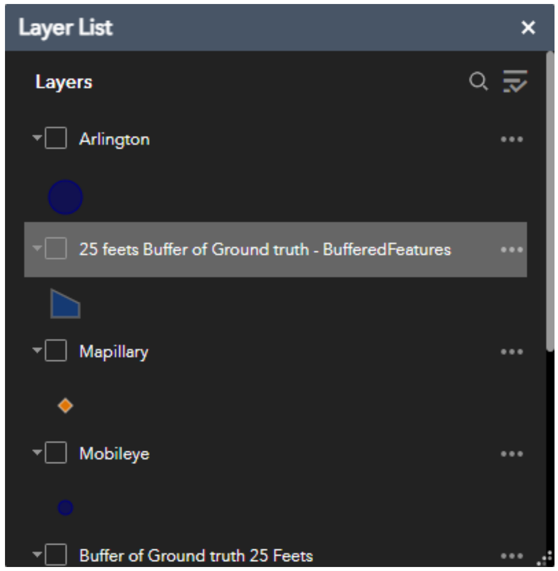


Figure 63. Interface of "Layer List" Widget.

- **Create Buffers widget**: This widget is part of the build-in-weight analysis, and it provides functionality consistent with what the project needs. Through this widget, the user can generate a buffer layer for the ground truth point layer for subsequent verification. The user can choose the point layer to generate the buffer and the size of the buffer. At the same time, the type of buffer is overlap or dissolve. The generated buffer will be saved in the ArcGIS Online server and automatically added to the layer (Figure 64).

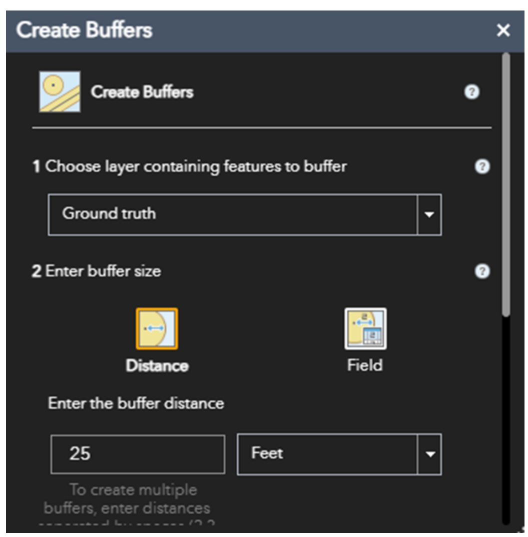


Figure 64. Interface of "Create Buffer" Widget.

- **Validation widget**: This is a self-developed custom widget, and its main function is to implement the data verification method mentioned in the above report in the web application. In order to complete this step, the user first needs to select the layer to be verified (target layer) and the buffer layer of the ground truth point generated in the create buffer section (ground truth layer). The widget will use the spatial relationship of the two layers to verify which points are spatially correct in the target layer.

- In order to verify whether the target point and the ground truth point are numerically correct, the user needs to select a field for verification. For example, select the spd_lt (speed limit) field of two layers at the same time for comparison verification (Figure 65).

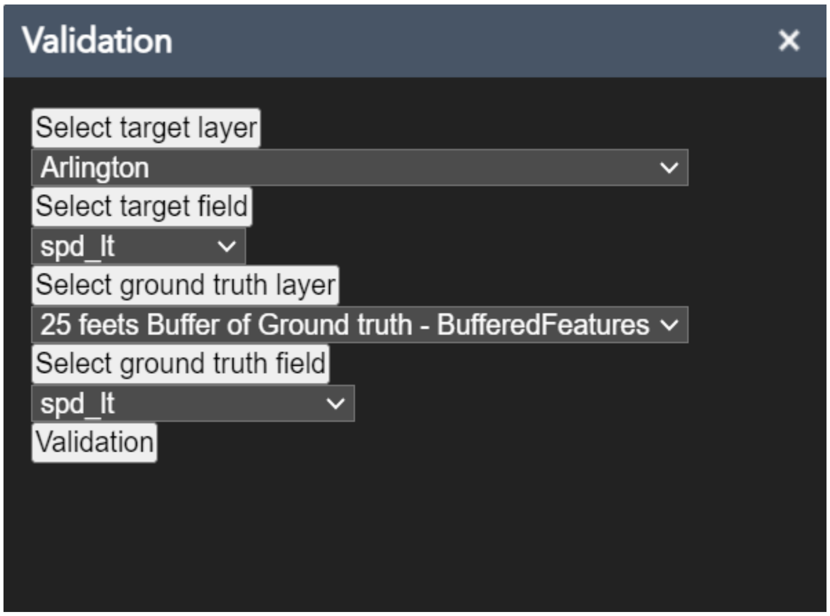


Figure 65. Select Layer and Field of Layer.

- When layer and field selections are complete, the user can press the validation button to run validation calculations, and the result will be displayed directly. In addition, the user can also click the add to layer button to generate a new layer with the correct point and add it to the map. By default, the new layer is named as the original layer name + correct, and the style is a black circle (Figure 66).

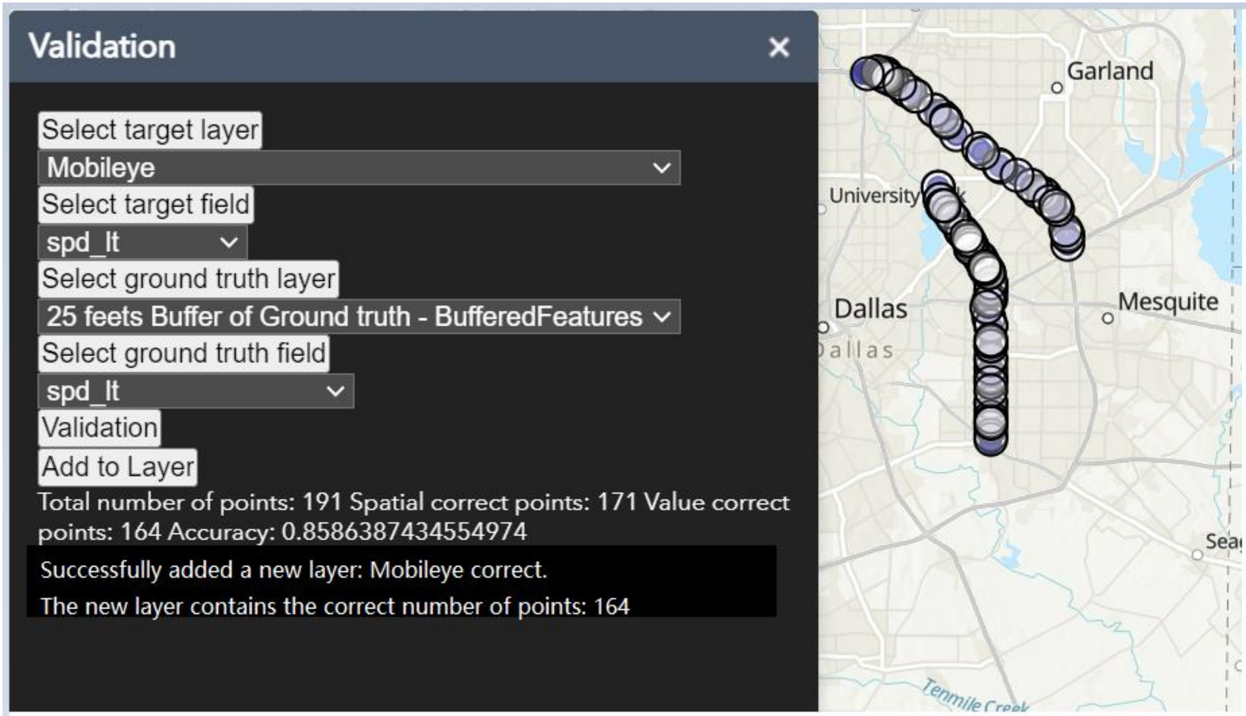


Figure 66. "Validation" Function and "Add to Layer" Function.

- Finally, the user can get the number of correct points of the target layer and the layer of correct points through this widget (Figure 67).

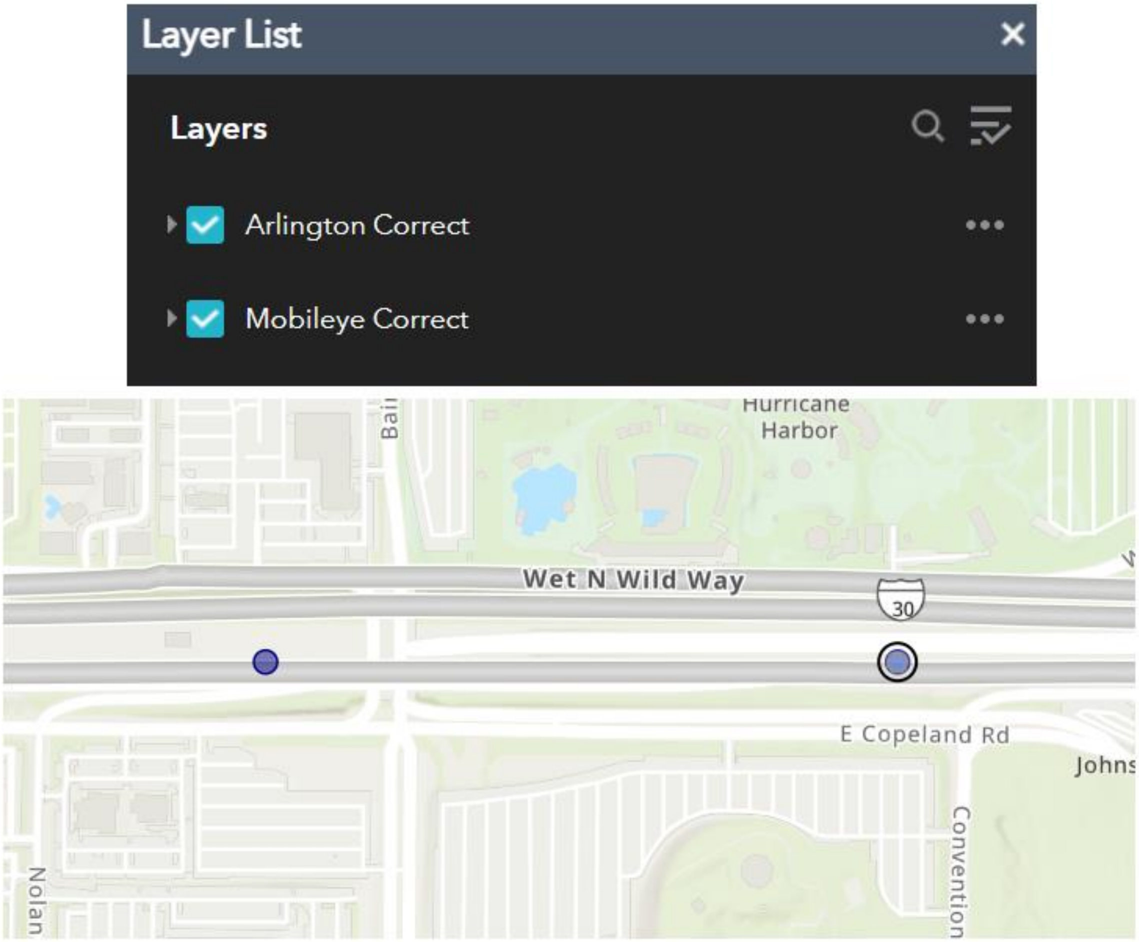


Figure 67. Results of Validation Widget.

- **Data fusion widget**: This is a self-developed customized widget. Its main function is to integrate the verified and correct data from different data sources into one layer. Users can select multiple layers and press the data fusion button to realize the function of merging multiple layers into one layer and adding it to the map (Figure 68).

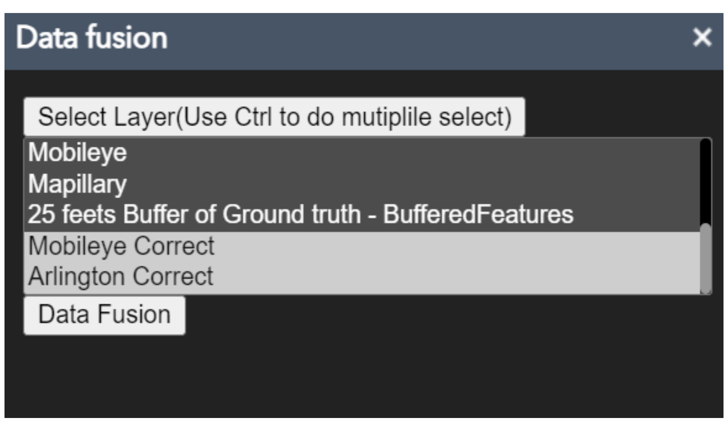


Figure 68. Interface of "Data Fusion" Widget.

- **Draw widget**: The Draw widget allows the user to draw simple graphics and text on the map (Figure 69). The user can also use it to add line distance or polygon area to the feature as text (Figure 70).

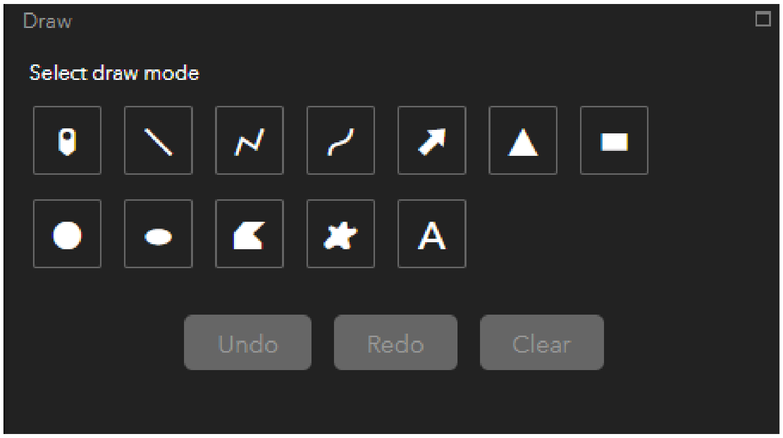


Figure 69. Interface of "Draw" Widget.

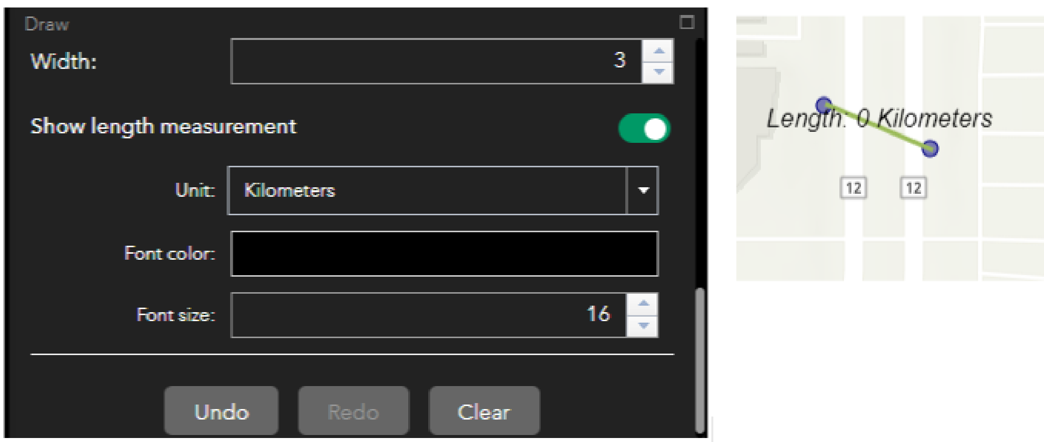


Figure 70. Example of Drawing a Line by Using This Widget.

- **Legend widget**: The Legend widget displays labels and symbols for layers in the map. It supports dynamic, tiled, image, feature, and KML layer types as well as WMS layers with an associated legend URL. The Legend widget can be set to automatically update when the visibility of a layer or sublayer changes. When no operational layers are rendered in the map, the Legend widget is blank (Figure 71).

Figure 71. Interface of "Legend" Widget.

- **Select Widget**: The Select widget enables users to interactively select features on the map and take actions on the selected features (Figure 72).

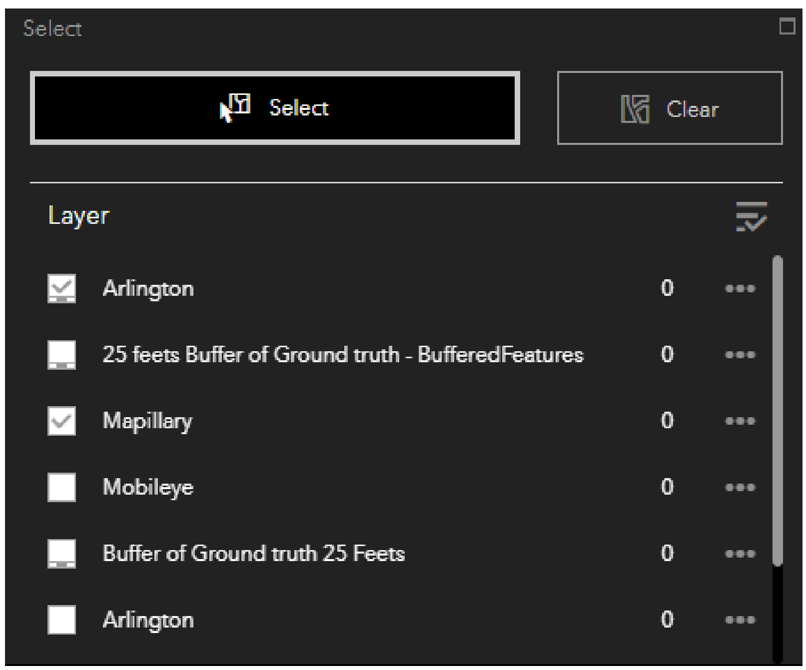


Figure 72. Interface of "Select" Widget.

- **Chart widget**: This widget is part of the build-in analysis, and it provides functionality consistent with what the project needs. Through this widget, the user can generate a buffer layer for the ground truth point layer for subsequent verification (Figure 73).

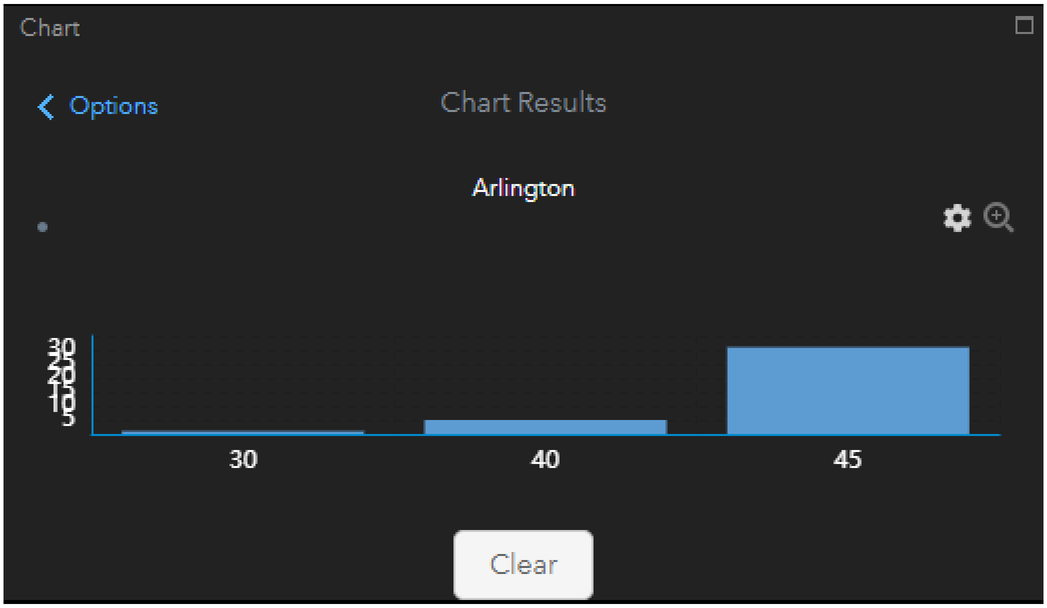


Figure 73. Interface of Chart Widget.

- **Info Summary widget**: The Info Summary widget allows the user to provide a count of features in the current map extent for each layer specified. Each layer in the widget panel can be expanded to show a list of features in the current extent, optionally grouped by a specified field. Point layers in the widget can be configured to display as clusters (Figure 74).

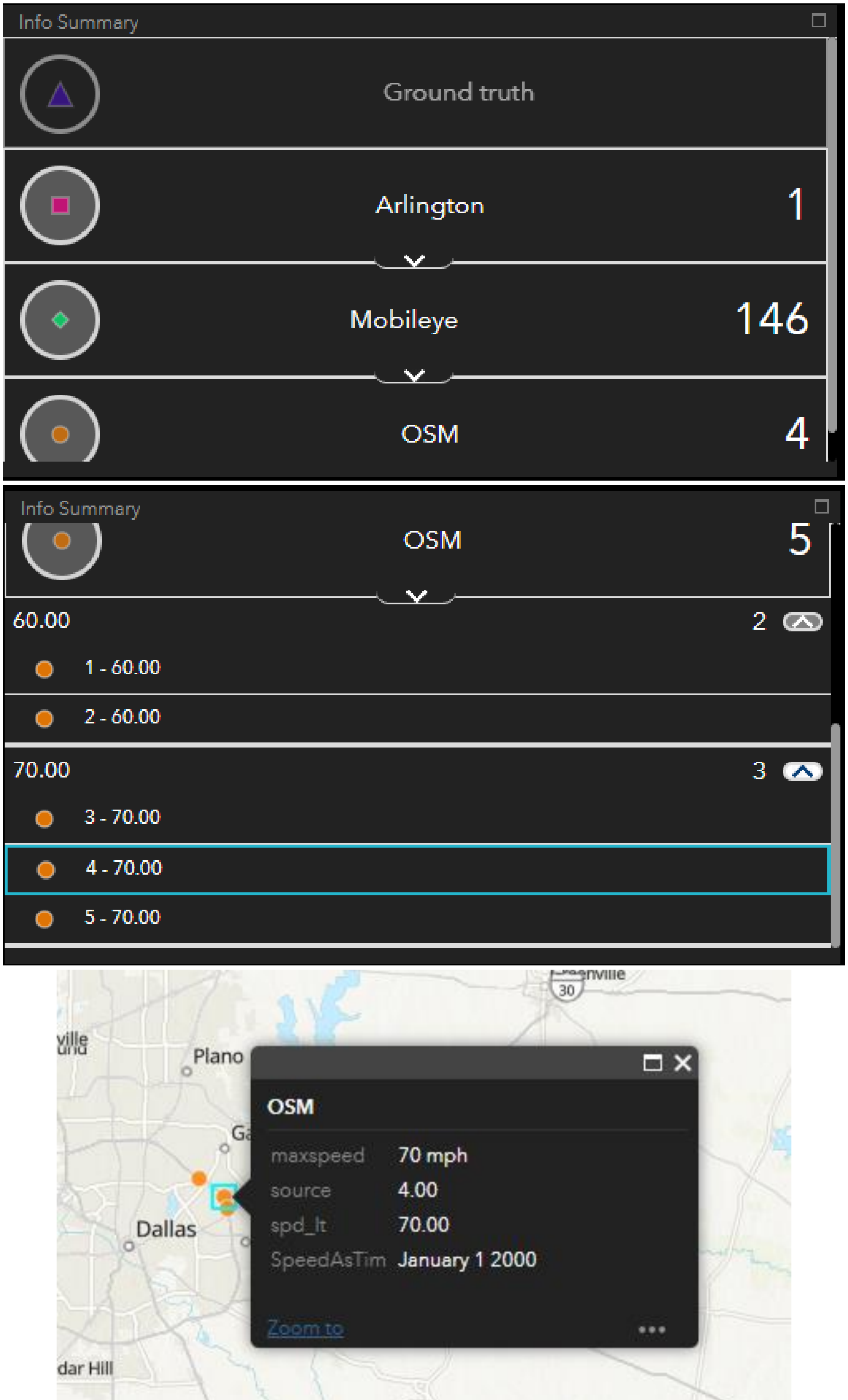


Figure 74. Interface of "Info Summary" Widget.

5.5 AV SIMULATIONS

5.5.1 Purpose

Evaluating the impact of the TCI digitization platform in a real-world setting with real-world AVs poses several challenges and risks. To mitigate these challenges and manage the associated risks, simulation plays an important role as an initial step.

- **Safety**: To mitigate risks to road users, it is essential to ensure the effectiveness of TCI digitized data in a simulated model.
- **Cost**: Real-world testing is expensive and time consuming. It involves deploying AVs equipped with digitized data processing capabilities, monitoring their performance, collecting data, and resolving any technical issues. Simulation allows for more cost-effective and efficient testing, reducing the need for physical prototypes and enabling quick iterations.
- **Limited Diversity of Test Environments**: Real-world testing is constrained by the availability of suitable test environments. Simulations can recreate a wide range of scenarios, including rare and dangerous situations, to evaluate the performance of digitized data–enabled AVs (DD-AVs). This allows for testing in conditions that might not be readily available or safe to reproduce in reality.
- **Regulatory Compliance**: Before DD-AVs can be deployed on public roads, they need to meet specific regulatory standards and safety requirements. Simulations help validate the benefits of digitized data utilization, demonstrating compliance and providing regulators with evidence regarding the technology's safety and performance.

Using MATISSE, the research team was able to simulate various real-world scenarios and evaluate the benefits of incorporating TCI digitized data in enhancing the safety measures of AVs. Through this validation process, the team gained valuable insights into the potential advantages and effectiveness of TCI digitized data utilization, further reinforcing its significance in promoting safe AV operations.

5.5.2 MATISSE Simulation System Enhancements

To meet the requirements of this project, several enhancements were made to the MATISSE simulator (see Figure 75), including the following.

1. **Simulation Performance**. In MATISSE, road networks are imported from OSM and converted into MATISSE graphs. However, the conversion process includes processing all OSM data, including irrelevant information, which affects the simulation performance for large road networks. To enhance simulation efficiency, it was essential to revise the MATISSE graph data structure and improve the OSM conversion algorithm. These

enhancements aimed to eliminate unnecessary details and retain only the essential nodes needed to generate more streamlined and simplified large road network graphs.

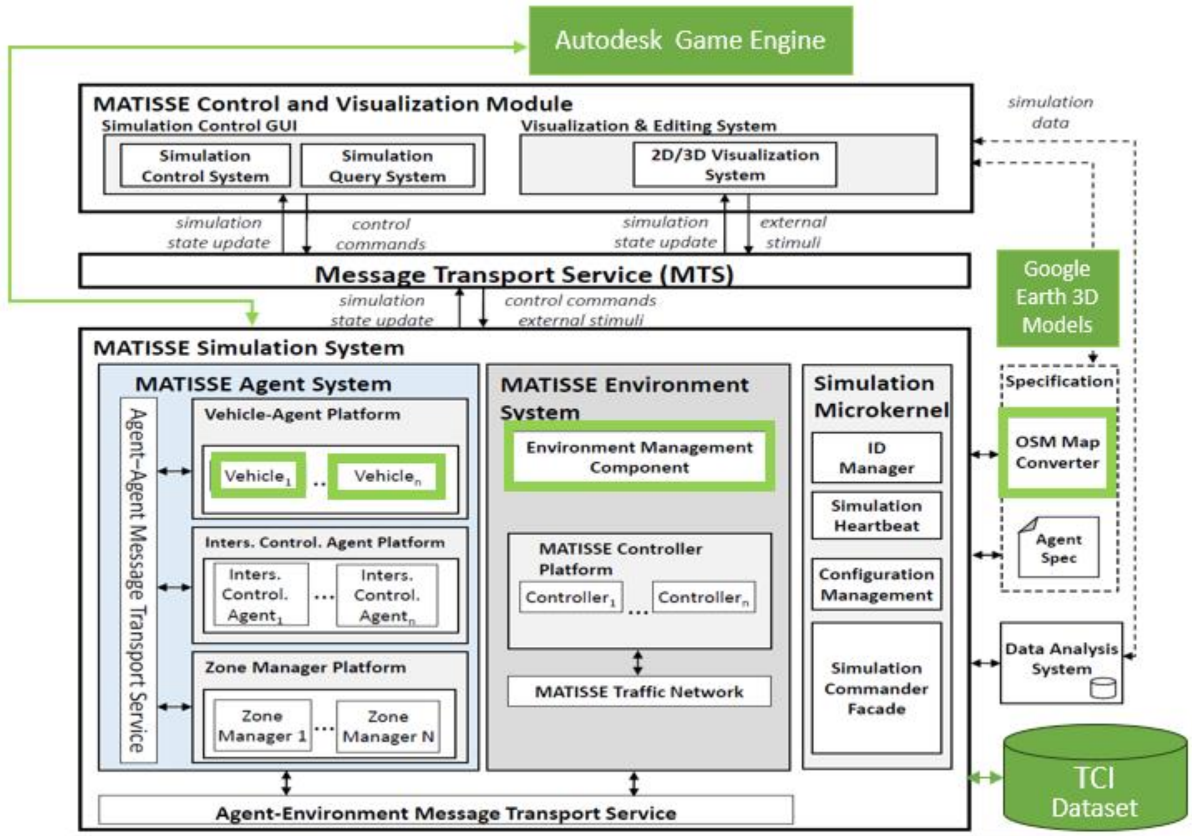


Figure 75. Component Modifications in the MATISSE Simulator (highlighted in green).

2. **Realistic Virtual Autonomous Vehicle Sensors**. In the original MATISSE simulator, the virtual AV models featured virtual sensors with a 360-degree detection range (Figure 76a). However, to create more realistic virtual AV models, it was necessary to develop virtual sensing devices that aligned with real-world specifications. As part of this effort, the research team developed virtual LiDARs, radars, and cameras to accurately mimic the functionality and capabilities of their real-world counterparts (Figure 76b). Additionally, the team had to undertake the development, implementation, and testing of sensor combination algorithms. These algorithms allow for the fusion of data from different virtual sensors, simulating the sensor fusion process observed in actual AVs. By incorporating these enhancements, the team ensured that the virtual AV models in MATISSE closely reflected the behavior and capabilities of real AVs, providing a more accurate and reliable platform for evaluating the integration of TCI digitized data for AV operation safety.

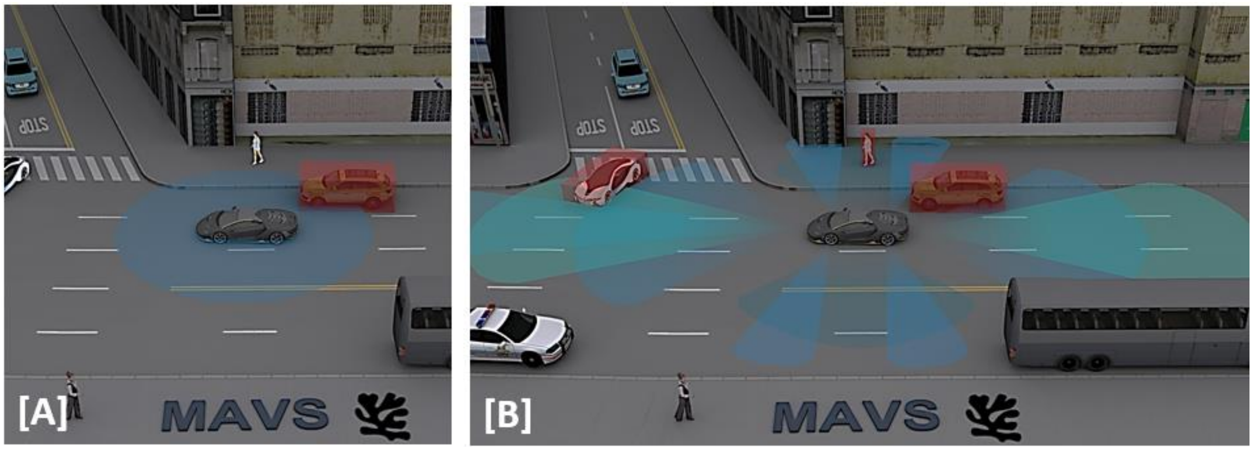


Figure 76. Virtual AV Sensing Capabilities in the Original (A) and Updated (B) MATISSE Models.

3. **Realistic 3D Models**. In the original MATISSE simulator, the environment models consisted of basic 3D buildings extracted from OSM and simple textured road models (Figure 77a). However, these generic models did not meet the project's requirement for realistic environment visualizations. To address this limitation, the research team developed an approach to systematically extract 3D models from Google Earth and seamlessly integrate them into MATISSE. This allows MATISSE to provide a more immersive and accurate representation of the environment (see Figure 77b).

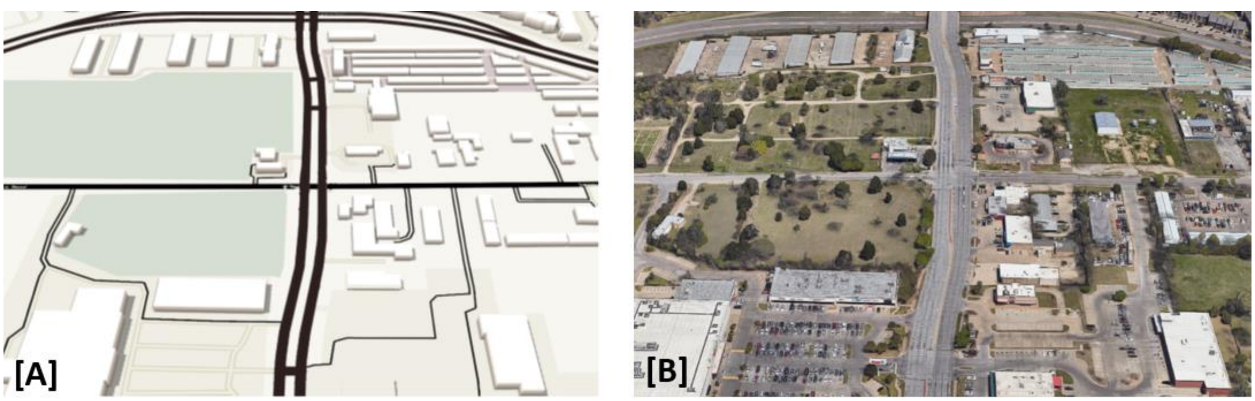


Figure 77. Original MATISSE Environment Models (A) and 3D Models Extracted from Google Earth (B).

4. **Visualization of Realistic 3D Models Extracted from Google Earth**. The original MATISSE visualizer did not have the capability to handle the visualization of 3D models extracted from Google Earth. To address this limitation, the research team explored various commercial visualizers and ultimately opted for Autodesk Game Engine (Figure 78). However, while the original MATISSE visualizer achieved real-time visualization, the current approach necessitates offline visualization. Under the current setup, after running the simulation, the simulation data are exported to Autodesk Game Engine for visualization purposes. Fully integrating Autodesk Game Engine with

MATISSE is indeed feasible, but it requires additional effort and time that fall beyond the scope of this project. Despite the shift to offline visualization, the advantage of the current approach lies in its ability to deliver enhanced visual fidelity.

Figure 78. Visualization of MATISSE Simulation Using Autodesk Game Engine.

5.5.3 TCI Dataset Utilization

In addition to the abovementioned enhancements, a critical aspect of evaluating the impact of TCI digitized data on AV operation safety involved developing a pipeline to read and process the data stored in the TCI Dataset repository. The integration of the TCI Dataset required the execution of several tasks:

1. **Data Pre-processing and TCI Model Generation**: The TCI Dataset repository contains multiple files containing information about individual types of traffic infrastructure components, such as traffic signs data and road boundaries data. Attempting to process these disparate datasets in real-time would be impractical. Therefore, it was necessary to develop an approach to systematically link these various components based on their geographical characteristics. This method enables the creation of a cohesive representation of the infrastructure and faster processing of the digitized data.
2. **TCI and MATISSE Model Alignment**: The TCI and MATISSE representations inherently differ from each other. To establish a meaningful correspondence between the two representations, algorithms were defined. These algorithms were designed with the purpose of facilitating effective mapping and synchronization of the TCI data with the corresponding elements within the MATISSE simulator.
3. **MATISSE Simulation with TCI Data**: Finally, new simulation features were implemented to enable virtual AVs to utilize the TCI data when necessary.

5.5.4 Workflow

To execute the simulations in MATISSE and visualize them using Autodesk Game Engine, a user follows several steps:

1. **Starting the Simulation Service**: Launch the simulation service and access the Control Panel to manage the Simulation Graphic User Interface (GUI) (Figure 79).

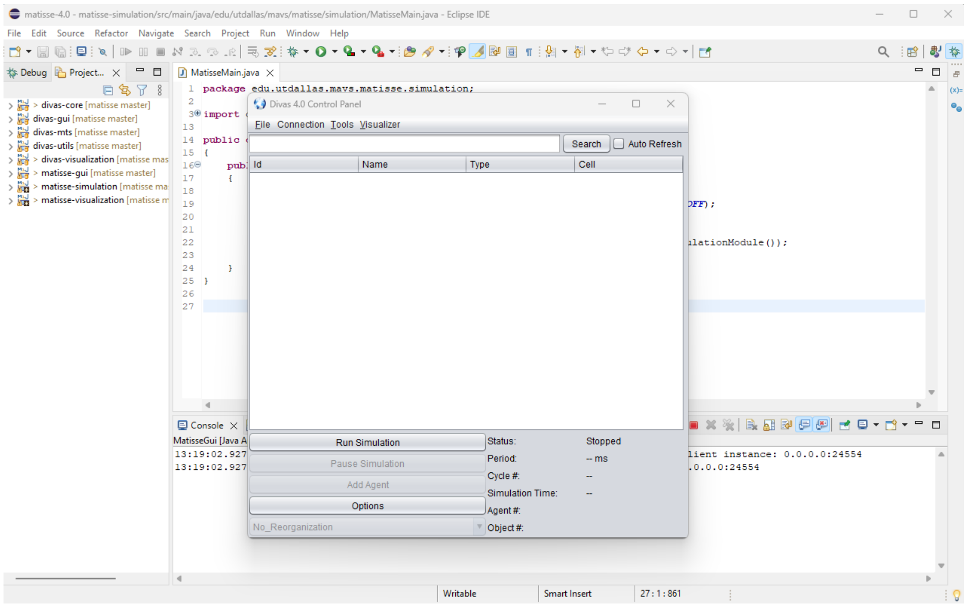


Figure 79. Control Panel of Autodesk Game Engine.

2. **Selecting the Traffic Environment**: Users have the flexibility to choose a road network from a list of user-defined traffic environment files or converted OSM files (Figure 80).

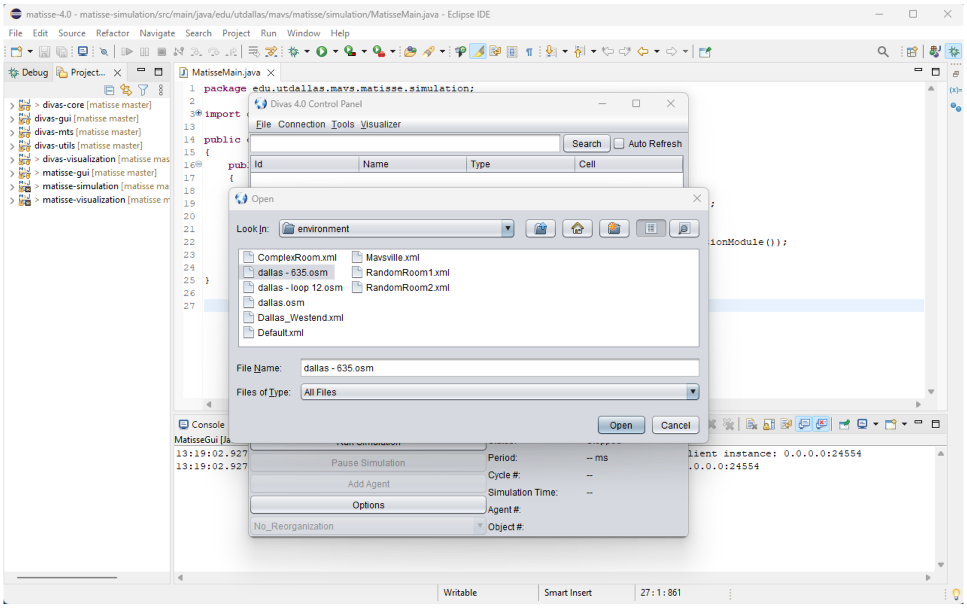


Figure 80. Traffic Environment Selection.

3. **Visualizing the Traffic Environment**: MATISSE offers an array of 3D models for buildings, roads, trees, and other elements, allowing users to create customized traffic environments (Figure 81). However, when utilizing OSM files, the visualization of real-world traffic environments is limited to simple road networks due to the unavailability of detailed 3D models.

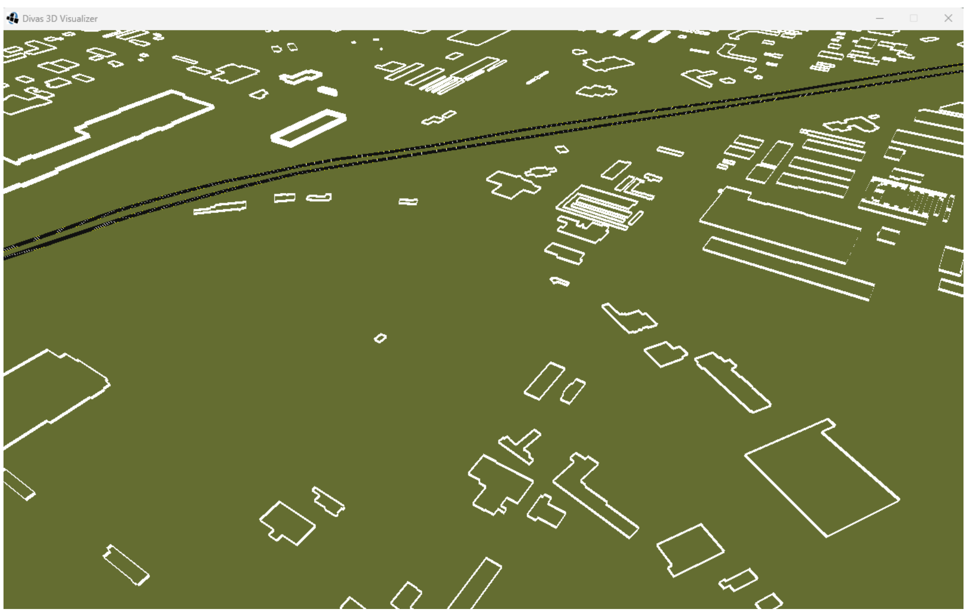


Figure 81. Model of Traffic Environment.

4. **Defining the Number of Vehicles and Parameters**: Specify the desired number of vehicles for the simulation. Users can also add vehicles and adjust their configurations

(e.g., sensor range, speed) at run time, enhancing the dynamic nature of the simulation (Figure 82).

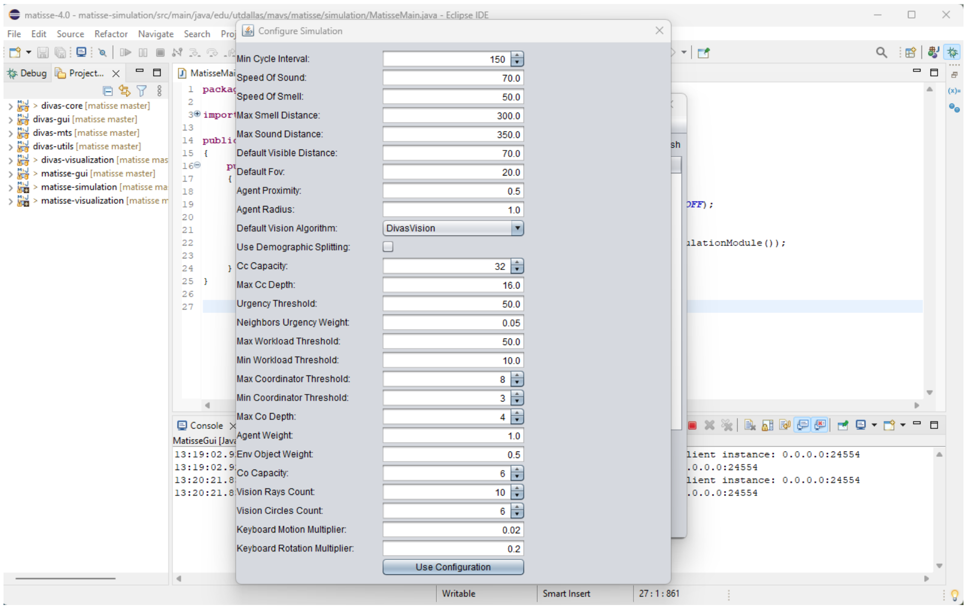


Figure 82. “Configure Simulation” Window.

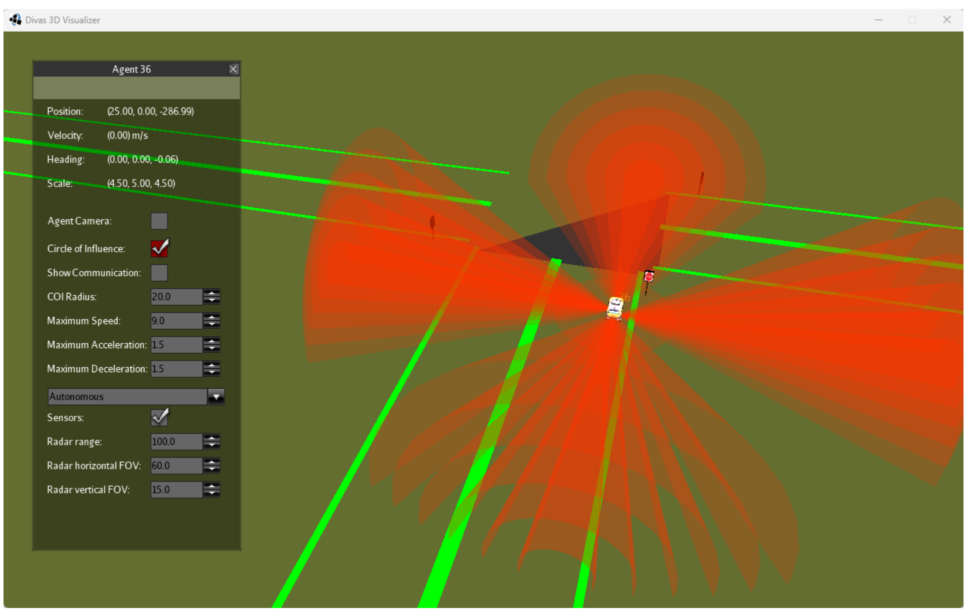


Figure 83. Configurations of Simulated Autonomous Vehicles.

5. **Running the Simulation**: Execute the simulation, and the results are saved in a JSON file.
6. **Visualizing the Simulation in Autodesk**: To achieve realistic visualizations of the simulation, import and execute the JSON file in Autodesk Game Engine.

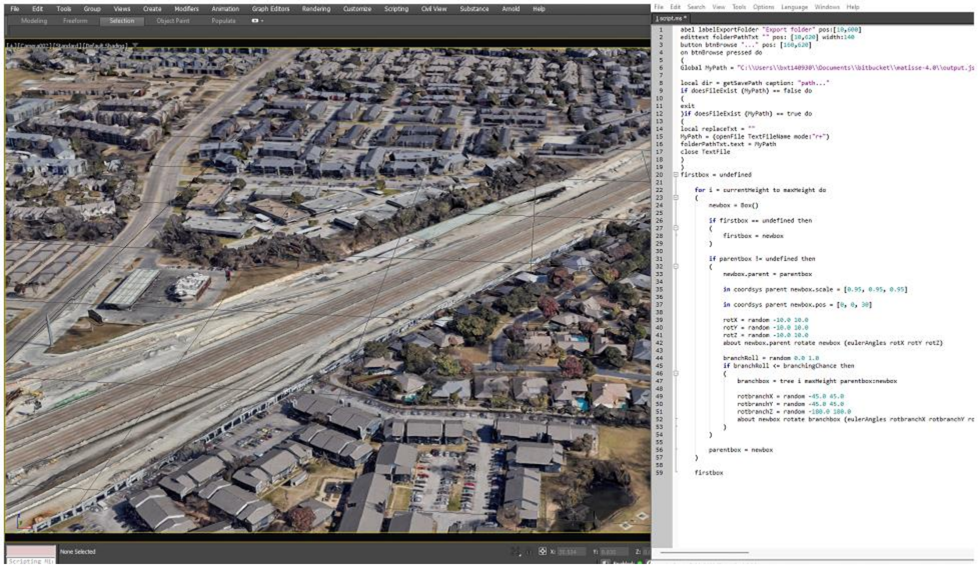
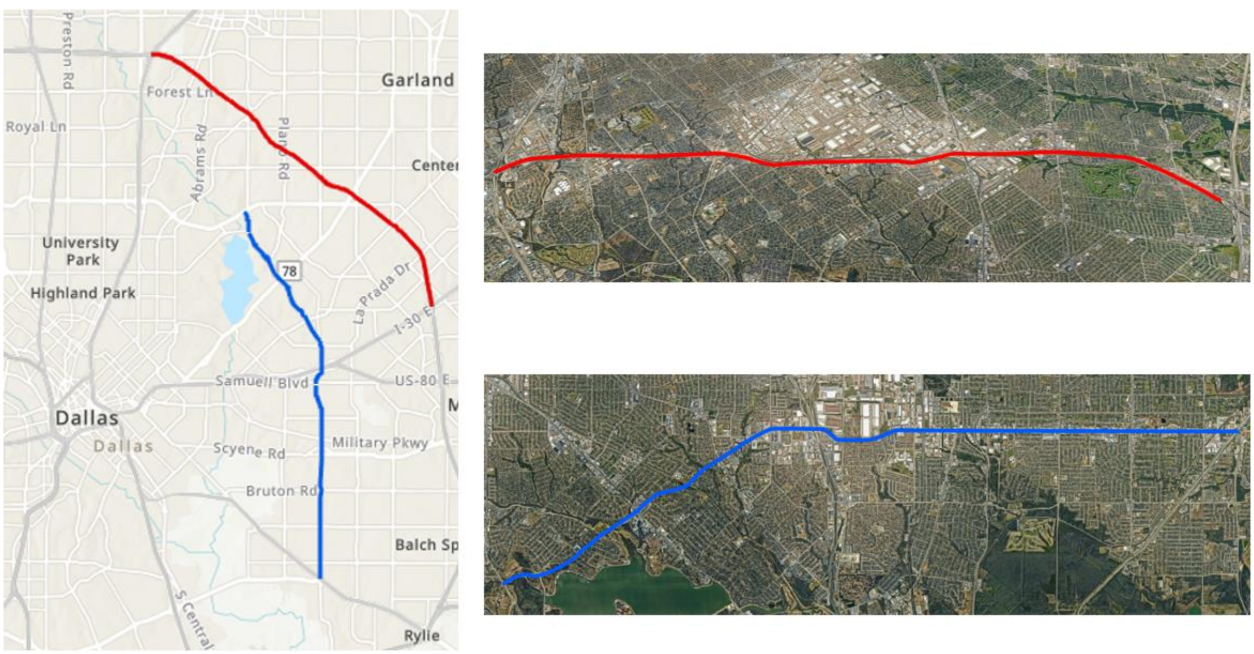


Figure 84. Realistic Visualization of Simulation Environment.

5.5.5 Simulation Setting

The road sections of interest for this project include IH-635 and Loop 12 (see Figure 85).



(a) **(b)**

Figure 85. (a) IH-635 (in red) Loop 12 (in blue); (b) Google Maps Visualizations of IH-635 and Loop 12.

To display the advantages of utilizing TCI digitized data, the research team focused on two simulation settings:

- **Simulation Setting 1**: The intersection at Loop 12/Buckner Blvd and I-30 Frontage Rd (Figure 86). This setting represents an intersection controlled by stop signs. For proper operation, an AV must possess knowledge of the road network configuration and the presence of stop signs.

Figure 86. Setting 1: Intersection at Loop 12/Buckner Blvd and I-30 Frontage Rd.

- **Simulation Setting 2**: A section of IH-635 presently under construction (Figure 87). To ensure proper operation, an AV must be aware of the road network configuration, work zone signs, work zone barriers and barrels, and the speed limit.

Figure 87. Setting 2: Intersection at Loop 12/Buckner Blvd and I-30 Frontage Rd.

For each of these settings, three scenarios were considered:

- **Scenario 1**: An AV’s sensor devices are operational, allowing it to perceive its surroundings. Furthermore, it can process its high-definition map, utilizing the detailed road network information it provides.

- **Scenario 2**: An AV's sensor devices are non-operational, resulting in the inability to sense the environment. Additionally, the AV is unable to process its high-definition map, losing access to the detailed road information it offers.
- **Scenario 3**: The AV's sensor devices are non-operational, thus lacking the ability to perceive its surroundings. It is also unable to process its high-definition map. However, the AV compensates by utilizing the TCI digitized data, which provides valuable information to aid in the AV's decision-making and navigation processes.

By examining these scenarios, it is possible to assess the impact and benefits of utilizing TCI digitized data on AV operation safety in real-world situations.

5.5.6 Simulation Results

5.5.6.1 Setting 1

- **Scenario 1**: In this scenario, the AV has fully operational sensor devices and can process its high-definition map. As it approaches a "STOP" sign, it accurately perceives the sign's presence and stops to allow vehicles on the service road to pass safely.
- **Scenario 2**: Here, the AV's sensor devices are non-operational, and it cannot process its high-definition map. Consequently, the vehicle fails to perceive the "STOP" sign and continues driving straight. This results in a collision with another vehicle traveling on the service road.
- **Scenario 3**: In this situation, the AV's sensor devices are also non-operational, preventing it from processing its high-definition map. However, the AV utilizes TCI digitized data as an alternative source of information. By considering the "STOP" sign in its decision-making process, the vehicle successfully avoids a collision with other vehicles on the service road.

5.5.6.2 Setting 2

- **Scenario 1**: In this scenario, the AV has fully operational sensor devices and can process its high-definition map. While traveling on the highway, it detects the "Work Zone—Speed Limit 40" sign and promptly reduces its speed in response.
- **Scenario 2**: Here, the AV's sensor devices are non-operational, and it cannot process its high-definition map. Consequently, the vehicle continues to travel at its original high speed, unaware of the "Work Zone—Speed Limit 40" sign.
- **Scenario 3**: In this situation, the AV's sensor devices are also non-operational, preventing it from processing its high-definition map. However, the AV utilizes TCI digitized data as an alternative source of information. By considering the "Work Zone—Speed Limit 40" sign in its decision-making process, the vehicle reduces its speed as expected.

5.6 DATA MANAGEMENT PLAN

The research team developed a comprehensive DMP that effectively oversees data maintenance throughout the project and beyond. The DMP encompasses the entire data lifecycle and outlines the necessary policies for efficient data management, ensuring the longevity and reliability of the data. The primary objectives of the DMP can be summarized as follows:

- Ensuring data adherence to standardized classifications.
- Verifying the validity and accuracy of the data.
- Upholding data integrity and internal consistency.
- Safeguarding and preserving primary data.
- Facilitating easy access to primary data.
- Streamlining data processing procedures to enhance efficiency.
- Enabling seamless integration of different datasets to enhance overall utility.

A core principle guiding the DMP is to retain all data in their original form as collected. This approach allows for flexibility in processing the data, such as applying filters, aggregating, or transforming them as needed.

The completed DMP will be provided as a separate document alongside the final report upon the project's conclusion. However, a draft version can be made available upon request at the current stage.

5.7 REFERENCE

[1] A. H. Ashara, S. Saleh, U. Hassan, and M. J. Kaura, "Traffic Management System in Abuja City Center, using Geographic Information Systems (GIS) and Global Positioning System (GPS). A case study of FCT, Nigeria," *J. Phys.: Conf. Ser.*, vol. 1529, no. 5, p. 052068, May 2020, doi: 10.1088/1742-6596/1529/5/052068.

[2] Y. Wen *et al.*, "Mapping dynamic road emissions for a megacity by using open-access traffic congestion index data," *Applied Energy*, vol. 260, p. 114357, Feb. 2020, doi: 10.1016/j.apenergy.2019.114357.

[3] "Learn About the Mobileye Mission and Vision | Mobileye AU." https://ims.mobileye.com/fleets/au/about/ (accessed Jun. 30, 2023).

[4] "Arlington Arlington » PlanUrStay," May 29, 2023. https://planurstay.com/st_location/united-states-of-america/arlington/ (accessed Jun. 30, 2023).

[5] S. O'Hear, "Mapillary, the crowdsourced database of street-level imagery, has been acquired by Facebook," *TechCrunch*, Jun. 18, 2020. https://techcrunch.com/2020/06/18/mapillary-facebook/ (accessed Jun. 30, 2023).

[6] "OpenStreetMap." https://www.openstreetmap.org/#map=5/38.007/-95.844 (accessed Jun. 30, 2023).

[7] J. Sander, M. Ester, H.-P. Kriegel, and X. Xu, "Density-Based Clustering in Spatial Databases: The Algorithm GDBSCAN and Its Applications," *Data Mining and Knowledge Discovery*, vol. 2, no. 2, pp. 169–194, Jun. 1998, doi: 10.1023/A:1009745219419.

CHAPTER 6. CONCLUSIONS AND RECOMMENDATIONS

6.1 FINDINGS AND CONCLUSIONS

TCIs play a vital role in guiding traffic flows (including AVs). How to effectively digitize TCIs and maintain a TCI inventory is an essential yet unanswered question for transportation agencies, not just in the United States but in all nations.

In this study, the research team developed an effective framework for the digitization, maintenance, and sharing of roadway assets, especially for TCIs. The research team explored existing solutions, potential resources/dataset, and possible legal issues for TCI digitization and inventory development. The team also investigated and digested the opinions from transportation agencies and AV companies on TCI data collection and utilization. The researchers proposed a web GIS-based framework for TCI digitizing and sharing and performed a comprehensive evaluation of the framework and selected TCI datasets. The team finally developed simulations based on various real-world scenarios and evaluated the benefits of incorporating TCI digitized data in enhancing the safety and operational measures of AVs. A web GIS-based application was developed based on the proposed TCI digitization framework, which provides a user-friendly interface to facilitate the TCI digitization process.

6.2 RECOMMENDATIONS

Based on this study, it is proved to beneficial if CAVs can digitally access the current state of traffic control, including static infrastructure, thereby adding redundancy to the function of correctly detecting and classifying traffic control units. In addition, integration of the live state of traffic control infrastructure in one place will greatly assist in traffic operations, optimization, and emergency response and recovery.

Asset inventory and condition are also key information that feeds into transportation asset management systems and the establishment of transportation planning and programming documents. Key elements of an asset management plan such as life cycle planning, financial plans, investment strategies, and risk analyses are all dependent on the asset inventory, condition, and performance information. Results of this research will allow transportation agencies (e.g., TxDOT) to communicate the type and locations of assets. It provides an easy-to-implement framework for digitizing and sharing the agency's TCIs. The new inventory data will also benefit roadway safety and planning-related research by providing accurately digitized asset datasets.

The insights derived from this study provide a clear path for future work, with the ultimate objective of enhancing the quality, comprehensiveness, and reliability of TCI data. Based on the current findings, future efforts can be geared towards the following areas:

1. ***Expanding Dataset Coverage***: Collaborate with the City of Arlington, Mobileye, and other potential data providers to extend the coverage of TCI data beyond the currently studied corridors.
2. ***Enhancing Open-Source Data Quality***: Efforts can be made to improve the data quality from open-source platforms like Mapillary and OSM through initiatives such as community drives for data collection and additional data validation measures.
3. ***Examining Additional TCI Elements***: The current study primarily focused on certain types of TCI data. Moving forward, the research team will evaluate other critical TCI elements such as traffic lights, traffic islands, and more. This will provide a more comprehensive understanding of traffic infrastructure.
4. ***Developing a Comprehensive Evaluation Framework***: The evaluation framework will be refined to be more robust and inclusive, incorporating additional performance metrics and considering factors like data freshness and update frequency. Also, the weight of each indicator can be discussed and adjusted based on project requirements.
5. ***Expanding the Study Area***: Given the local variability in TCI data quality and coverage, the research team aims to expand their study to encompass broader areas, possibly at a city-wide scale.
6. ***Investing in Proprietary Data Collection***: To ensure data reliability and control over the data collection areas, the team will explore investing in proprietary data collection efforts.
7. ***Leveraging Advanced Technologies***: The use of advanced technologies like AI and machine learning will be explored for automating the TCI data collection and evaluation process.
8. ***Fostering Collaborations***: The research team can seek partnerships with other stakeholders such as academic institutions, government bodies, and private companies to pool resources and share expertise to enhance TCI data quality and coverage.

APPENDIX C. DATA MANAGEMENT PLAN

TABLE OF CONTENTS

Terms and Acronyms

Acronym	Definition
AV	Automated Vehicles
DMPP	Data Management and Privacy Plan
FHWA	Federal Highways Administration
FIPS	Federal Information Processing Standards
FOIA	Freedom of Information Act
GIS	Geographic Information System
ITD	TxDOT Information Technology Division
ME8	Mobile Eye Sensor System
OBU	Onboard Unit
OEM	Original Equipment Manufacturer
PII	Personally Identifiable Information
PKI	Public Key Infrastructure
SCMS	Security Credential Management System
SPaT	Signal Phase Timing
SQL	Structed Query Language
SSL	Secure Sockets Layer
TCI	Traffic Control Infrastructure
TxDOT	Texas Department of Transportation
USDOT	United States Department of Transportation
WAVE	Wireless Access in Vehicular Environments devices

1 Introduction

Asset inventory and condition are key information that feed into planning and operations products such as transportation asset management systems, transportation planning and programming documents, and automated vehicle (AV) routing and operating systems. Key elements of an asset management plan such as life cycle planning, financial plans, investment strategies, and risk analyses are all dependent on the asset inventory, condition, and performance information. This research will allow the Receiving Agency to communicate the type and location of assets by providing a framework for digitizing and sharing the Receiving Agency's Traffic Control Infrastructure (TCI). The new inventory data will also benefit roadway safety and planning-related research by providing accurately digitized traffic control data.

1.1 Purpose

The Data Management and Privacy Plan (DMPP) for the project outlines the data that will be collected and how the data will be managed and shared with stakeholders and other entities throughout the project and beyond. It will also be used by the project team to ensure data are available to support framework development, design, deployment, and evaluation activities.

The DMPP also describes the needs of the TCI digitization framework deployment to protect the privacy of users, ensures secure communications, and outlines the plan to address these needs. It also describes how the project team, developed framework, and any system developed from it will address and "scrub" any personally identifiable information (PII) from data to be collected to meet the goals of the project. The plan also describes, at a high level, how the framework for an automated vehicle-generated TCI inventory and asset management system will provide security.

1.2 Document Organization

This document is organized as follows:

- Section 2—Overview of Digitized TCI Framework data and data considerations.
- Section 3—Overview of a high-level plan to manage distribution of TCI digitization mapping data.
- Section 4—Overview of security and privacy considerations for TCI digitization mapping data.
- Section 5—Overview of a high-level plan to ensure the security and privacy of AV-based TCI digitization mapping data inputs and outputs.

1.3 Proposed AV-to-TCI Asset Management System Overview

The 0-7128 project is split into five tasks over a 24-month period. At a very high level, the project scope includes the following deliverables:

- Development of a system concept of operations document and a framework for use of AVs to digitize and distribute TCI digitization mapping data for various public and private-sector operations and planning uses.
- Comprehensive framework evaluation plan with evaluation test results.

The TCI digitization mapping system conceptual framework contains seven modules, including:

- Project description and data need.
- Data acquisition.
- Data management.
- Quality check and legal issue assessment.
- TCI inventory creation.
- TCI digitization mapping data update.
- Data sharing and visualization.

Figure 88 shows a conceptual framework for TCI inventory establishment and management.

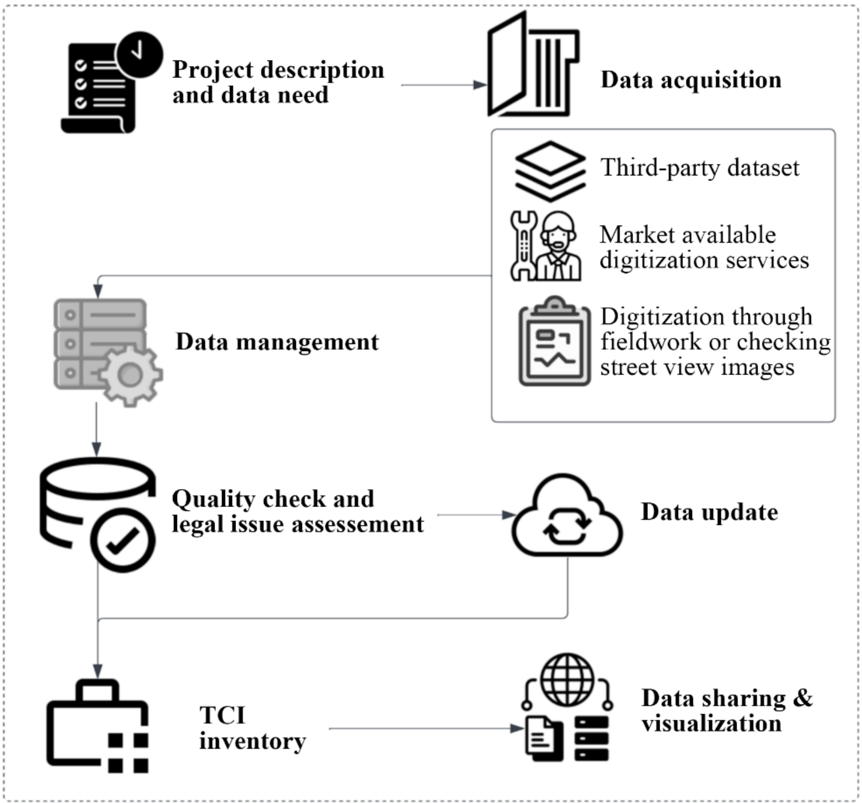


Figure 88. Conceptual Framework for TCI Inventory Establishment and Management.

The proposed framework is built using ArcGIS Online. Esri developed ArcGIS Online, a cloud-based geographic information system (GIS) platform (see Figure 89). With ArcGIS Online end users can apply create, store, share, and manage TCI digitization mapping data sets for maps, data analyses, and content generation within a geospatial information and services platform. ArcGIS Online offers various tools and functionalities for data analysis, mapping, and visualization. These include custom map creation, spatial analysis, and data and map sharing. ArcGIS Online also grants access to a vast repository of geospatial data and pre-built maps and apps. Individuals, organizations, and government agencies can utilize the platform for a wide range of applications, including the TCI digitized inventory in this study.

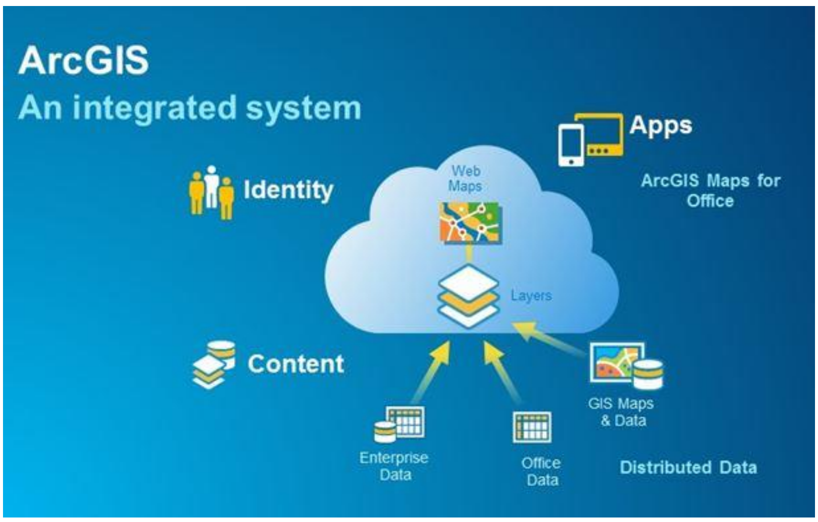
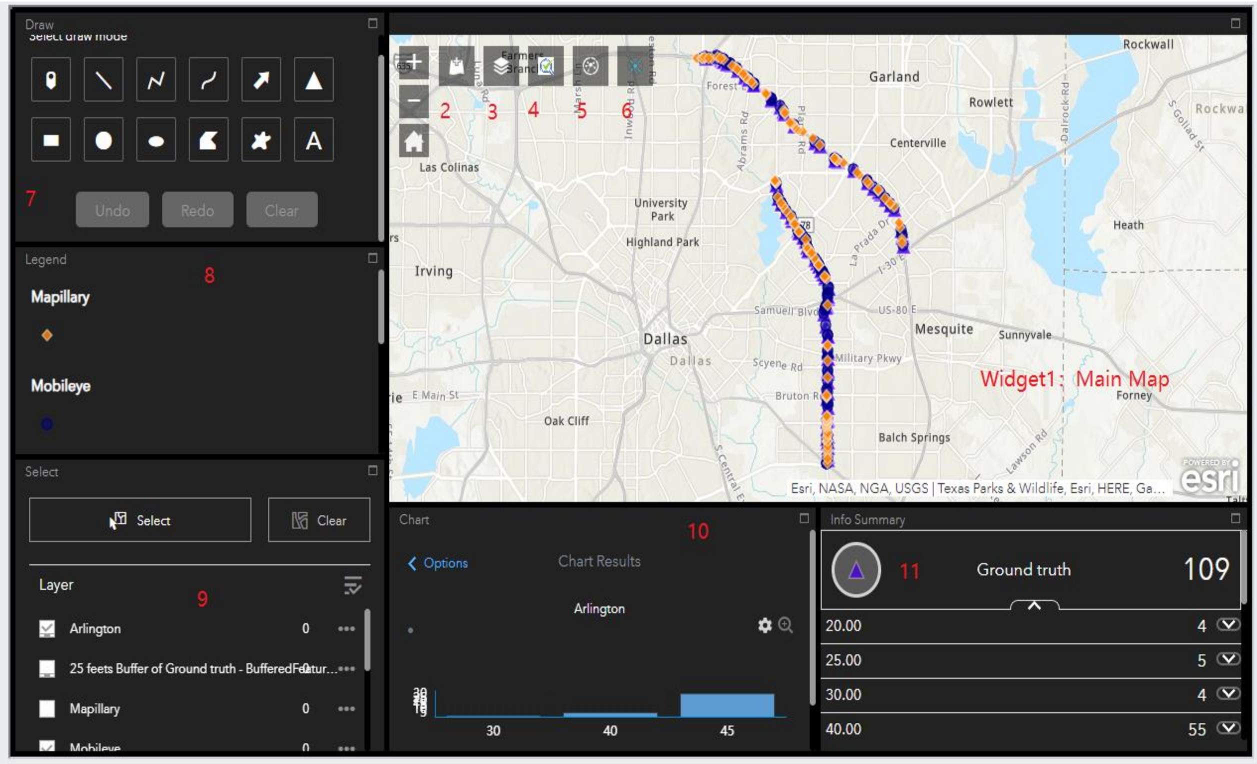


Figure 89. ArcGIS Online Platform

The TCI dataset is created by fusing data sources from the Texas Department of Transportation (TxDOT) and local governments' TCI inventory data, and third-party TCI digitization mapping data through the ArcGIS Online platform. Third-party data providers, such as Mobileye, supply some of the TCI digitization mapping data, which are then imported into ArcGIS Online as TCI digitization maps. An interactive dashboard within ArcGIS Online visually presents the data, allowing users to access, analyze, and interact with maps, charts, graphs, and other elements in a single view. The dashboard is customizable and supports various data visualization tools and widgets, such as heat maps, bar charts, pie charts, and tables. ArcGIS Online can access data from multiple sources, including ArcGIS Online maps, external sources, and cloud-based platforms. The dashboards can be easily shared and embedded in websites and other platforms for seamless data sharing and management. See Figure 90 for the developed platform.



(Powered by Esri ArcGIS Online Cloud Platform & Web App Builder)

Figure 90. Developed TCI Framework.

2 Data Management Approach

Once acquired, TCI digitization mapping data must be stored and managed. As part of this project, the research team is developing a data management plan, which will provide guidance about data protection, storage, and access. For the simulation project, only researchers and administrators will have access to the data. Should the project be implemented, the datasets may be publicly accessible. The standards for both the short- and long-term data management plan should acknowledge and accommodate the legal issues arising from data management.

2.1 Data Sharing

The ultimate product of this research is a TCI dataset rather than a digital map installed in and used by AVs. Therefore, this research project will not simulate the dissemination of the data to AVs. The research will compare how AVs drive with and without the data but will not address how the AV captures the data.

2.1.1 Data Sharing Agreements

TxDOT will enter into individual agreements with third-party data providers and AV stakeholders that outline the data to be installed, time period to which the data applies, and estimated accuracy of TCI. In general, this will involve the third-party data providers and AV stakeholders providing access to data from their integrated systems, potentially including TCI digitization mapping data and data from their AV and telematics systems. Data sharing agreements will also involve the project team sharing data from the TCI dataset, including new available data update alerts, among other information. These data-sharing agreements are not yet in development but they will bound what data will be received from and shared with the partners.

2.2 Intellectual Property

The project team has reviewed project data and determined that no intellectual property issues exist. Data generated and collected with internal systems will be the property of TxDOT. Negotiations and the aforementioned data sharing agreements with third-party data partners and AV stakeholders will grant TxDOT right, license, and privilege to use all or portions of the data shared by AV partners.

2.3 Archiving and Preservation

The TCI digitization mapping data asset and access system design is still underway for servers, data storage, and preservation, but this section provides a summary of potential approaches for archiving and preserving TCI digitization mapping data.

The ArcGIS Online web-based GIS application will be a central point through which TCI digitization maps will flow. TCI digitization maps from the dashboard will be updated periodically and available for transmission to AV stakeholder fleets operating in Texas, where they can process or collect the maps for integration as backup TCI maps envisioned in scenario three of the Task 4 simulation. TCI digitization mapping data can be access-restricted per the needs of

TxDOT to ensure security, and eventually have a two-way flow with data obtained directly from AV partners. The TCI digitization mapping data will be collected and stored on both local and online servers dedicated for statewide deployment. Some third-party service providers who obtain their TCI digitization mapping data from private probe-source vehicles may have concerns about TCI digitization mapping data being stored on public agency servers and potentially being open to Freedom of Information Act (FOIA) requests. If necessary, data will only be temporarily stored on TxDOT servers to facilitate transmission to private servers in ArcGIS Online. Project team members plan to access portions of the TCI digitization mapping data for testing, performance measures, and self-evaluation, maintaining separate research servers for data storage.

Preliminary efforts are underway to migrate some TxDOT intelligent transportation systems and associated servers to be hosted in the cloud. Project team members may also have remote access to third-party data during system integration and operation and may periodically collect data directly from those datasets for testing and evaluation. Project team members will follow TxDOT processes and procedures to acquire access to TxDOT servers and systems.

3 Data Management Plan

This section outlines the data to be collected and provides details on the management of that data.

3.1 Data Types and Sources

The data that will be collected to construct the TCI digitization mapping system and framework can be grouped into the following categories:

- TxDOT and Local Government TCI Asset Data—TCI digitization mapping data collected from comma-separated values (CSV) and shape file formats.
- Third-party Commercially Available Data—TCI digitization mapping data collected from third-party service providers that include Mobileye, Blyncsy, and Nexar.
- Raw Data—Street view imagery in photographic format.

Table identifies sample data elements from one third-party data service provider as an indicator of the type of TCI digitization mapping data that may be collected. For each road and utility feature identified, properties are depicted.

3.1.1 TxDOT and Local Government Data

TxDOT and local government TCI asset inventory data showing the location of TCI is primarily geographic data in CSV and shape file formats. These data are collected manually or by using vehicular equipment.

3.1.2 Third-Party Commercially Available Data (e.g., Mobileye)

Third-party commercially available data consist of inventory data similar to those owned by TxDOT, as well as raw photographic data (e.g., street view imagery) and connected vehicle trajectory data. The inventory data are collected differently depending on the company providing the data, but can include digitized photographic street view images. Regardless of collection method, the data are delivered to TxDOT in the form of CSV files. Trajectory data from connected vehicles are processed before delivery to TxDOT so that PII about private individuals is protected. Like inventory data, trajectory data will be delivered to TxDOT as CSV files consisting of a series of points that describe vehicle trajectory, and accessed through a license, which will be renewed each year. Third-party data will be purchased from a data aggregator who will send the data to a cloud storage service. For the research simulation project, the contractual relationship will be between the Texas A&M Transportation Institute and the aggregator. Should the project move forward to implementation, TxDOT would contract directly with the aggregator.

3.1.3 Raw TCI Digitization Mapping Data

Raw TCI digitization mapping data include street view imagery and will function as a back-up option to complement existing datasets with imagery. These will be delivered to TxDOT as photographic files. Their content will consist of public and exterior spaces. These will be acquired and paid for on a one-time, rather than subscription, basis.

Table 14 provides a list of survey and interview data elements and descriptions.

Table 14. Sample TCI Third-Party Data Service Provider Features and Properties.

Mobileye Road Asset Data		**Mobileye Utility Asset Data**	
Features	**Properties**	**Features**	**Properties**
Road edges	Type (guard rails, curb etc.)	Streetlight poles*	• Type (Mobileye Sensor System (ME8) and select Original Equipment Manufacturers (OEMs) • Tilt Angle**
Road markings	Type (e.g., left-road arrow)	Power poles*	
Lane markings	Type (e.g., solid, dashed, etc.)	**Mobileye Pavement Anomaly Detection**	
Lane width	Lane width	Potholes*, **	• Width • Length • Area • Image (option)**
Traffic signs	Type (overhead, speed limit, stop, etc.) Azimuth	Mobileye Road Image Capture	
Traffic lights	Type (vertical 2 spot, pedestrian, etc.)	Road Image Capture***	Time of Image Location of Image
Poles	Type (ME8 and select OEMs) Tilt Angle**		
Crosswalks	Type (Zebra Crossing, Solid Crossing, etc.)		

**Only available through Mobileye 8 Connect. **Experimental capability. ***Faces and license plates are obfuscated for privacy.*

3.2 Data Collection and Transmission

Data sources and sample features are outlined for TCI digitization mapping data elements in the above sections. This project data will be collected and stored to support the development of the TCI digitization mapping data system and framework.

3.2.1 Data Storage

Collected public, third-party, and raw data will be stored on secured and encrypted servers hosted and maintained by TxDOT, ArcGIS Online, and team members. Third-party data are collected, stored, and periodically updated onto servers dedicated for statewide TCI digitization map deployments shared with TxDOT districts and AV stakeholders to be further aggregated by the project team for further performance measures and evaluation. Some data may be accessible and recordable by project team members on separate data servers.

3.2.2 Frequency

The project team will collect as much TCI digitization mapping data as practical within system, budget, and time constraints to meet project objectives, performance measures, and self-evaluation. The frequency for these updates to the TCI digitization mapping data will be determined as the system is further completed and as further constraints arise that impact schedules. It is anticipated to meet the demands of both AV stakeholders and TxDOT stakeholders that the collection and distribution frequency will range from nearly continuous to as needed or requested. For instance, data from TxDOT and local government TCI asset data will be continuously updated through the TCI digitization mapping system; however, third-party data may be delivered on an annual update basis or on a more frequent basis given funding constraints of any potential partner agreement.

3.2.3 Relationship to Performance Measures

Performance measures will be informed by the comprehensive evaluation plan in Task 5. The project team will ensure performance is connected to each data feature. Some data features, such as traffic light and lane marking logs, likely support evaluation of multiple performance measures. Other data features are specific to one or a small number of measures.

3.3 Baseline Data

Baseline data will be associated with each data feature that is prioritized by TxDOT and AV stakeholders. Baseline data will be collected before the TCI digitization mapping system becomes operational to serve as a point of comparison for evaluating the hypotheses identified for each performance measure.

4 Data Security and Privacy

Security and privacy of applications and data is critical for the acceptance of AV systems. This section describes the security and privacy concepts and considerations for the TCI digitization mapping system.

4.1 IEEE 1609.2

The IEEE 1609.2 standard defines secure message formats and processing for Wireless Access in Vehicular Environments devices, potentially including the digitization mapping data that will be incorporated by TxDOT vehicle fleets and AV stakeholders as part of the project, to create, decode, sign, and verify responsive AV functions to the TCI environment in real time when automation is engaged. It also outlines functions to administer and support core security functions. The project team intends to procure third-party data and webtool services that conform to and implement these security standards.

4.2 Security Credential Management System

The Security Credential Management System (SCMS) is an important component of a secure and private AV ecosystem. The SCMS employs the use of a public key infrastructure where a central authority issues certificates to enrolled devices (i.e., onboard units [OBUs] on vehicles). These certificates are used to sign messages prior to transmission and are changed on a regular basis to protect end-user privacy, to assure the message recipient that the transmitter is authorized, and to ensure the integrity of the message. In general, enrollment of devices in an SCMS is implemented in such a way that no single entity has sufficient information to identify a specific device. It would take the cooperation of multiple entities (i.e., in response to a court order) to re-identify an enrolled device.

SCMSs also employ misbehavior detection to detect malfunctioning or compromised devices (e.g., OBU broadcasting basic safety messages with unrealistic vehicle speed, etc.). If a misbehaving device is identified, its certificates that are no longer trusted can be placed on a certificate revocation list or the device can be internally blacklisted by the SCMS vendor.

The project team is evaluating available SCMS solutions to support TxDOT's procurement of third-party service providers for TCI digitization mapping. Blyncsy and Mobileye do not have any SCMS components as of the date of this report, while Nexar does. The project team will consider whether an SCMS Evaluation Plan is necessary for the purposes of the development of the TCI digitization mapping system and framework. If this is warranted, the project team will conduct technical evaluations of available SCMS solutions, including enrollment of data services, and verification of loading certificates from those services, among other things.

4.3 Privacy Concerns

The project team wants to ensure that end-user privacy is not sacrificed while realizing the operations, maintenance, safety, and mobility benefits afforded by the TCI digitization mapping system. The TCI digitization mapping system is being designed and implemented to maintain privacy to the highest extent possible.

4.4 RISK ANALYSIS

Risk analysis helps identify and mitigate threats to information, resources, processing, disclosure, communications, integrity, and availability. The following categories of risk have been identified for the TCI digitization mapping system:

- Unauthorized Access.
- Viewing Confidential/Private Data.

4.4.1 Unauthorized Access

Unauthorized access to the TCI digitization mapping system could allow it to be used for malicious purposes, including:

- Sending invalid or inaccurate digitization mapping updates to end users.
- Manipulating data coming from third-party service providers and associated vehicles on the road network.

4.4.2 Viewing Confidential/Private Data

The project team does not anticipate PII data being collected or stored; however, some data may be considered private or confidential and some anonymized data always contain the risk of being re-identified. Potential risks of exposure of private or confidential information could include:

- Unauthorized use of contact information associated with system login accounts.
- Unauthorized use of server information.
- Use of online server login information to manipulate configuration or data streams.

5 Data Security and Privacy Plan

This section describes the approach to protecting the privacy of users and ensuring secure communications throughout the TCI digitization mapping. Core components of the system will be implemented within the ArcGIS Online and the TxDOT business network with which it sources for data inputs; these policies and measures largely follow the TxDOT Information Resources and Security Requirements (*1*).

5.1 Control Measures

The TCI digitization mapping system will include an ArcGIS online component that consumes information from public and third-party data services and distributes information to various end users including potential AV stakeholders for their incorporation into routine automation functions. To protect the underlying TxDOT network infrastructure from unauthorized access, misuse, malfunction, modification, destruction, or improper disclosure, logical security measures are in place, mostly implemented via the TxDOT Information Technology Division.

This section describes the technical and policy controls used to ensure security and privacy within the TCI digitization mapping system.

5.1.1 Technical Controls

The following technical control measures will be implemented to ensure security and privacy of project data.

5.1.1.1 Access Control

TCI digitization mapping system users include TxDOT and non-TxDOT personnel who access the system through the TxDOT Business Network. Access to the TxDOT TCI digitization mapping system network infrastructure and applications will be restricted by role-based authorization and authentication, with minimum required privileges provided to users and administrators. Permission levels will be determined by TxDOT administrators and may include:

- Read-only Permission—The authenticated user is allowed to view data in the system but does not have permission to create, delete, update, or control data.
- Read, Create, Delete, Update, Control Permission—The authenticated user is allowed to view and alter data in the system, including public and third-party data sets.
- Administrative Permission—The authenticated user can make global modifications in the system.

To mitigate unauthorized access to the TCI digitization mapping system, access will be restricted by and limited to appropriate users. Once access has been granted, an authenticated user can access the system and perform actions allowed by the user permissions. TxDOT administrators will be responsible for maintaining and monitoring the list of authenticated users and should remove users who should no longer have access.

5.1.1.2 Encryption

All data distributed between TxDOT, ArcGIS Online servers, project team members, and end users will be considered for potential encryption using secure sockets layer encryption according to TxDOT network policies. Data flowing through external interfaces (e.g., SCMS) should also be encrypted.

5.1.1.3 Database

TCI digitization mapping data at rest will be stored and accessible on secure structured query language server databases maintained by TxDOT or other members of the project team for conveyance to ArcGIS Online servers at appropriate times. Physical servers will be located at TxDOT or project team member facilities and behind network firewalls. Access to these database(s) will be further restricted by requiring authenticated users to log in. TxDOT and/or project team member organization administrators will maintain and monitor the list of authenticated users.

5.1.2 Policy Controls

TxDOT and the rest of the project team employ industry best practices for ensuring a safe and secure computing and networking environment. Each project team member defines policies and implements processes and procedures for information technology security, including for:

- Malware defenses.
- Incident management.
- Device and software inventories.
- Skills and training.
- Network controls.
- Admin privileges.
- Audits.

TxDOT's policies and procedures relating to the access of computers and data are governed by statutes, codes, and procedures, including among others Texas Government Code Title 5, Subtitle A, Chapter 552—Public Information (*2*).

5.1.2.1 Use of Collected Data

The project team will anonymize and sanitize any information that is derived from TCI digitization mapping data to be documented in project materials. This includes results from third-party data services, which may contain sensitive and PII elements of vehicle and system data.

5.2 COMPLIANCE

The project team and its member organizations are responsible for ensuring that all TCI digitization mapping system team members and project participants and partners, especially those that have access to TCI digitization mapping data and potential PII, abide by the policies identified in the DMPP. All project team members are employees of their respective organizations.

6 References

[1] TxDOT (2023) 'Information Resources and Security Requirements'. Available at: https://ftp.txdot.gov/pub/txdot/itd/cybersecurity/information-resources-and-security-requirements.pdf (Accessed: 31 July 2023).

[2] Texas State Legislature (1993) *Government Code Title 5. Open Government: Ethics; Subtitle A. Open Government; Chapter 552. Public Information; Subchapter A. General Provisions*. Available at: https://statutes.capitol.texas.gov/Docs/GV/htm/GV.552.htm (Accessed: 31 July 2023).

APPENDIX D. VALUE OF RESEARCH

Following the procedures outlined in TxDOT's *University Handbook*, the research teams assessed the potential value of TxDOT Research Project 0-7128: "***Digitizing Traffic Control Infrastructure for Autonomous Vehicles (AV)"***. Table 15 shows the areas where this research project would be expected to generate benefits for TxDOT.

Table 15. Selected Benefits Areas Associated with Project 0-7128.

Selected	Functional Area	QUAL	ECON	Both	TxDOT	State	Both
X	Level of Knowledge	X			X		
X	Quality of Life	X					X
X	Customer Satisfaction	X					X
X	Infrastructure Condition	X					X
X	Intelligent Transportation Systems	X			X		
X	Safety			X			X

The value of research assessment included qualitative and economic assessments of the potential benefits. This section presents the findings from the research team's assessment of research value.

1.1 QUALITATIVE BENEFITS

The research team identified the following qualitative benefits associated with the outcomes of this research project:

- Improved safety.
- Increased level of knowledge of TxDOT personnel.
- Increase quality of life for Texas motorists and TxDOT personnel.
- Improved customer satisfaction.
- Use of Intelligent Transportation System (ITS) technologies.

The following provides a detailed explanation of the benefits identified by the research team.

1.1.1 Improved Safety

Improving safety is perhaps the most significant benefit of this research. Through the implementation of advanced traffic control infrastructure (TCI) digitization techniques and approaches, this research project contributes to a safer environment for AVs. By optimizing traffic control systems and communication, AVs can navigate roadways with improved reliability and reduced risk of accidents.

1.1.2 Level of Knowledge

One of the primary benefits of this research project was to augment the knowledge and expertise of TxDOT personnel in the field of TCI digitization. By equipping TxDOT professionals with the latest insights and best practices, the state agency can better adapt to the evolving transportation landscape and play a pivotal role in ensuring the success of AVs on Texas roads.

1.1.3 Quality of Life

The research has the potential to improve the quality of life for Texas motor vehicle operators. As AVs become increasingly prevalent, the quality of life for both Texas motorists and TxDOT personnel will see significant improvements. Reduced traffic congestion, safer roadways, and increased efficiency in transportation operations will contribute to a more enjoyable and convenient daily life for all stakeholders.

1.1.4 Customer Satisfaction

This research also has the potential to improve customer satisfaction with TxDOT. With the implementation of TCI digitization technologies and the seamless integration of intelligent transportation system (ITS) technologies, customer satisfaction among individuals utilizing AVs on Texas roads will be greatly improved. These enhancements in road infrastructure and traffic management will ensure a smoother and more satisfying travel experience.

1.1.5 Intelligent Transportation System

This research illustrates the value of TxDOT's use of ITS technologies. As part of this research, the research team developed an effective framework for the digitization, maintenance, and sharing of roadway assets, especially for TCIs. The research team evaluated available solutions (commercial, open-source, and public), investigated potential legal issues, and proposed new approaches by leveraging emerging data sources and techniques. Simulations based on various real-world scenarios were developed to evaluate the benefits of incorporating TCI digitized data in enhancing the safety and operational performance of AVs. By harnessing the developed asset digitization techniques, Texas can stay at the forefront of transportation innovation and infrastructure development, fostering sustainable and efficient mobility solutions.

1.2 ECONOMIC ASSESSMENT

The research team conducted an economic assessment of the benefits associated with the outcomes of this research project. The research team identified one functional area where TxDOT may benefit the most from this research: Safety.

The research team identified the following potential safety benefits as an outcome of this research effort:

- Reductions in AV crashes.

The following provided a summary of the data and assumptions used to compute the economic value of this research project.

1.2.1 Reduction in AV Crashes

NHTSA has issued a Standing General Order (the General Order)[1] requiring identified manufacturers and operators to report to the agency certain crashes involving vehicles equipped with automated driving systems or SAE Level-2 advanced driver assistance systems. After that, NHTSA maintains a database of ADS crash data that the agency received under the General Order. As of August 15th, 2023, 386 ADS crashes were reported in the United States and 24 of them happened in Texas. 1,112 Level-2 ADAS crashes were reported in the United States and 96 of them occurred in Texas.

Table 16. Number of AV Crashes Occurring in the United States (July 2021 to August 2023)

Year	Fatal (K)	Incapacitating (A)	Non-Incapacitating (B)	Complaint (C)	Unknown Injury (O)
ADS Crashes	0	1	6	32	329
Level-2 ADAS Crashes	28	15	27	44	998
ADS & ADAS Crashes in Total	28	16	33	76	1,327
Yearly Average	13.442	3.841	15.843	36.486	637.062

The research team assumed that one percent of those crashes listed in Table 16 were addressable by this research to estimate benefits associated with reducing AV-related crashes.

The researcher developed the following potential reductions in crash types because of this research:

- Fatalities (K) = 0.1344 / year
- Incapacitating Injuries (A) = 0.0384 / year
- Non-Incapacitating Injuries (B) =0.1584 / year

- Compliant of injuries (C) = 0.3649 / year
- Unknown injury (O) = 6.371/year

Figure 91 shows the Value of Reduced Fatality recommended by the US Department of Transportation for estimating the value of reducing crashes of certain types. These values are in 2017 dollars. The research team adjusted to 2023 dollars using standard inflation rates based on the Consumer Price Index. After making this conversion, the research team estimated the potential cost saving caused by reducing AV-related crashes. Table 17 shows the results of the potential benefits associated with enhancing safety for AVs.

Recommended Monetized Value(s)

MAIS Level	Severity	Fraction of VSL	Unit value ($2017)
MAIS 1	Minor	0.003	$28,800
MAIS 2	Moderate	0.047	$451,200
MAIS 3	Serious	0.105	$1,008,000
MAIS 4	Severe	0.266	$2,553,600
MAIS 5	Critical	0.593	$5,692,800
Fatal	Not Survivable	1.000	$9,600,000

KABCO Level	Monetized Value
O – No Injury	$3,200
C – Possible Injury	$63,900
B – Non-incapacitating	$125,000
A – Incapacitating	$459,100
K – Killed	$9,600,000
U – Injured (Severity Unknown)	$174,000
# Accidents Reported (Unknown if Injured)	$132,200

References and Notes

Guidance on Treatment of the Economic Value of a Statistical Life in U.S. Department of Transportation Analyses (2016) https://www.transportation.gov/office-policy/transportation-policy/revised-departmental-guidance-on-valuation-of-a-statistical-life-in-economic-analysis

Note:
The KABCO level values shown result from multiplying the KABCO-level accident's associated MAIS-level probabilities by the recommended unit Value of Injuries given in the MAIS level table, and then summing the products. Accident data may not be presented on an annual basis when it is provided to applicants (i.e. an available report requested in Fall 2011 may record total accidents from 2005-2010). For the purposes of the BCA, is important to annualize data when possible.

Source: Federal Highway Administration[1].

Figure 91. Estimated Cost of Collisions, by Collision Severity

Table 17. Estimates of Potential Crash Reduction Benefits.

Crash Type	Estimated # of Crashes	Cost per Crash Type (2017)	Cost per Crash Type (2022)	Value of Research
Fatality (K)	0.1344	$ 9,600,000	$12,096,000	$1,625,702
Incapacity Injury (A)	0.0384	$ 459,100	$578,466	$22,213
Non-Incapacitating Injury (B)	0.1584	$125,000	$157,500	$24,948
Possible Injury (C)	0.3649	$ 63,900	$80,514	$29,380
No Injury (O)	6.371	$ 3,200	$4,032	$25,688
			TOTAL	**$1,727,931**

Notes: The Consumer Price Index is 1.26 when converting the 2017 dollar to the 2023 dollar (Bureau of Labor Statistics)

1.2.2 Value of Research Computation

Using these estimates, the research team computed the value of this research as follows based on a ten-year horizon:

The benefit-cost ratio of this project is:

$$\frac{B}{C} = \frac{\$1{,}727{,}931 * 10\ Years}{\$359{,}394\ (Project\ Costs)} * \left(\frac{120}{1496}\right) = 7.71$$

Notes: The portion of AV related crashes in Texas is calculated by the NHTSA crash numbers by states.