RESEARCH

UDOT

Keeping Utah Moving

Utah Department of Transportation - Research Division
4501 South 2700 West - P.O. Box 148410 - SLC,UT 84114-8410

Report No. UT-26.14

A GAP ANALYSIS OF EXISTING DATABASE MANAGEMENT PRACTICES AND ONTOLOGY DEVELOPMENT FOR UDOT TIER 1 SAFETY ASSETS

Prepared For:

Utah Department of Transportation
Research & Innovation Division

Final Report
June 2026

DISCLAIMER

The authors alone are responsible for the preparation and accuracy of the information, data, analysis, discussions, recommendations, and conclusions presented herein. The contents do not necessarily reflect the views, opinions, endorsements, or policies of the Utah Department of Transportation or the U.S. Department of Transportation. The Utah Department of Transportation makes no representation or warranty of any kind, and assumes no liability therefore.

ACKNOWLEDGMENTS

The authors acknowledge the Utah Department of Transportation (UDOT) for funding this research, and the following individuals from UDOT on the Technical Advisory Committee for helping to guide the research:

- Chris Whipple, UDOT Asset Management Program Manager
- Jennifer Volkening, UDOT Data Analytics and Governance, Director
- Katherine Colburn, UDOT Assistant Freeway Operations Engineer
- Kevin Nichol, UDOT Research Project Manager
- Marcelo Manzo, UDOT Data Scientist
- Tim Predmore, UDOT Asset Management Data Scientist

TECHNICAL REPORT ABSTRACT

1. Report No. UT-26.14	2. Government Accession No. N/A	3. Recipient's Catalog No. N/A
4. Title and Subtitle A Gap Analysis of Existing Database Management Practices and Ontology Development for UDOT Tier 1 Safety Assets		5. Report Date June 2026 6. Performing Organization Code
7. Author(s) Brig Dean Adams, Mohsen Zaker Esteghamati		8. Performing Organization Report No.
9. Performing Organization Name and Address Utah State University Department of Civil and Environmental Engineering Old Main Hill Logan, Utah, 84321		10. Work Unit No. 5H094 86H 11. Contract or Grant No. 25-8903
12. Sponsoring Agency Name and Address Utah Department of Transportation 4501 South 2700 West P.O. Box 148410 Salt Lake City, UT 84114-8410		13. Type of Report & Period Covered Final Feb 2025 to June 2026 14. Sponsoring Agency Code PIC No. UT24.405
15. Supplementary Notes Prepared in cooperation with the Utah Department of Transportation and the U.S. Department of Transportation, Federal Highway Administration		

16. Abstract

This study evaluates transportation asset data management practices within the Utah Department of Transportation (UDOT) and identifies opportunities to improve data governance, integration, and interoperability through ontology-informed enterprise data management principles. Current practices were assessed through stakeholder surveys, asset database reviews, and a gap analysis comparing existing conditions with best practices from the literature on enterprise data management, ontology engineering, and data lake architecture. Findings revealed challenges including fragmented data storage, inconsistent terminology, limited metadata documentation, unclear stewardship responsibilities, and reliance on unstructured condition reporting. These issues reduce data usability, hinder integration efforts, and limit advanced analysis capabilities. To address these challenges, a proof-of-concept ontology and standardized schema were developed for selected transportation safety assets, including barriers, pavement, sign faces, and sign assemblies. The proposed schema introduces structured metadata categories and new fields for asset status, condition, and damage classification to improve consistency and reduce reliance on free-text data entry. Ontology development principles informed the organization of asset concepts, relationships, and metadata, creating a semantic foundation for future enterprise applications. The framework was implemented and validated within Google Cloud Dataplex using business glossaries, metadata catalogs, and data quality rules. Results demonstrated the feasibility of ontology-informed data management practices for improving data discoverability, governance, interoperability, and decision support, while establishing a foundation for future enterprise-wide transportation asset data integration at UDOT.

17. Key Words Transportation Safety Asset Management, Data Governance, Ontology, Ontology, Data Integration		18. Distribution Statement Not restricted. Available through: UDOT Research Division 4501 South 2700 West P.O. Box 148410 Salt Lake City, UT 84114-8410 www.udot.utah.gov/go/research		23. Registrant's Seal N/A
19. Security Classification (of this report) Unclassified	20. Security Classification (of this page) Unclassified	21. No. of Pages 94	22. Price N/A	

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ACRONYMS

API	Application Programming Interface
EDM	Enterprise Data Management
EDSS	Enterprise Data Storage System
ELT	Extract-Load-Transform
ETL	Extract-Transform-Load
FAIR	Findable, Accessible, Interoperable, and Reusable
UDOT	Utah Department of Transportation
TAMP	Transportation Asset Management Plan

EXECUTIVE SUMMARY

The Utah Department of Transportation (UDOT) manages a large and diverse portfolio of assets in its transportation network, and its operations rely on vast amounts of data to support effective decision-making. However, existing transportation asset data environments are often fragmented across multiple systems, formats, and management practices, limiting the ability to consistently leverage data across organizational domains. A survey-based review of current UDOT data management practices for safety assets identified several challenges affecting data usability and consistency. Key issues included fragmented data storage, inconsistent terminology, limited metadata documentation, informal governance processes, unclear stewardship responsibilities, and heavy reliance on unstructured comment fields for recording condition information. These limitations reduce confidence in the data, complicate integration efforts, and limit the ability to perform advanced querying and analysis across datasets. Therefore, this project presents a proof-of-concept ontology-informed schema intended to improve semantic consistency, metadata organization, governance practices, and interoperability within UDOT transportation safety asset datasets.

To address existing data management challenges, the developed framework introduces a standardized schema to improve consistency and interoperability. The development process involved evaluating existing asset inventories, identifying data gaps and inconsistencies, and incorporating stakeholder input. The schema organized asset data into several use-based categories and incorporated new structured fields, including asset status, condition type, and damage classification. These additions reduced reliance on inconsistent condition reporting and created opportunities for more effective querying, filtering, and future analytical applications. The schema also emphasized the role of metadata management and governance in supporting long-term enterprise data quality and accessibility. Ontology development principles informed the schema's structure and metadata organization by defining relationships among assets, their properties, and associated governance information. The resulting schema demonstrates how ontology-driven approaches can create a semantic foundation for future enterprise applications. Establishing standardized relationships and terminology improves the ability to integrate

heterogeneous datasets and creates opportunities for enhanced decision support, automated reasoning, and future digital twin applications.

The project focused on a few selected transportation safety assets, including barriers, pavement, sign faces, and sign assemblies, to demonstrate how structured metadata and ontology principles can support improved data management capabilities. The framework was implemented in Google Cloud Dataplex using tools such as *business glossaries, metadata catalogs*, *aspect types*, and *aspects* to establish a more standardized, semantically aligned data environment. These components were used to organize and define asset information related to governance, geolocation, collection processes, physical properties, and condition reporting. Validation activities demonstrated that the proposed schema and metadata structures could be successfully operationalized within Dataplex. Testing demonstrated the ability to create and manage glossary entries, metadata aspects, schema categories, and data quality rules within the platform environment. These capabilities support improved discoverability, consistency, and governance of transportation asset information while also establishing a scalable framework for future expansion. This project concludes that ontology-informed enterprise data management practices can significantly improve the organization, governance, and usability of transportation asset data. Despite the limited scope of this project, it lays the foundation for future work involving broader dataset integration, enhanced semantic relationship modeling, standardized lifecycle management practices, and more advanced analytical capabilities. Continued development of these systems and broader efforts to develop ontologies for other assets and data needs have the potential to improve interoperability, strengthen data-driven decision-making, and support the long-term modernization of transportation asset management at UDOT.

1.0 INTRODUCTION

1.1 Problem Statement

Due to the increasing number of assets managed and the volume of asset data collected by the Utah Department of Transportation (UDOT), modernized asset management practices have become necessary (Foxxe et al., 2026; U.S. Department of Transportation, 2026). Current practices lead to disparate data management across groups, including inconsistent schemas, variable data-tracking methods, and conflicting understandings of data stewardship. These issues lead to a lack of interoperability between datasets and within UDOT groups, increasing the time and resources required to leverage valuable data.

Data is the lifeblood of any organization's decision-making processes. If data is siloed, uninterpretable, or unreliable, it becomes more difficult to make effective decisions. Every large organization faces different challenges in managing its data, ranging from volume to complexity. Given the nature of data management, there is no universal process for designing and implementing effective management principles. To determine which practices or solutions need to be implemented, a review of asset databases, a survey of UDOT staff, and a gap analysis against the literature must be conducted to identify barriers and solutions for the current data management landscape.

1.2 Project Context

Utah is a rapidly growing state, ranking in the top five of the United States in population, economy, and infrastructure growth (Bureau, 2018; Foxxe et al., 2026; *Infrastructure Report Card | Utah Section*, 2025). As Utah continues to grow in these dimensions, the demand for robust data management practices within state agencies increases accordingly. The annual UDOT Statistical Summary for 2024 reports approximately 61.7 billion dollars in assets managed by the agency, excluding 1.68 billion dollars in projects for 2025 (*UDOT Announces New and Ongoing 2025 Construction*, 2025). Because of the value and volume of these assets, it is necessary to ensure that relevant data is collected at each stage of each asset's lifecycle to track currently relevant information (e.g., asset value, asset condition) as well as prepare UDOT

for future artificial intelligence applications, such as predictive maintenance and smart city integration (Pandey & Esteghamati, 2026; Wu et al., 2025; Yan et al., 2023). Furthermore, it is critical that the groups responsible for these assets during each phase of the asset's lifecycle can pass along data in a way that is understandable between groups without creating a need to trace where the data comes from, how it is being leveraged, and why it is being collected in the first place.

Legacy systems and inconsistent practices at UDOT create barriers to effective data utilization. Particularly significant challenges arise from inconsistent taxonomy, a lack of standard quality control mechanisms, and gaps in data lifecycle management. Nevertheless, various methods are available to overcome some of these difficulties (Lei et al., 2022; Nambiar & Mundra, 2022; Rashid et al., 2020). Addressing these barriers in a timely manner will prevent further organizational discord and enable UDOT to integrate data more effectively in the future.

The implementation of ontologies within transportation agencies remains an emerging area of practice. While some agencies have adopted some form of enterprise ontology, this approach remains relatively uncommon (Boadi et al., 2026; Sudre et al., 2008; Xu et al., 2019). Existing efforts are often focused on specific applications such as asset management or digital twin applications (Du et al., 2023a; Ren et al., 2019; Wu et al., 2025). As transportation agencies continue to generate larger volumes of data from increasingly diverse sources, interest in ontology-driven approaches has grown due to their potential to improve interoperability, data quality, and cross-system integration. However, challenges related to governance, stakeholder alignment, technical complexity, and long-term maintenance have limited widespread adoption, leaving the field in the early stages of development.

1.3 Objectives

This project aims to address three main objectives:

1. Review existing database management practices, including data collection, metadata, schemas, storage, and usage of transportation safety asset data.

2. Perform a gap analysis on current data management practices to understand how a formal ontology can be developed to unify data management practices across UDOT.

3. Develop a proof-of-concept initial schema for a limited set of UDOT assets, informed by the findings from Objectives 1 and 2.

1.4 Scope

This project aims to understand current database management practices for traffic and safety asset data and to outline a path toward more unified data governance across UDOT groups through ontology-driven data management. The datasets selected for this report include barriers, pavement, sign faces, and sign supports. The main tasks include a review of the literature on asset data management in related fields, a review of current internal data management practices, a gap analysis between an idealized state of data management and current practices, the acquisition and analysis of stakeholder feedback, and the development of a proof-of-concept ontology and corresponding schema for the selected datasets.

1.5 Outline of Report

This report is organized into six chapters. Chapter 1 introduces the study by describing the motivating factors for adopting modern data management practices, defining the problem statement, stating the study objectives, and outlining the project's scope. Chapter 2 reviews the existing literature on data management, ontology development, and data lake architecture in engineering, data science, and related fields. Chapter 3 details UDOT's current data management practices and compares them with an ideal state. Chapter 4 describes the development process for the proof-of-concept management framework and schema, providing insight into the iterative changes made and their purposes. Finally, Chapter 5 offers a roadmap for implementing the report's findings and details recommendations for the practical application of ontology.

2.0 LITERATURE REVIEW

2.1 Overview

This chapter provides a literature review of previous work on enterprise data management (EDM), ontologies, and data lake architecture. The EDM, ontology, and data lake architecture sections review how other research has approached database management and taxonomy in relevant fields. Drawing on principles developed across engineering, data science, and database management research, this chapter develops a framework to guide effective data management practices for the study.

2.2 Enterprise Data Management

Enterprise data management (EDM) is the organizational standard that ensures data is managed in line with the enterprise's goals (Otto et al., 2024). By developing a high-level framework, data organization can become more consistent, reliable, and accessible. EDM has evolved greatly over its history, but has remained a mainstay in the operations of large organizations since its early beginnings (Nambiar & Mundra, 2022). The core pieces of EDM are governance, integration, storage, access, and quality (Al-Badi et al., 2018; Brous et al., 2016). Determining the properties of these components sets the trajectory for data practices across organizations and tracking them ensures continued effectiveness. The following subsections will discuss critical components of EDM, including data governance, integration, storage, quality, and access.

2.2.1 Data Governance

Data governance establishes the framework for decision rights and accountability to ensure effective data use. The framework should include methods for aligning data to organizational needs, monitoring and enforcing compliance policies, and ensuring a common understanding of data across groups (Brous et al., 2016). Governance responsibilities can be assigned in several ways, but the ultimate decision depends on the enterprise's needs. Across the literature, the strategies invoked have functioned more as suggestions than as strict, prescriptive rules. The reason for this approach is that every organization faces different infrastructure

challenges, mission parameters, unique cultures, and financial constraints (Office of Management and Budget, 2020). Large enterprises, such as state or federal agencies, will require very different governance guidelines than small enterprises, such as small businesses or cities.

A highly critical aspect of data governance is that it must be viewed as a collaborative endeavor rather than an act of control (Benfeldt et al., 2020). Governance mediates organizational discord, an inherent part of any large group, by assigning decision-making responsibilities and accountability to individuals or groups. When assigning these responsibilities, several factors should be considered. First, a great deal of attention should be paid to the systems associated with the data in question (United Nations Conference on Trade and Development, 2025). When discussing data systems, there are two pertinent sides. The two sides, EDM design and operations, work together to inform how data governance should work in an organization. The people associated with those groups must work together in order to ensure reliable and effective policies (Brous et al., 2016; Janssen et al., 2020). To bring these groups together, a trusted framework (e.g., an accessible data dictionary readable by humans and machines), reliable data sharing between groups (e.g., a consistent taxonomy), and an understanding of and compliance with data-sharing policies (e.g., data transparency vs. data privacy) must be in place. By implementing these steps, an effective data management culture can be established (Benfeldt et al., 2020).

2.2.2 Integration

Data collection, definitions, uses, stewards, and individual stakeholders must work in harmony to enable consistent, usable, and trustworthy data integration across the enterprise (Brous et al., 2016; Janssen et al., 2020). Integration enables effective decision-making support and prevents fragmentation and silos. Fragmentation refers to data or data processes that lack information about origin, use, stewardship, or other relevant details. Silos describe data or data processes that should be accessible to members of an organization but are isolated to one place or person due to ineffective or nonexistent communication.

Data integration extends beyond the technical realm in the context of EDM, even though the technical aspect is equally important. The socio-technical aspects of data integration must be viewed through the lens of the respective organization. Chief officers of data, information,

technology, or engineering (or their equivalents) hold the final say on all data governance decisions, thereby establishing decision rights and accountability within an organization (Al-Badi et al., 2018; Office of Management and Budget, 2020). By establishing these responsibilities, technical integration becomes a more streamlined process. Once roles are defined, metadata and standards can be formalized and implemented within an organization's data architecture. By doing so, schemas can be integrated across systems and users more efficiently (Bakry et al., 2016).

In addition to silos and fragmentation, common data integration challenges include inconsistent formats and standards, lack of stewardship, data quality issues, and difficulties integrating across legacy systems (Al-Badi et al., 2018; Bakry et al., 2016; Janssen et al., 2020). Because data is often collected before policies are implemented, it is frequently stored in different formats and across various repositories, making lineage difficult to track. When this sequence occurs, it is often a significant hindrance to implementing EDM. Because of the investment required to implement such changes, some organizations choose to maintain the status quo in data practices. As these practices continue, they can contribute to the existence of a *data swamp* (Nambiar & Mundra, 2022). A *data swamp* is an environment where data is lost, misused, or otherwise poorly managed, leading to difficulties implementing machine learning or artificial intelligence methods in the future (Janssen et al., 2020).

Legacy systems that lack interoperability and are difficult to update are often used because of their relevance and the user base's existing understanding. While this is a hindrance to organization-wide data integration, these systems remain necessary and convenient. The complete overhaul of these systems would be time-consuming and expensive. However, the lack of interoperability reduces the usefulness of data for organizational decision-making, which requires data integration across systems (Boadi et al., 2026). Continued use of legacy systems suggests the need to implement an enterprise-level "data warehouse" management strategy. This approach improves data access but does not immediately solve data integration problems without support from ontologies, taxonomies, metadata catalogs, and other relevant data management tools.

Establishing warehouse architecture and schemas for the data in question lays the groundwork for data standards and useful metadata definitions. Using the established framework to build a common understanding of data is essential for data governance and integration, which can be done with a formalized, accessible data dictionary (Brous et al., 2016). By standardizing data definitions, metadata structures, and naming conventions, organizations create a common reference point that enables data from disparate systems to be mapped, exchanged, and interpreted consistently. Ensuring that the developed framework captures all necessary data and metadata ensures that data accuracy, consistency, timeliness, and completeness are sufficient across an enterprise. Integrating data across organizations without quality control can spread poor-quality information, further confusing groups (Janssen et al., 2020).

Establishing the foundation for enterprise-level integrated data enables asset managers to access trustworthy, understandable data for decision-making. Integrated data enables better asset condition assessment, lifecycle planning, and risk-based decision making. Integration of multiple asset datasets extends these benefits, enabling the establishment of more robust decision-support systems (Bakry et al., 2016; Brous et al., 2016).

2.2.3 Storage

Data storage methods are plentiful, but only some are suited to enterprise-level data management. When asset data is managed in consumer-grade spreadsheet or document applications, it becomes very easy for data to become siloed on an employee's desktop or fragmented through undocumented processes or changes to the data (Benfeldt et al., 2020; Shrestha & Sheikh, 2022). While these methods may have their place within an enterprise, enterprise data storage systems (EDDS) present a necessary solution for big data management. Support for asset management, analytics, and decision-making relies on a cross-departmental approach to data storage. There is an abundance of EDSS solutions available across providers. Google, Amazon, Oracle, and Dell are just a few of the brands in the current market. As UDOT begins migrating to Google solutions, all research conducted will use the terminology used across Google. Tools, plug-ins, and platforms referenced may be platform-specific; however, in the event of future platform migration, the principles and methods discussed should remain relevant (Shrestha & Sheikh, 2022).

When storing data, democratization should be a focus. This practice has been somewhat sparsely conceptualized in research due to its recency. However, it is defined as "*the process of empowering a group of users spanning beyond established data experts to access and use data, by removing obstacles to data exploration and sharing in enterprises*." (Labadie, Eurich, et al., 2020). This definition represents the implementation of FAIR (Findable, Accessible, Interoperable, and Reusable) data principles in an enterprise context ("FAIR Principles," 2016; Labadie, Legner, et al., 2020). Democratization and FAIR data principles necessitate the use of a central data repository (or repositories) with relevant guiding documentation readily available to any user in need. Further discussion of data storage architecture, management strategies, ontologies, taxonomies, and other relevant information is discussed in sections 2.4 and 2.5.

The use of EDSS also addresses several important integration challenges. By moving data from individually chosen storage methods to the enterprise level, standards can be applied to enable interoperability between datasets. This storage method further breaks down silos and mends fragmentation. A centralized data repository enables integrated decision-making across asset types, groups, and departments (Rashid et al., 2020).

2.2.4 Access

Data access within an enterprise involves determining who should have access to specific data, what level of access they should receive, and how that access should be managed. Different groups, including planners, engineers, maintenance teams, leadership, consultants, contractors, and other external stakeholders, require access to different types of information based on their roles and responsibilities. Establishing appropriate access levels is therefore a critical component of data governance (Bernardo et al., 2024). Transportation agencies rely on access to a wide range of information, including asset inventories, maintenance records, GIS data, condition ratings, financial data, and project information (France-Mensah & O'Brien, 2019). Timely and reliable access to these datasets is essential for supporting safety initiatives, budgeting decisions, project prioritization, risk management, and long-term planning. However, while data should be accessible to those who need it, access controls must also ensure that sensitive or restricted information is protected through clearly defined permissions and role-based access policies (Brous et al., 2016).

Role-based access and permissions tie organizational responsibilities together. Assigning different levels of access and permissions based on job function, authority, and needs lays the foundation for a strong EDM implementation. Creating decision rights and accountabilities for data use through clear organizational structures for data stewards, data owners, system administrators, and end users ensures data quality and improves coordination between data architects and operational staff (Brous et al., 2016; Rashid et al., 2020).

Access also refers to the interpretability and traceability of data. Data interpretability can take the form of definitions, use cases, and related data. Metadata, or descriptive information about data, including definitions, field descriptions, lineage, quality indicators, and business context, is a critical aspect of improving data accessibility (Rashid et al., 2020). Well-developed metadata enables users to better understand where the data originated, how it has been modified over time, who is responsible for maintaining it, and any limitations that may exist. This improves discoverability by helping users identify relevant datasets more quickly and determine whether the information is appropriate for their intended use.

Semantic data dictionaries, controlled vocabularies, and ontology-based relationships can further improve accessibility by connecting similar terms, standardizing definitions, and identifying relationships between datasets that may otherwise appear unrelated. These tools support greater interoperability across systems and departments by ensuring that users interpret data consistently, regardless of its source. Improved metadata and discoverability practices, therefore, make enterprise data more accessible, understandable, reusable, and trustworthy (Labadie, Legner, et al., 2020).

2.2.5 Quality

Data quality refers to the degree to which data is usable. Some characteristics of data quality include completeness, relevance, reliability, timeliness, and value added (Wang et al., 2024). Additional dimensions of data quality often include accuracy, consistency, validity, uniqueness, and accessibility. Accuracy refers to whether the data accurately reflects real-world conditions, while consistency refers to whether similar information is consistently represented across multiple systems or departments. Timeliness is particularly important in transportation agencies, where outdated information may no longer reflect current asset conditions,

maintenance activities, or project statuses. Ensuring that these characteristics are present helps organizations avoid the consequences of poor-quality data. Pitfalls of poor-quality data can manifest as duplicated work, inefficiency, poor decisions, and a possible loss of user confidence in the data. For transportation agencies, poor-quality data may also result in inaccurate condition ratings, missing maintenance history, conflicting asset inventories, inconsistent GIS locations, or incomplete financial records. These issues can reduce the effectiveness of safety analysis, budgeting, project prioritization, and long-term planning efforts (Katsumi et al., 2022).

Data quality must also be maintained throughout the entire data lifecycle, including collection, storage, integration, analysis, sharing, and archiving. Errors introduced during manual entry, file transfers, system integration, or inconsistent naming conventions can degrade data usability over time. As a result, organizations should implement validation rules, standardized templates, metadata requirements, and regular quality checks to ensure that data remains accurate and fit for use (Lei et al., 2022; Wu et al., 2025). Establishing clear data ownership and stewardship responsibilities is also important to ensure that quality issues can be identified, corrected, and consistently monitored across the organization.

2.3 Ontology

Ontology is a method of defining a common vocabulary for research or enterprise use to share information in a domain (Noy & McGuinness, 2001). The purposes of ontology development include creating a common understanding, enabling the reuse of domain knowledge, asserting domain assumptions, and analyzing domain knowledge. The primary benefit of an ontology is the creation of a formalized understanding of domain knowledge that is semantically interpretable by humans and machines (Firdaus Law et al., 2019; Lei et al., 2022). At its core, an ontology helps an organization give meaning to data, enabling improved decision-making (Boadi et al., 2026). In this section, the core components of ontology, the benefits of ontology in the context of EDM, its applications in transportation and infrastructure, and ontology development methodologies are discussed. A simplified family ontology is visualized in Figure 2.1.

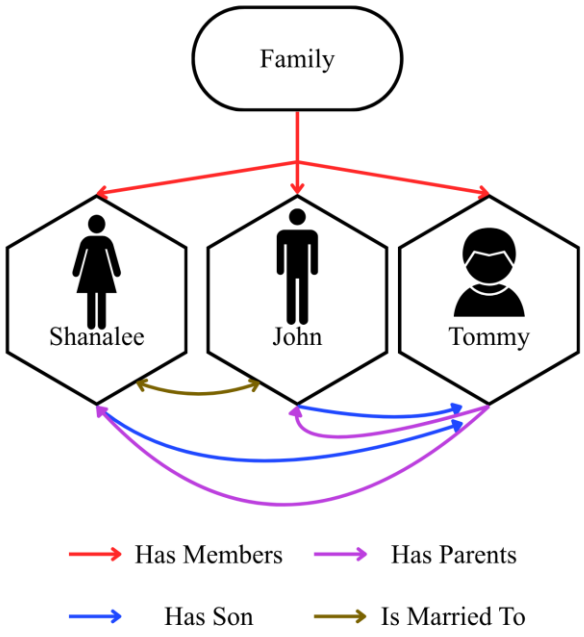


Figure 2.1 A visualization of a simple family ontology.

2.3.1 Ontology Components

Ontologies, regardless of domain or application, are typically developed through a set of generalized steps. These include defining classes, arranging the classes in a class-subclass hierarchy, specifying properties (or slots) and their allowable values, and assigning values to instances within the ontology (Noy & McGuinness, 2001). Defining classes within an ontology formalizes domain concepts. Within classes, there may be subclasses that can further narrow a description. For example, an asset class may exist, and within that class, subclasses of barriers, signs, and bridges would exist.

Instances within an ontology are not always restricted to a single class or subclass; rather, they may belong to multiple classes when appropriate. For example, a material such as gravel may be associated with both a *concrete mix* class and a *road base* class, provided its defining properties are consistent across contexts. This flexibility allows ontologies to more accurately reflect real-world relationships, reuse concepts across domains, or differentiate between similar concepts.

The process of defining and structuring classes enables ontology developers and subject matter experts to distinguish between entities that may share similar names but differ in meaning, and those that have similar meanings but are labeled differently. By arranging classes into a hierarchical structure, ontologies establish relationships, such as "is-a" relationships, that enhance interpretability and semantic clarity. For example, defining that a *cast-in-place barrier* "is a" type of *barrier* demonstrates how hierarchical organization supports increasingly granular levels of information (Öhgren, 2004).

Properties are another critical piece of an ontology. While a class may define the "which" of a subject, the properties set forth the "what about." Defining the properties associated with an object provides valuable information to a user. Not only do these properties allow the ontology developer to set constraints that allow for the automation of quality control mechanisms, but they also allow the developer to define what a property means, the relevant units, or use cases for a property (Du et al., 2023; Le et al., 2018; Lei et al., 2022; Noy & McGuinness, 2001). In the domain of traffic asset management, assets often have similar or identical properties, which may mitigate some of the quality-control benefits of an ontology. This necessitates the implementation of a field or fields that can differentiate assets from one another (Wunder et al., 2011).

Within an ontology, relationships and axioms are critical components that go beyond simple classification, enabling the representation of meaningful connections and domain-specific rules. Relationships, often defined as properties, describe how classes and instances interact, allowing the ontology to capture real-world dependencies between objects. These relationships may link entities to other entities (object properties), such as an asset being located on a roadway segment, or connect entities to data values (data properties), such as an asset having a specific material type or condition rating. By formally defining these connections, ontologies move from static representations of concepts to dynamic models of how systems function (Katsumi et al., 2022; Lei et al., 2022; Noy & McGuinness, 2001).

Axioms further enhance this structure by introducing logical constraints and rules that govern how classes and relationships behave. These may include restrictions on allowable values, definitions of disjoint classes, or requirements that certain relationships must exist.

Through these constraints, axioms ensure that the ontology remains internally consistent and semantically meaningful, while also enabling automated reasoning and validation (Firdaus Law et al., 2019; Wu et al., 2025; Yan et al., 2023).

Using domain knowledge to establish relationships and axioms adds significant depth to an ontology, particularly in complex systems such as transportation infrastructure, where assets, processes, and data are highly interconnected. The benefits of incorporating these elements include reducing ambiguity in data interpretation, ensuring data quality and consistency, enabling advanced decision-support systems through inference, and supporting interoperability across disparate systems. Together, relationships and axioms transform an ontology into a robust knowledge framework capable of supporting enterprise data integration and data-driven decision-making (Noy & McGuinness, 2001).

2.3.2 Ontology Development Methodologies

A variety of approaches to ontology development exists, ranging from highly prescriptive processes to general guidelines. Whether prescriptive or general, all ontologies set out to define a scope, identify concepts and relationships, develop a taxonomy, and implement and validate the ontology. These methodologies provide a structured or semi-structured process for building consistent, reliable, and reusable ontologies. These methodologies aim to ensure scalability, interoperability, and clarity across datasets. Ontology development is inherently iterative, collaborative, and domain-driven. Several of the foundational ontology methodologies applied in this study are summarized in Table 2.1.

One of the most widely adopted approaches to ontology development is the methodology proposed by Noy and McGuinness (2001). This approach provides a practical, step-by-step framework that emphasizes clarity and flexibility, making it particularly suitable for domain-specific applications. The methodology begins by defining the domain and scope of the ontology, then identifying key concepts and organizing them into a hierarchical class structure. Subsequent steps involve defining properties (or slots) and their allowable values, establishing relationships among concepts, and populating the ontology with instances. A key strength of this methodology is its accessibility and adaptability, allowing developers to refine the ontology as domain understanding evolves iteratively. The structured yet flexible nature has made it a

Table 2.1 A Summary of the Ontology Development Methodologies Employed in This Study

Methodology	Focus	Strengths	Limitations	Contribution
Uschold & King	Structured ontology engineering	Simple and Foundational	Less implementation detail	Provides strong conceptual alignment
Noy & McGuinness	Iterative ontology creation	Accessible and practical	Less focused on EDM principles relevant to UDOT applications	Strong foundation for a proof-of-concept ontology
Fernandez et al.	Formal development cycle	Comprehensive documentation	Highly complex	Helpful to consider as an ontology scales

foundational reference in ontology engineering, particularly in projects where domain knowledge must be progressively formalized and validated (Noy & McGuinness, 2001; Stadnicki et al., 2020).

METHONTOLOGY (M. Fernandez et al., 1997)is a comprehensive, lifecycle-based ontology development methodology that emphasizes rigorous documentation, validation, and iterative refinement throughout the development process. Originally developed to support knowledge-based systems, this methodology structures ontology development into several key phases, including specification, conceptualization, formalization, implementation, and maintenance. During the specification phase, the ontology's purpose and scope are clearly defined. In contrast, the conceptualization phase focuses on organizing domain knowledge into structured representations, such as concept hierarchies and relationship models. Formalization translates these conceptual models into formal representations, which are then implemented using ontology languages such as OWL. A distinguishing feature of METHONTOLOGY is its strong emphasis on evaluation and documentation at each stage, ensuring both consistency and traceability of decisions. This methodological rigor makes it particularly suitable for complex,

large-scale domains such as infrastructure systems, where maintaining clarity and consistency across evolving datasets is critical (M. Fernandez et al., 1997; Firdaus Law et al., 2019; Öhgren, 2004; Stadnicki et al., 2020).

The methodology presented by Uschold and King (1995) is among the earliest structured approaches to ontology development, with a strong focus on enterprise integration and organizational use. This methodology begins by identifying the ontology's purpose and intended users, followed by ontology construction, which includes knowledge capture, coding, and integration with existing systems. Knowledge capture involves eliciting domain expertise and identifying key concepts and relationships, while coding translates this knowledge into a formal representation. The methodology also includes evaluation and documentation stages to ensure that the ontology meets its intended objectives and can be effectively reused. A key strength of this approach lies in its alignment with organizational and enterprise needs, making it particularly relevant for applications such as transportation agencies, where ontologies must integrate with existing workflows, databases, and decision-making processes (Firdaus Law et al., 2019; Katsumi et al., 2022; Stadnicki et al., 2020; Uschold & King, 1995).

Ontology development represents domain knowledge through classes, properties, and relationships that define concepts and their interactions. These components create a semantic framework that enables consistent interpretation, integration, and interoperability of data across systems, making ontologies particularly valuable in complex domains such as transportation and infrastructure (Rashid et al., 2020; Stadnicki et al., 2020). Even though the discussed components provide the foundation for ontology, the development process is inherently iterative and collaborative. Effective ontology development depends heavily on domain experts who provide essential knowledge of relevant concepts, classifications, and relationships. Initial ontology structures are rarely complete or fully accurate (Lei, 2023). They require refinement, which progressively improves through evaluation, feedback, and revision. This ongoing process allows developers to resolve inconsistencies, address knowledge gaps, and adapt the ontology to changing domain requirements.

Ontology developers and subject matter experts continuously work together to ensure technical rigor and practical relevance are maintained throughout the ontology's lifecycle. In

domains with complex interdependencies, such as infrastructure systems, this ongoing collaboration is essential for accurately capturing real-world relationships. The integration of structured development practices with continuous expert input results in ontologies that are robust, flexible, and capable of supporting advanced, data-hungry applications (Farghaly et al., 2023; Wu et al., 2025; Yan et al., 2023).

2.3.3 Ontology in the Context of EDM

Within an enterprise context, ontologies enhance enterprise data management by providing a semantic framework that enables interoperability, data integration, discoverability, governance, and decision-making. By assigning meaning to otherwise isolated data elements, ontologies transform stored data into structured knowledge and establish a shared vocabulary that can be understood consistently across departments, systems, and stakeholders. This semantic layer addresses common organizational challenges such as siloed data, inconsistent terminology, and fragmented information systems, allowing organizations to derive greater value from their data assets and support more informed decision-making (Firdaus Law et al., 2019).

A key distinction exists between syntactic and semantic interoperability. Syntactic interoperability enables systems to exchange data using consistent formats and structures, whereas semantic interoperability ensures that the meaning and context of data are interpreted consistently by both humans and machines (Nandini & Shahi, 2018; Wang et al., 2024). Ontologies facilitate semantic interoperability by establishing standardized definitions and relationships among concepts, reducing ambiguity and improving communication across organizational boundaries.

These capabilities are particularly valuable in environments characterized by heterogeneous systems and diverse data sources. As organizations adopt new technologies and maintain legacy systems, differences in schemas, formats, and terminology create barriers to integration. Ontologies function as a semantic integration layer, enabling data from structured, semi-structured, and unstructured sources to be mapped and connected through shared concepts. This approach improves data accessibility, reduces duplication, supports enterprise-wide governance, and provides a more unified view of organizational information (Bernardo et al., 2024; Rashid et al., 2020).

Ontologies also improve data discoverability by enriching metadata with semantic meaning. Unlike traditional keyword-based searches, ontology-driven systems enable users to retrieve information based on conceptual relationships and meaning. This supports more effective data cataloging, faster information retrieval, and improved usability for both technical and non-technical stakeholders. For example, a query such as "all assets with structural damage" can leverage semantic relationships across multiple systems, enabling more comprehensive and accurate analysis (Bakry et al., 2016; Esteghamati et al., 2024; Esteghamati & Flint, 2021; Farghaly et al., 2023; Katsumi et al., 2022; Zaker Esteghamati et al., 2025). These capabilities result in faster data retrieval, improved usability for non-technical stakeholders, and increased overall utilization of enterprise data.

2.3.4 Ontology Applications in Transportation and Infrastructure

Ontologies have seen increased adoption in the transportation and infrastructure domains to address challenges posed by large, heterogeneous, and semantically inconsistent datasets (Boadi et al., 2026; S. Fernandez et al., 2016; Le et al., 2018; Nandini & Shahi, 2018; Wu et al., 2025; Yan et al., 2023). These applications focus on improving data integration, enabling advanced analytics, enhancing decision-making processes, and facilitating knowledge sharing across organizations.

Ontologies in the domain of transportation infrastructure provide a formalized framework for representing knowledge through defined entities, relationships, and attributes, enabling a shared semantic understanding across users and systems (Nandini & Shahi, 2018). Unstructured, semi-structured, or structured data with inconsistent naming conventions can pose challenges to extracting meaningful insights. These challenges are exacerbated when semantic inconsistencies across organizations are not inherently identifiable or understandable. Within transportation and infrastructure applications, such methodologies support the unification of fragmented data systems and provide a foundation for improved decision-making, asset management, and system-level analysis (Lei et al., 2022). A high-level view of an ontology and its respective concepts is shown in Figure 2.2.

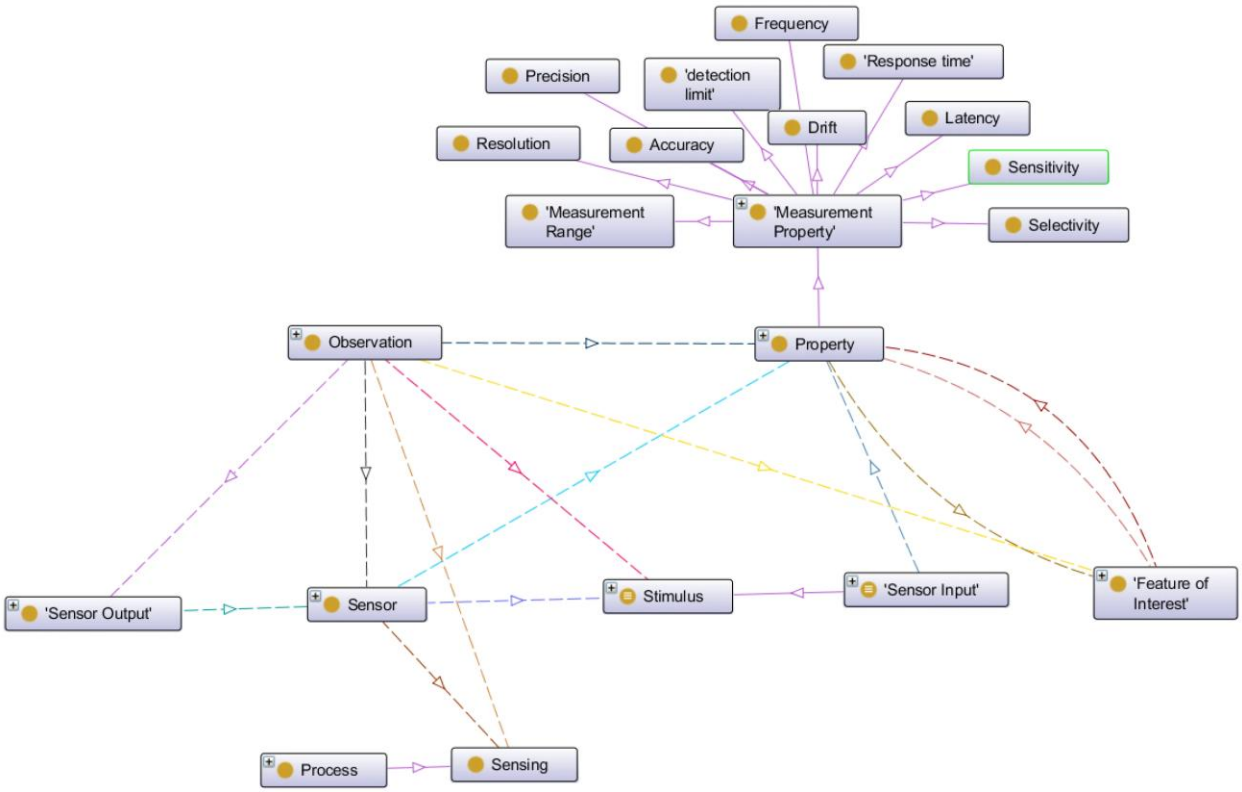


Figure 2.2 An example ontology demonstrating relationships between concepts in an intelligent transportation system (S. Fernandez et al., 2016).

A primary application of ontologies in infrastructure systems is their role in asset management and data integration. Infrastructure agencies often rely on a mix of legacy and cutting-edge systems that store asset information across disparate databases, resulting in fragmented, difficult-to-access data environments. Asset management ontology research highlights that asset management activities, such as condition monitoring, maintenance planning, and lifecycle cost estimation, are inherently data-intensive and can be significantly hindered by these siloed systems (Adams & Esteghamati, 2026; Katsumi et al., 2022; Lei et al., 2022; Pokhrel et al., 2025, 2026). Data is also collected at various stages of the asset lifecycle, as shown in Figure 2.3, creating distinct data types associated with individual assets. Ontologies provide a semantic layer that enables the integration of these systems into a unified data hub, allowing for standardized data access and improved interoperability (Farghaly et al., 2023). By defining shared vocabularies and relationships among assets, roles, and processes, ontology-backed systems enable users to query data across multiple sources through a single, coherent interface or to navigate disparate systems independently. This integration reduces reliance on

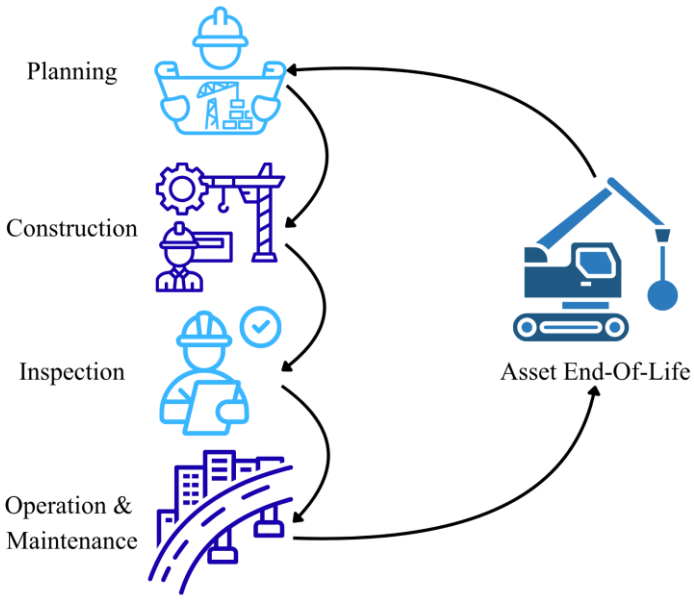


Figure 2.3 The asset lifecycle.

technical specialists for data retrieval and supports more comprehensive tracking of asset histories and lifecycle performance.

Beyond integration, ontologies play a critical role in supporting data analytics and visualization in transportation systems. Transportation asset datasets are often large, complex, and hierarchical, containing millions of records and numerous interrelated attributes. Traditional visualization techniques struggle to effectively represent these hierarchical relationships. Ontology visualization research shows that ontology-based approaches can guide the automated selection of relevant data attributes and enable the visualization of hierarchical asset data through techniques such as heat trees (Le et al., 2018). In this context, the ontology encodes domain knowledge and semantic relationships, allowing systems to identify which variables are meaningful for analysis and visualization. The resulting simplification reduces the cognitive burden on analysts and enables more efficient exploration of patterns, trends, and correlations within large datasets. Ultimately, ontology-driven visualization enhances practitioners' ability to interpret complex data and make efficient, informed decisions. Another significant application of ontologies lies in decision support and knowledge management for infrastructure operations, particularly in areas such as bridge maintenance (Bakhtiari et al., 2026; Ren et al., 2019; Xu et al., 2019). Infrastructure maintenance processes involve complex decision-making that requires

integrating knowledge from multiple domains, including structural, environmental, cost, and operational constraints. Research on bridge maintenance demonstrates that ontology-based knowledge systems can formalize these diverse knowledge sources into a unified framework, enabling automated reasoning and rule-based decision support (Ren et al., 2019). By incorporating semantic relationships and rule languages, these systems can perform tasks such as damage evaluation, maintenance planning, and supplier selection while ensuring consistency with established standards and regulations. Additionally, ontologies facilitate improved communication and knowledge sharing among stakeholders by providing a common conceptual framework, reducing ambiguity, and supporting cross-disciplinary collaboration.

Infrastructure ontologies provide a structured framework for representing the complex interdependencies that exist among urban assets, enabling a more integrated approach to infrastructure management. Cities function as interconnected "systems of systems," with components such as roads, buried utilities, and the ground that are not isolated but constantly influence one another through shared physical, environmental, and operational processes. Traditional management approaches, which treat these systems independently, often fail to account for cascading effects, where a failure in one asset can trigger a chain of impacts across multiple systems. To address this limitation, ontology-based models that capture not only the classifications of infrastructure assets, but also their properties, processes, and interrelationships have been developed (Du et al., 2023b). These models incorporate key domains, including roads, water pipes, ground conditions, human activities, and environmental phenomena, allowing both direct and indirect interactions to be formally represented and machine interpretable. For example, environmental factors such as rainfall can influence soil properties, which, in turn, affect pipe corrosion and road deformation, illustrating the layered and interconnected nature of infrastructure deterioration processes. By encoding these relationships within an ontology, the framework enables improved data integration, supports automated reasoning, and enhances the ability to anticipate system-wide impacts, ultimately providing a foundation for more informed and proactive infrastructure decision-making.

Another application of ontology in transportation infrastructure has been to support integrated highway planning by addressing fragmented data systems and siloed decision-making within state highway agencies. Multiple groups are usually involved in traditional planning

practices (e.g., maintenance, safety, planning). These groups often operate independently, using incompatible systems that lead to redundant efforts and scheduling conflicts. Developed ontologies aim to provide a shared, formalized representation of asset information, maintenance, and rehabilitation, as well as inter-project coordination, to enable cross-functional integration (France-Mensah & O'Brien, 2019). By structuring relationships among pavement assets, funding constraints, project schedules, and coordination processes, the ontology enables the identification and resolution of conflicts between projects that overlap in schedule. The ontology also supports advanced querying through semantic web technologies, enabling users to retrieve information across otherwise siloed datasets. Ontology-driven approaches can achieve high levels of precision and recall in representing domain knowledge, illustrating the effectiveness of ontology in supporting data-driven, proactive, and system-level planning in infrastructure systems. (France-Mensah & O'Brien, 2019)

Collectively, these applications illustrate that ontologies function as a foundational technology for addressing key challenges in transportation and infrastructure systems. They enable a transition from semantic discord to semantically consistent systems that support advanced analytics and decision-making. As infrastructure systems continue to grow in scale and complexity, the role of ontologies is expected to expand further, particularly in supporting digital twins, predictive modeling, and intelligent asset management systems (Wu et al., 2025; Yan et al., 2023). Table 2.2 summarizes the applied ontologies referenced in this study.

Table 2.2 Summary of applied ontologies, including Author, Year, Domain, Development Method, and Key Takeaway

Author(s)	Year	Domain	Development Method	Key Takeaway
France-Mensah & O'Brien	2019	Highway Planning and Transportation Asset Management	Expert interviews, literature review, ontology reuse, competency questions, and ontology validation through SPARQL queries	Developed an ontology to improve interoperability and integrated highway planning across transportation functions.
Le et al.	2018	Transportation Asset Data Visualization	Domain ontology development supported by taxonomies and semantic modeling of asset data	Used an ontology to organize and visualize large transportation asset datasets.
Nandini & Shahi	2019	Transportation Systems	DERA methodology (Domain–Entity–Relation–Attribute), competency questions, upper ontology alignment (DOLCE), and SPARQL validation	Created a transportation ontology to support semantic querying and information retrieval.
Ren, Ding, & Li	2019	Bridge Maintenance	Knowledge acquisition from literature and experts, OWL ontology development, SWRL rule implementation, and multi-method validation	Developed a bridge maintenance ontology to support knowledge sharing and decision-making.
Xu et al.	2019	Highway Construction Inspection	Knowledge extraction from specifications and agency documents, ontology modeling, OWL implementation, SPARQL querying, and SWRL reasoning	Applied ontology-based knowledge management to automate and improve highway construction inspection.

2.4 Data Lake Architecture

Data lakes are a method of big data management that has emerged recently in the 21st century (Russom, 2020). The motivation for developing data lakes stems from the rapid expansion of data generated by sensors, enterprise systems, digital platforms, and increasingly data-centric organizational operations. Volume, velocity, variety, veracity, and value (commonly referred to as "the 5 Vs") are the characteristics of enterprise data that necessitate the development of a data lake (Harby & Zulkernine, 2025). Traditional data storage methods, such as data warehouses, present challenges for management and development in the current data management environment, which requires an agile approach.

In contrast, data lakes provide a centralized repository that stores data in its raw, native format, supporting structured, semi-structured, and unstructured data on a single platform. This approach is enabled by a shift from schema-on-write to schema-on-read, where data structure is applied only when it is accessed rather than when it is ingested. A comparison of schema-on-read and schema-on-write approaches is shown in Figure 2.4. As a result, data lakes adopt an extract-load-transform (ELT) paradigm, allowing for greater flexibility in data processing and analysis.

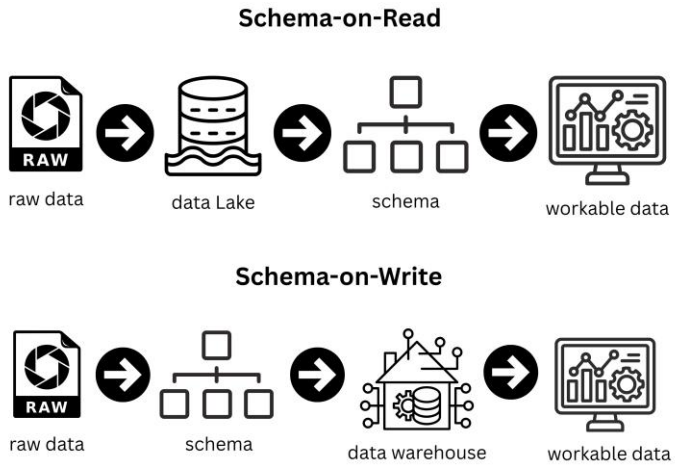


Figure 2.4 Comparison of schema-on-read processing in data lakes and schema-on-write processing in data warehouses.

This capability is particularly important for supporting advanced analytics, machine learning, and real-time decision-making applications. However, the flexibility of data lakes also introduces challenges related to data governance, quality, and organization, as poorly managed systems risk becoming "data swamps" that hinder rather than support analysis (Nambiar & Mundra, 2022; Nargesian et al., 2019). To effectively realize their potential, data lakes are implemented using layered architectures that manage data ingestion, storage, processing, and access, as discussed in the following sections.

2.4.1 Core Components

Several components have traditionally been a part of data lakes regardless of architecture. Every data lake platform uses slightly different semantics to describe a component, but the core purpose remains the same. The components include ingestion, storage, cataloging, processing, and access layers. While sometimes these layers are combined into a single layer, they will be discussed individually here as follows:

- The *data ingestion* layer serves as the landing spot for all data entering the lake. This layer supports batch, real-time, and streaming ingestion, enabling large volumes of data to be ingested simultaneously. Sensors, asset databases, GIS systems, and other collected data are captured in this layer with minimal processing. Depending on the platform used, tools and technologies can automate certain steps in the ingestion process. These tools are needs-based and are often determined by the enterprise utilizing the system (Harby & Zulkernine, 2025).

- The *data storage* layer is where ingested data is stored until it is ready to undergo further processing or cataloging. This layer is intended to be a scalable repository for structured, semi-structured, and unstructured data. Storage methods are platform-dependent, ranging from cloud-managed solutions to on-premises servers or drives. Raw data is preserved in this layer to preserve flexibility and reusability (Freitas et al., 2026; Harby & Zulkernine, 2025).

- The *cataloging layer* is the most relevant for data governance applications. This layer utilizes tags, data dictionaries, or ontology to improve discoverability and

governance. As data passes through this layer and is assigned the proper tags, it can be indexed, tracked, and semantically enriched. The cataloging layer is also the most critical layer for processing all data. Effective lake architecture requires strong data governance mechanisms to ensure quality, security, and usability. Governance functions include access control, data validation, metadata management, and lifecycle tracking, all of which help maintain the system's integrity. Without proper governance, data lakes risk becoming disorganized and difficult to navigate, often referred to as “data swamps.” As a result, governance is increasingly integrated into cataloging and processing layers to support consistent and reliable data management across the system (Freitas et al., 2026; Nargesian et al., 2019; Sawadogo & Darmont, 2021).

- The *processing layer* is responsible for transforming raw data into usable, structured formats that support analysis and decision-making. Unlike traditional systems, which rely on extract-transform-load (ETL) processes, data lakes use extract-load-transform (ELT), storing data in its raw form and transforming it as needed. This layer supports batch and real-time processing, enabling data cleaning, aggregation, enrichment, and integration across multiple sources. Processing frameworks enable organizations to derive meaning from large, complex datasets while maintaining flexibility in how they use data. This layer plays a crucial part in bridging the gap between raw data storage and actionable insights (Freitas et al., 2026; Hai et al., 2023; Labadie, Legner, et al., 2020; Nambiar & Mundra, 2022).

- The *data access* layer enables users to retrieve, query, and analyze data from the data lake. This layer supports a variety of access methods, such as structured query interfaces, application programming interfaces, and integration with business intelligence and visualization tools. In this layer, analysts, engineers, and decision-makers can run queries, generate reports, and develop insights from data. The flexibility of the access layer allows multiple users to interact with the same data in ways that suit their unique needs (Bernardo et al., 2024; Freitas et al., 2026; Sawadogo & Darmont, 2021).

Together, these components provide the foundation for a layered architecture in which data flows efficiently from ingestion to access. The scalability, flexibility, and navigability of such systems require implementing all the discussed layers in one form or another.

2.4.2 Zone Architecture

Several data management methods are discussed in the literature; however, UDOT employs the medallion architecture. Therefore, solely the medallion architecture will be discussed in this section. The medallion architecture, shown in Figure 2.5, is a data-refinement model that organizes data into progressively more structured zones (bronze, silver, and gold). It is a widely adopted approach that improves data quality, increases accessibility, enables scalability, and opens the door for modular data pipelines (Mohna et al., 2022). This architecture represents a shift away from rigid data warehouse systems toward a flexible, cloud-based data ecosystem capable of supporting AI and large-scale analytics (Pamula, 2026). The medallion architecture represents a zone-based data lake design, enabling the refinement of data from raw ingestion to actionable insight.

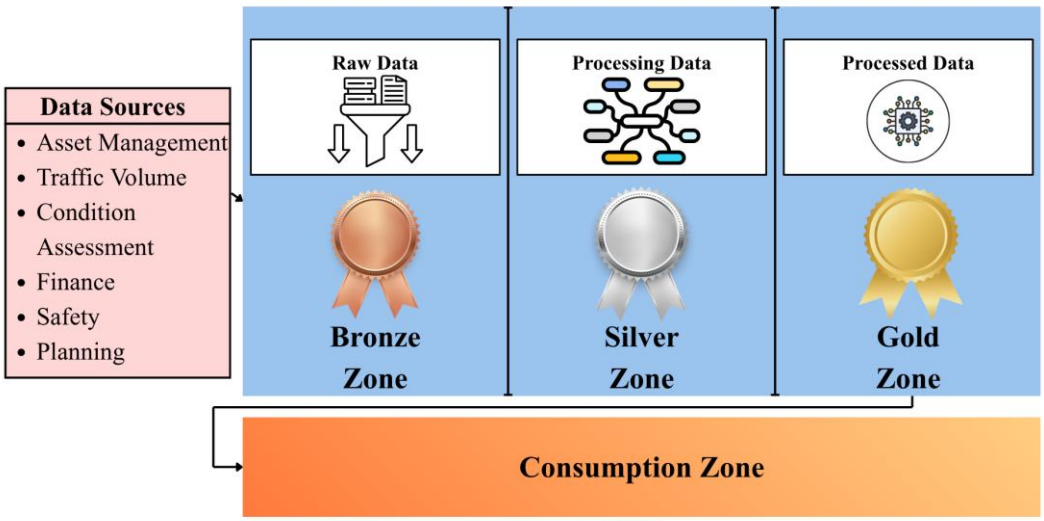


Figure 2.5 The medallion architecture utilized by UDOT.

Within the medallion architecture, data is gradually “matured” as it passes through each zone. The zones typically house raw, cleaned, and curated data, with each layer representing increases in data quality, structure, and usability. Each layer serves as a staging environment where users can work at different levels without corrupting downstream data. The implementation of these zones also aligns with the immutability, reproducibility, and auditability required in modern data systems (Mohna et al., 2022; Ravindran, 2025).

- The bronze layer serves as the initial landing spot for all incoming data. It stores raw, unprocessed data in its original format. This is often implemented through append-only logs or schema-on-read storage. It allows for high-throughput ingestion with minimal transformation. The bronze zone also enables the full history of the data to be traced, simplifying data auditing. The bronze layer also enables the schema to evolve as needed and creates a flexible ingestion pipeline (Mohna et al., 2022; Pamula, 2026). Reproducibility, data lineage, and data reprocessing are the bronze layer's most important contributions.

- The silver layer is where data is initially cleansed and processed. Schema enforcement, data cleansing, deduplication, and dataset integration are primary functions of the silver layer. These functions convert semi-structured data into reliable, queryable formats. This layer handles null fields, standardizes data, and adds semantic context. Data in this layer is structured and validated, but not yet fully optimized for use. The silver layer also prevents the propagation of poor-quality data and establishes a trusted intermediate layer (Gujjala, 2024; Mohna et al., 2022).

- The gold layer is where high-value, analytics-ready data resides. Here, data can be aggregated, used for feature engineering, or otherwise modeled. The data in this layer is optimized for business intelligence, machine learning, and real-time dashboard population. Performance techniques such as partitioning, materialized views, and query optimization can occur here. All data in this zone is highly structured and domain specific. Decision-making, predictive analytics, and AI

model inputs all pull data from this zone (Gujjala, 2024; Mohna et al., 2022; Sawadogo & Darmont, 2021).

Within the medallion architecture, data progresses sequentially through the bronze, silver, and gold layers. This progression represents increasing levels of refinement, structure, and usability. This staging approach enables controlled transformation, in which each layer applies defined processes to improve data quality and consistency. The bronze layer retains raw data in its native format, preserving data fidelity and enabling future reprocessing. In the silver layer, it undergoes cleansing, standardization, and validation to establish a reliable intermediate dataset. Finally, the gold layer provides curated, analytics-ready data. This structured flow ensures that data evolves in a controlled and reproducible manner, supporting scalability and consistency across large and complex datasets (Gujjala, 2024; Hai et al., 2023; Nambiar & Mundra, 2022; Ravindran, 2025).

This data flow is managed through automated pipeline frameworks that coordinate the movement and transformation of data between layers. These pipelines enable scheduling, dependency management, and incremental updates, ensuring that data remains current and that transformations are applied consistently across the system. By modularizing tasks, orchestration frameworks enhance the flexibility of the architecture, allowing individual components to be updated or scaled without disrupting the overall system (Mohna et al., 2022).

Governance mechanisms are deeply intertwined with these data flows. Metadata management plays a crucial role in this integration, capturing critical information such as origin, structure, transformations, and usage. This metadata enables comprehensive data lineage tracking, enabling data to be traced through a system or across multiple systems. Such traceability is paramount in auditing, debugging, and reproducibility of results. Additionally, embedded governance controls enforce data quality standards through validation checks, schema enforcement, and anomaly detection, particularly as data transitions from the Bronze to Silver and Gold layers. These mechanisms ensure that only high-quality, consistent data is propagated to downstream applications, reducing the risk of errors in analysis and decision-making.

Beyond quality control, governance frameworks also address security, compliance, and access management. Role-based access controls and policy enforcement mechanisms govern

who can view or modify data at each layer, ensuring sensitive information is protected while enabling appropriate access for analysis. Logging and monitoring systems track data access and transformations, supporting accountability and regulatory compliance requirements. Collectively, these governance practices prevent the degradation of the data lake into a disorganized and unusable "data swamp," instead maintaining a structured and reliable data environment (Nambiar & Mundra, 2022; Rashid et al., 2020).

The integration of data flow orchestration and governance within the medallion architecture is also important for enabling advanced analytics and artificial intelligence applications. By combining controlled data transformation with comprehensive metadata and governance frameworks, the medallion architecture provides a scalable foundation for modern data systems. However, while these mechanisms effectively manage the structure, quality, and accessibility of data, they do not inherently define the semantic relationships between data elements. This limitation underscores the need for ontology-based approaches that provide the semantic layer needed to interpret and relate data across systems and domains.

2.4.3 Capabilities and Advantages

The medallion architecture is a structured, scalable approach for managing large, heterogeneous datasets by organizing data into progressively refined layers. The balance of flexibility and structure in this approach allows raw data to be ingested without predefined constraints while enabling transformation into analytics-ready formats (Gujjala, 2024). This layered design supports structured, semi-structured, and unstructured data, providing an appropriate framework to support complex environments where data originates from a variety of sources (Lei, 2023).

A key capability of the medallion architecture is its scalability and flexibility in data processing. Decoupling storage from computation and adopting an ELT approach enables data to be processed on demand. This enables both batch and real-time data streams while allowing individual pipelines or layers to be updated without system disruption (Pamula, 2026; Shrestha & Sheikh, 2022). Data quality is also improved through staged refinement, in which data is progressively cleansed, validated, and transformed in each zone. This reduces errors and ensures that high-quality data is delivered downstream (Gujjala, 2024; Harby & Zulkernine, 2025).

The medallion architecture also supports strong data governance and traceability through integrated metadata management and lineage tracking. These features enhance transparency, reproducibility, and auditability across the data lifecycle (Gujjala, 2024; Labadie, Legner, et al., 2020). Interoperability is also supported through a unified framework for integrating data from disparate sources. Finally, the gold layer enables advanced analytics by delivering curated datasets prepared for querying, reporting, and machine learning (Pamula, 2026). Overall, the medallion architecture establishes a foundation for scalable, governed, and analytics-ready data.

2.4.4 Challenges and Limitations

Despite its advantages, the medallion architecture does present several challenges that must be addressed during implementation. One primary limitation is the increased complexity of the architecture. Managing multiple data layers and transformation pipelines requires significant data engineering knowledge. This can lead to higher development costs, particularly when transitioning from legacy platforms (Mohna et al., 2022; Ravindran, 2025).

Another limitation is data duplication and storage overhead. Data retained across multiple layers at different levels of refinement means the same dataset may appear in three layers. While this allows for excellent traceability and reprocessing potential, it increases storage requirements. The effectiveness of the architecture also depends heavily on strong data governance practices. Without robust metadata management, validation processes, and access controls, systems risk becoming disorganized “data swamps” that limit usability (Nambiar & Mundra, 2022; Ravindran, 2025).

Orchestrating the pipeline also presents challenges, as coordinating data movement across layers requires careful management of scheduling, dependencies, and failures as they arise. Delays in processing can introduce latency between data ingestion and availability for analysis, which may limit real-time applications. While the medallion architecture improves data structure and accessibility, it does not inherently capture semantic relationships between data elements. This limits its ability to support deeper data interpretation across systems without supporting infrastructure, such as data dictionaries and ontology (Gujjala, 2024; Hai et al., 2023; Nambiar & Mundra, 2022).

Integrating legacy systems can be difficult, especially in domains such as transportation, where data is often siloed, inconsistent, or of poor quality. Significant preprocessing may be required, and the cost associated with maintaining a multi-layered architecture can be substantial. These limitations highlight that while the medallion architecture is a powerful framework for data management, it does require careful design, governance, and complementary approaches to support it fully (Firdaus Law et al., 2019; Lei, 2023; Mohna et al., 2022).

2.4.5 Relevance to Transportation and Infrastructure Systems

Transportation agencies operate in increasingly data-intensive environments, where effective management of large, diverse datasets is critical for informed decision-making (Boadi et al., 2026). These datasets originate from a wide range of sources, including inventories, inspections, geospatial systems, and real-time sensor data. The continued evolution of transportation systems results in increased data volume and complexity. This evolution increases the demands placed on existing data management systems and practices. This trend aligns with broader observations in data-intensive environments, where increasing volume, velocity, and variety of data create challenges in storage, processing, and integration (Hai et al., 2023). Traditional approaches limit the integration of data across domains and hinder the development of comprehensive, system-level insights.

Transportation data is inherently complex and exhibits several characteristics that present challenges for legacy data management systems. It is heterogeneous, encompassing structured, semi-structured, and unstructured data formats. This data frequently originates from distributed and independent systems. These characteristics are consistent with those identified in modern data lake research, where heterogeneous data sources and formats necessitate flexible storage and schema-on-read approaches (Hai et al., 2023; Nargesian et al., 2019). Transportation data is also strongly spatiotemporal and lifecycle-oriented, requiring systems that can track asset performance and condition over time. As data volumes continue to grow, organizations are increasingly required to adopt new data management strategies to support real-time, large-scale analytics (Ravindran, 2025; Shrestha & Sheik.

Generally, data management systems within transportation agencies struggle to address these challenges because they rely on legacy infrastructure. Isolated databases, inconsistent data

formats, and a lack of standardized terminology across departments are typically characteristic of these systems. Legacy data management architectures are often rigid, costly to maintain, and poorly suited for integrating diverse, high-velocity data sources or supporting advanced analytics (Brous et al., 2016; Pamula, 2026). Because legacy platforms remain prominent, data integration requires significant manual effort.

Data lake architecture offers a promising solution to these challenges by enabling centralized, flexible data storage and processing. The data lake enables the ingestion and processing of raw data from heterogeneous sources in their original formats, supporting schema-on-read processing and reducing the need for extensive preprocessing (Hai et al., 2023; Ravindran, 2025). More advanced applications, such as lakehouse architecture, further enhance these capabilities by integrating features of both data lakes and data warehouses. These advanced methods enable transactional consistency and improve support for both batch and real-time analytics (Ravindran, 2025).

The data lake architecture enables effective support in data governance and transparency through metadata management, data lineage tracking, and fine-grained access control. These capabilities enable organizations to monitor how data is ingested, transformed, and used throughout its lifecycle, improving accountability and ensuring compliance with regulatory requirements (Pamula, 2026; Wu et al., 2025; Yan et al., 2023). Strong governance practices are essential in transportation systems, where decisions must be both reliable and defensible. The adoption of a medallion-based lake architecture further aids data management, improving integration across departments and systems. By providing a centralized platform for data storage and processing, fragmentation can be reduced, and cross-functional data sharing becomes a reality. This supports a more holistic approach to infrastructure management, allowing agencies to align data practices across multiple domains (Office of Management and Budget, 2020; Ravindran, 2025).

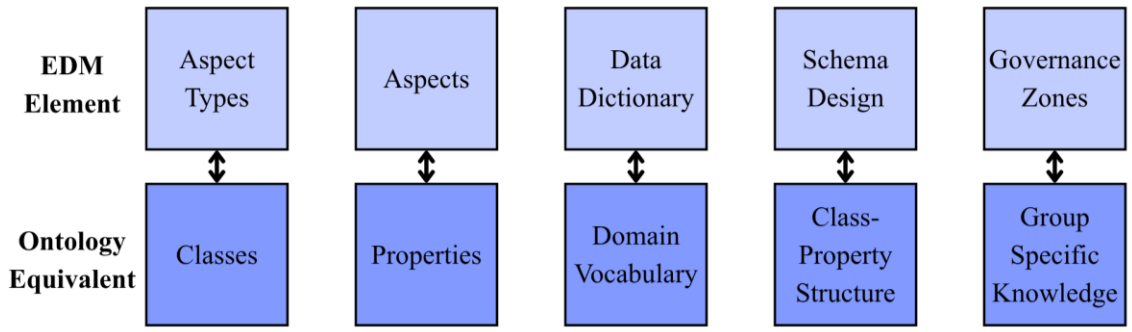


Figure 2.6 Alignment between enterprise data management elements and their ontology engineering equivalents.

2.5 Summary

Transportation agencies operate within increasingly complex, data-intensive environments where traditional data management approaches struggle to support integration, scalability, and advanced analytics. The heterogeneous and spatiotemporal nature of transportation data limits the ability to generate system-level insights and efficiently support decision-making. As discussed throughout this section, these challenges highlight the need for a modern data architecture that can manage large, diverse datasets in a unified, flexible manner. There is a natural fit between ontology development methodologies and EDM principles, with several core concepts in both domains possessing direct equivalents. Some of these equivalents are shown in Figure 2..

Data lake and lakehouse architecture, particularly when implemented using a medallion framework, provide a structured yet adaptable solution to these challenges. By organizing data into progressively refined layers, agencies can maintain the integrity of raw data while systematically improving data quality and usability. This layered approach supports the integration of diverse data sources and enables more consistent, reliable datasets for analysis, thereby addressing many of the limitations associated with legacy systems.

Implementing such an architecture also enhances core transportation functions, including infrastructure asset management, performance evaluation, and maintenance planning. By enabling lifecycle tracking and improving access to integrated datasets, agencies are better

positioned to perform comprehensive analyses and support data-driven decision-making. Additionally, strong governance mechanisms, including metadata management and data lineage tracking, improve transparency, accountability, and trust in data across the organization.

Finally, the centralized, modular nature of the medallion-based data lake architecture facilitates improved collaboration across departments and supports enterprise-level data management goals. By reducing fragmentation and enabling cross-functional data sharing, these systems lay the foundation for more coordinated, efficient infrastructure management. Together, these advancements position transportation agencies to improve their data for strategic planning, operational efficiency, and long-term system performance.

3.0 GAP ANALYSIS

3.1 Overview

This chapter discusses four components of the gap analysis: the current state of asset data management practices, the desired or ideal future state of asset data management practices, the gaps between the current and desired states, and the underlying causes of those gaps. Chapter 5 presents proposed solutions and implementation strategies to address the identified gaps. These components were identified and informed by the literature review, meetings with UDOT staff, and a survey of additional UDOT stakeholders. The full survey and its results can be found in Appendix A.

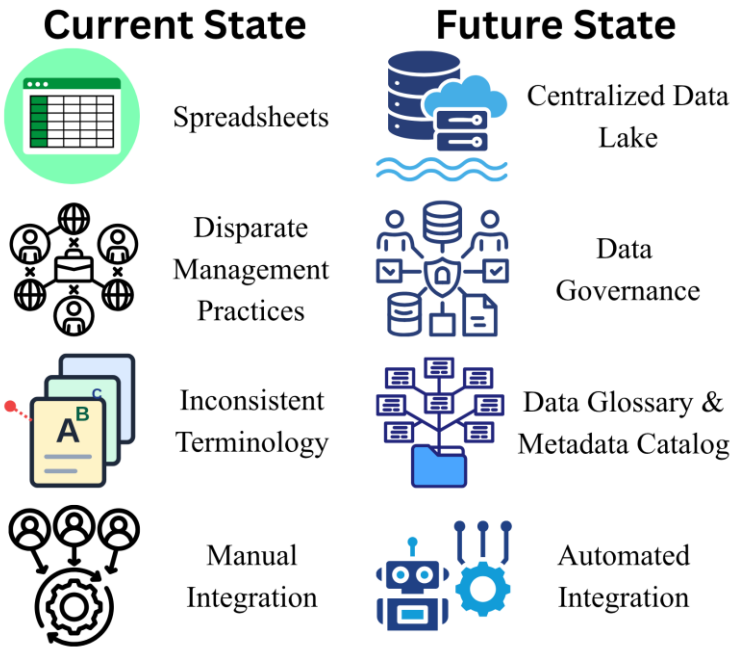


Figure 3.1 A summarization of the current and future desired state of data management at UDOT.

3.2 Current Asset Data Management Practices

Current asset data management practices for UDOT are guided by the UDOT Transportation Asset Management Plan (TAMP), a published document that informs organizational decision-making and asset management strategies (*Transportation Asset Management Plan (TAMP) - 2026*, 2025). The TAMP presents performance and risk-based strategies for the management of transportation system assets (*Transportation Asset Management Plan (TAMP) - 2026*, 2025). The objectives of the TAMP include developing a formal, data-driven approach to managing and funding UDOT transportation assets; integrating asset management into short- and long-term planning; incorporating risk into resource allocation decisions; and providing a dynamic asset management tool that supports connected, performance-based decision-making.

Regarding organizational data management goals, the TAMP provides an adequate framework for guiding decisions about what data should be collected, how it should be collected, and why it is needed. However, there is currently no formalized method for sharing this information or ensuring consistent use of the data across UDOT groups. To better understand group-level data management practices, this project included a survey of UDOT staff who work directly with, or adjacent to, selected asset data. The questions discussed in this section are categorized under survey demographics and data inventory, data governance, and stakeholder sentiment and challenges. The survey results are visualized in their entirety in Appendix A.

3.2.1 Survey Respondent Profile and Data Inventory

The survey questions in this section focus on respondent demographics and roles, current data management practices, data workflows, and the methods used for storing and maintaining collected data. Questions 1, 2, 3, 5, 6, 7, 12, 13, 18, and 21 from Appendix A provide a representative view of data usage and user information that lays the foundation for further discussion in this chapter.

Questions 1, 2, and 3, shown in Figure 3.2 provide important context for the rest of the survey. These questions provide information about the groups represented, the data they use, and

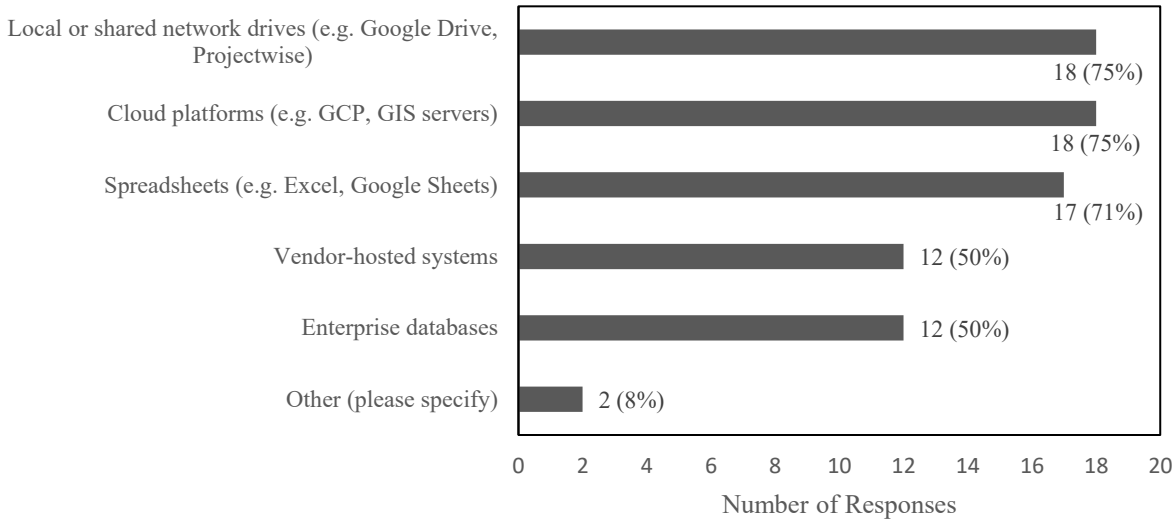


Figure 3.2 Question 3: Where is this data currently stored?

the sources of that data. Traffic & Safety and Technology & Innovation had 4 and 5 responses respectively, but overall, 13 groups had at least one respondent. Traffic & Safety, Technology & Innovation, Right of Way, Preconstruction, Planning, Materials & Pavement, Maintenance, Data & Analytics, Central Maintenance, Central GIS, Bridge, Asset Management, and one unspecified "other" group were represented in the dataset. These groups worked with inventory, construction, maintenance, finance, operations, and employee data. The data in question is primarily stored on network drives, cloud platforms, spreadsheets, and vendor systems.

Questions 5, 6, and 7 provide more context regarding usage habits and more granular information about the datasets users are accessing. The data respondents use varies in update frequency, according to Question 5. 33% of respondents reported that data is updated as needed, while 50% reported that their data was updated daily, weekly, or in real time. These responses demonstrate that 88% of respondents work with data that is changing frequently. Question 6 shows that 92% of respondents use Pathway data in their work. From these responses, it can be inferred that Pathway data is used widely across multiple departments. Question 7 further clarifies how the groups use Pathway data. The uses of the Pathway data are highly varied, with quality control, asset information, measurements, and project tracking among the applications reported. It is also noted that not all applications of the dataset are known to the respondents, suggesting that there are likely additional uses that have not been accounted for.

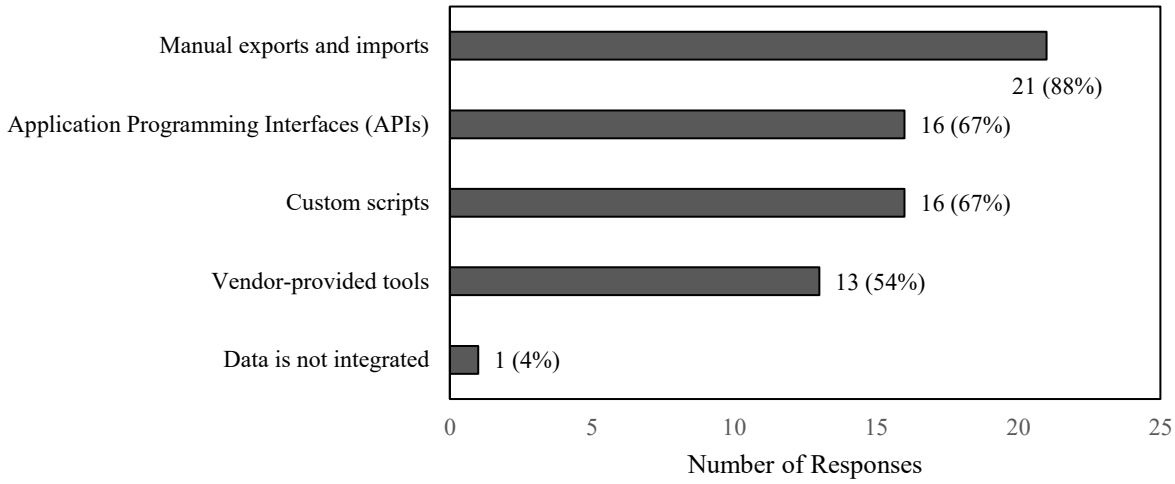


Figure 3.3 Question 13: How is data typically integrated today?

Questions 12, 13, 18, and 2 help determine how data users manipulate, integrate, or otherwise access data. Questions 12 and 13 provide valuable insights into data integration habits. 50% of respondents combine data from multiple systems either weekly or daily, with an additional 42% integrating data once every 1-3 months. With 92% of respondents reporting that they integrate data as part of their daily responsibilities, data integration appears to be a widespread practice across organizational departments. While integration is common, the methods employed are more varied. Only one respondent reported not integrating data, while 88% reported integrating data via manual imports and exports. Users also implemented vendor-provided tools, application programming interfaces (APIs), and custom scripts. From these results, shown in Figure 3.3, it can be inferred that most respondents are using a combination of manual imports and exports, APIs, custom scripts, and vendor-provided tools to integrate data.

Question 18 summarizes the levels of access permitted for datasets, based on responses where participants could select all applicable options. Results indicate that data access is primarily internal, with most datasets being shared within UDOT (88%) and across internal groups (79%), while a smaller portion is accessible to external stakeholders. Specifically, 58% of respondents reported managing data restricted to their own group, whereas 54% reported responsibility for publicly accessible datasets. This overlap highlights the coexistence of both restricted and open-access data within the current data management environment. Question 21

allowed respondents to select all methods that they employ to work with their data. The methods include spreadsheets (96%), geographic information systems (GIS) (92%), dashboards (75%), statistical or scripting tools (58%), asset management systems (50%), and other tools primarily associated with a specific vendor (8%). Spreadsheets often require the lowest bar to entry so it is not unexpected to see such a high volume of responses, however over reliance on spreadsheets can lead to data, lineage, and governance losses that may be preventable by making other methods more accessible.

3.2.2 Data Governance

Survey questions in this section address metadata, data stewardship, data sharing, and taxonomy-related practices. Questions 4, 8, 9, 10, 17, and 20 provide insight into UDOT's current data governance practices.

Questions 4, 17, and 20 provide insight into the current state of data governance practices regarding responsibility. Question 4 represents a significant gap in the understanding of responsibility for different data sets. 13% report that datasets have no single owner or an unclear owner, while the rest of the respondents were split between datasets with informal owners (46%) and datasets with a clearly defined owner (42%). This shows that some groups may be formalizing data stewardship assignments, while others may have an individual assigned as the dataset's responsible party. Question 17 shows a similar divide amongst respondents, with most (54%) resolving conflicting data issues on a case-by-case basis. 25% of respondents reported that conflicts are resolved based on data ownership, 13% reported that there was no clear process, and 8% reported that a formal rule of some kind exists. The wide range of responses indicates that there is no agreed-upon method for resolving data conflicts across groups. Question 20 continues the trend of highly varied responses to questions about data governance. Survey results show that 39% of respondents have formal policies regarding data sharing and use that are unclear, 30% report having no formal policies, 22% report having well-understood policies, and 9% are unsure whether policies exist. These responses are visualized in Figure 3.4. A lack of understanding, awareness, and use of these policies suggests they are not universal, and it is unlikely that respondents can locate or interpret information about them without help.

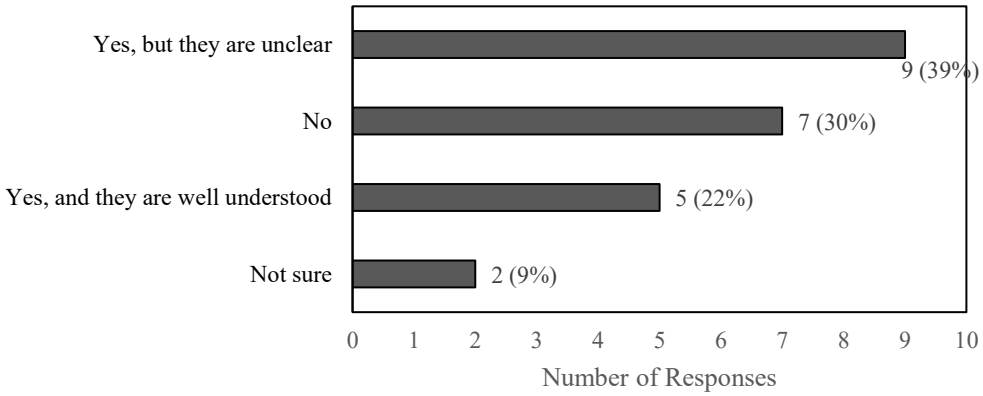


Figure 3.4 Question 20: Are there formal policies governing data sharing and use?

Questions 8, 9, and 10 are intended to develop an understanding of the interpretability of data across groups and of the methods used to support metadata documentation. Question 8 asks if standard definitions are used for key terms within datasets. 63% of respondents indicated that formalized standard definitions are maintained. 17% of respondents reported that the definitions used vary by system; another 17% reported that the definitions exist but are informal; and 1 respondent, representing 4% of the survey takers, was unsure of the practices employed by their group. These results are shown in Figure 3.5. This question initially suggested that perhaps more formal efforts were underway within groups to formalize definitions. However, upon further discussion with UDOT stakeholders, it was determined that the "formalized" definitions were more likely than not to be managed using text documents, spreadsheets, or other methods that

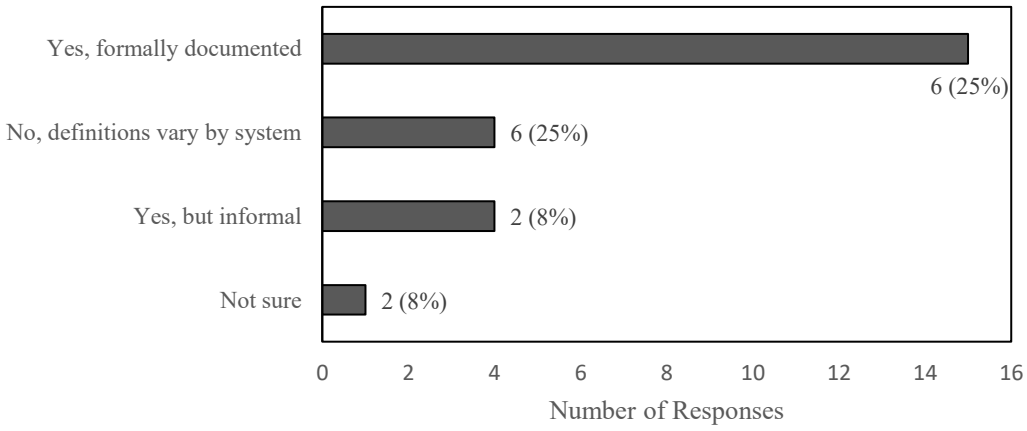


Figure 3.5 Question 8: Are standard definitions used for key terms

would silo data from other groups, creating further distance between groups managing similar data across different systems.

Question 9 asked respondents if multiple systems used different names or definitions for the same concept. Only 1 respondent (4%) reported that this never occurs, with 46% reporting sometimes, 25% reporting frequently, and another 25% reporting rarely. 96% of respondents navigating different naming conventions across systems report that semantics are inconsistent, an issue that can lead to further confusion without clear data stewardship and a unified data management approach. Question 10 is about the current state of metadata management. Encouragingly, no respondents reported that no metadata was available. Metadata across datasets varies in quality and availability: 58% of respondents report partial or missing metadata, and 17% report that all metadata is present but outdated or irrelevant. Only 25% of respondents reported that all metadata is present and up to date. These responses indicate a clear opportunity to improve metadata management, ensuring the relevance and completeness of metadata and their associated catalogs.

3.2.3 Stakeholder Sentiment

Survey questions in this section allowed respondents to share common problems they face, attitudes towards data practices, and sentiments about data. Questions 11, 14, 15, 16, 19, 22, and 23 inform this study on the current problems faced by survey respondents, their data literacy, and their quality of life.

Questions 11, 15, 19, and 22 allowed respondents to report the level of ease or difficulty associated with data in their work. Question 11 allowed respondents to describe how difficult it would be for a new staff member to navigate and interpret their data. No respondents thought that no assistance would be needed to interpret their data with the appropriate resources and documentation. 46% of respondents believed that direction and support would be necessary for most, if not all, processes. 29% thought that some support and direction would be sufficient, while another 25% believed that extended oversight and training would be necessary for a new staff member to understand the data. The volume of respondents believing that significant oversight or mentorship is necessary suggests that the availability and quality of data navigation resources leave room for improvement. Question 15 allowed respondents to report their

confidence level in the accuracy of their data. 17% of responses indicated that the quality of data was dependent on the dataset, where some datasets required stringent accuracy and others were less meticulous. 58% were somewhat confident, and 8% were not confident, indicating that a significant portion of respondents had some concerns about data accuracy, whether those inconsistencies were inaccurate values, duplicates, or missing values. 17% of respondents reported a high level of confidence, likely to represent users who primarily work with more meticulous datasets.

Question 19 asked about the level of difficulty in sharing data with other DOT groups. Responses were mostly skewed towards data sharing being easy, with 39% of respondents saying it was somewhat easy and 22% reporting that it was very easy. A small, but still significant portion of respondents (22%) reported that sharing data was somewhat hard and that this process hinders their work. One respondent (4%) described the sharing of data as very hard. Data sharing is generally not the easiest process for most respondents, however there is some documentation to aid those who need help. The documentation can help some users but may lack clarity for those who are not fully data-literate. Question 22 aligns with the results of question 19. 48% of respondents described the ability of their group to read, understand, use, and communicate with data as "moderate" and 43% described the ability of their group as "high." Only 9% of respondents reported their ability as "low." Groups with "moderate" data literacy would likely be able to follow most data adjacent documentation, with more complex documentation potentially proving confusing. The visualization of the responses to question 22 are shown in Figure 3.6.

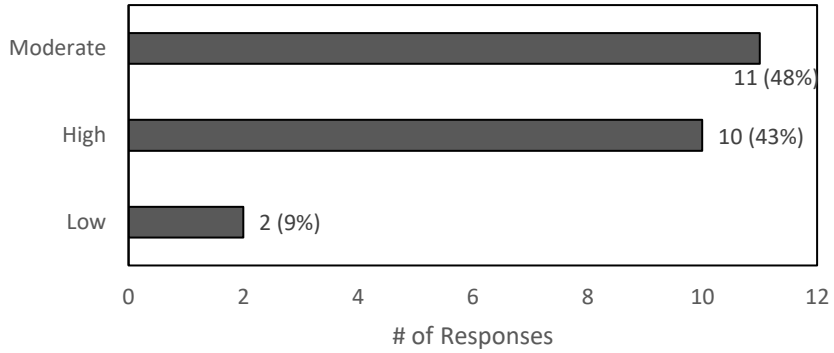


Figure 3.6 Question 22: How would you rate your group's ability to read, understand, use, and communicate with data for better decision-making?

Questions 14, 16, and 23 discuss the different barriers each group faces with its data. Question 14 assessed the different barriers to data integration across systems. The most frequently reported issues were limited time or resources (75%), vendor restrictions (67%), different definitions for the same data (63%), and inconsistent formats (63%). Less reported but still significant barriers included legacy systems (42%), unclear ownership (38%), and other reported problems such as poor location information, poor data access, and lack of documentation (17%). Many of the issues reported in this question would likely be addressed by improved quality control and more unified data management practices. Improved metadata collection and specifications are critical to those processes. The results for question 14 are shown in Figure 3.7. Question 16 provides similar input to question 14, this time about common data quality issues. All responses were given more than 40% of the time. Missing values (71%), inaccurate or out-of-data information (63%), inconsistent schemas between systems (54%), unknown data sources (46%), and duplicate records (42%) were all reported as common data quality issues. Once again, these issues would likely be resolved through stronger quality control and more unified data management.

Question 23 asked what limits respondents' ability to manage better or use data. Respondents reported limited staff or time (75%), organizational silos (63%), technical limitations (50%), training gaps (38%), and lack of standards (38%) as frequent barriers. 2

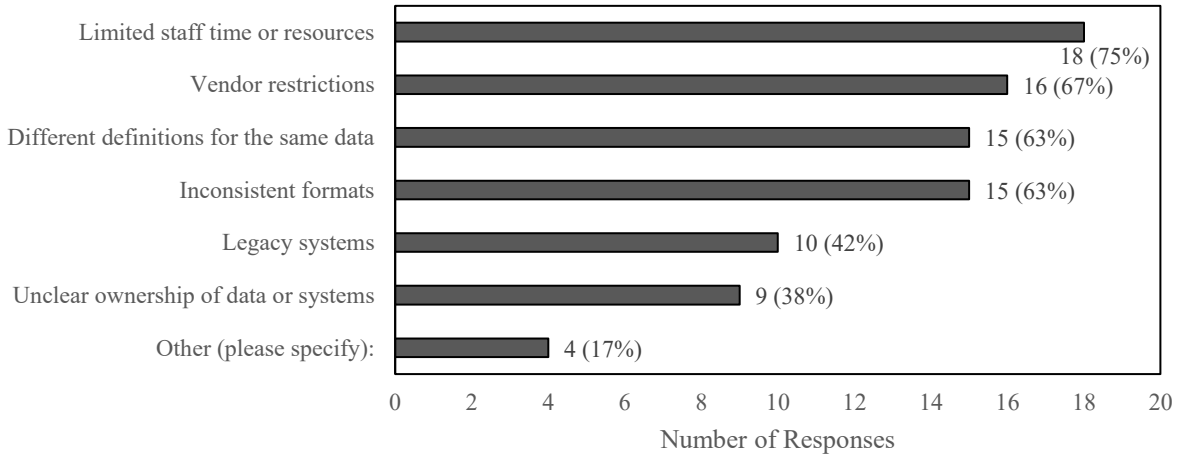


Figure 3.7 Question 14: What are the biggest barriers to integrating data across systems? (Select all that apply)

respondents (8%) reported that the barriers faced depended on the datasets in question. Based on further discussion with UDOT stakeholders, it was determined that the lack of staff or time responses was primarily due to the time required to make the data interpretable or otherwise usable. Following standards, tracking lineage, and locating metadata are just some of the challenges associated with data use across UDOT groups. A unified data management framework would require consistent tracking and timely updates to reduce the prevalence of these challenges in the future.

3.3 Desired or Ideal Data Management Practices

Many of the questions about the current state of data management practices did not require open-text fields. The approach taken to establish the future, idealized state of data management allowed for greater flexibility in responses, enabling capture of exactly what stakeholders think is needed. Questions 24, 25, 26, and 27 allowed stakeholders to provide input and raise concerns regarding an organization-wide data management solution.

3.3.1 Future State

Questions 24, 25, and 27 allowed respondents to identify needs and potential improvements in UDOT's current data management approach. These responses were used to inform the development of a proof-of-concept schema and relevant supporting management tools. Questions 24 and 25 do not have a free-response component, whereas question 27 encouraged respondents to write their own answers. Question 24 asks which improvements would have the greatest positive impact on respondents' work, provides a list of 5 impacts, and limits respondents to 3 selections. Centralized data access (57%), improved data quality (57%), easier system integration (52%), better documentation (48%), and clearer data definitions (48%) were all selected by users approximately equally. Every benefit listed is an outcome of ontology and EDM implementation (Bernardo et al., 2024; Boadi et al., 2026; Lei, 2023; Office of Management and Budget, 2020; Wang et al., 2024). The uniform distribution of responses, shown in Figure 3.8, indicates that these improvements are broadly valued, each contributing meaningfully to enhancing usability and effectiveness across the organization. Question 25 asked respondents how valuable a DOT-wide data vocabulary or organized and defined data structure

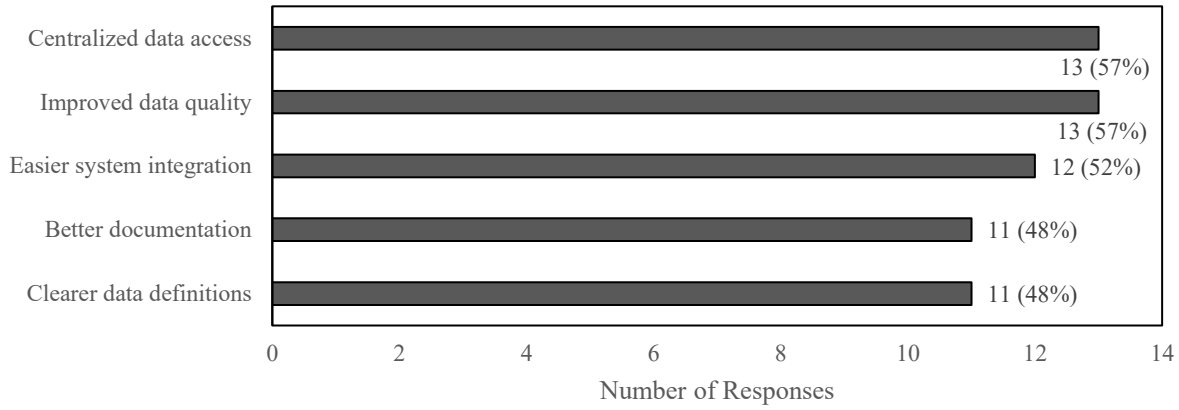


Figure 3.8 Question 24 Responses: Which improvements would have the biggest positive impact on your work?

would be. Extremely valuable (75%) and somewhat valuable (17%) were the most frequent answers. Neutral (8%) was a less common sentiment, and no respondents reported it as not valuable.

Question 27 allowed respondents to report what they considered most useful for their work in the context of a DOT-wide method for organizing and structuring data. Survey responses consistently emphasized the need for greater standardization, transparency, and integration within UDOT's data management environment. A dominant theme was the desire for centralized, authoritative data sources, with respondents highlighting issues arising from duplicate datasets, siloed systems, and inconsistent access across divisions. Many respondents stressed the importance of clear data definitions, schema standards, and accessible documentation, including explanations of data fields, decision-making processes, and relationships between datasets. Closely related to this was the need for clear data ownership and governance, including who is responsible for data, how it can be shared, and the quality standards that must be met. Notably, several responses specifically identified the need for a universal asset identifier to provide a consistent mechanism for linking datasets, reducing duplication, and enabling interoperability across systems.

Another key theme was data quality and reliability, with frequent references to the need for improved QA/QC processes, up-to-date information, and validation workflows to reduce duplication and outdated records. Respondents also identified the importance of data

discoverability and accessibility, expressing a desire for systems that make it easier to locate and understand available data without extensive manual searching or reliance on institutional knowledge. Finally, several responses pointed to more advanced capabilities, including system integration, linked data structures, and digital twin concepts, which would enable more comprehensive analysis, support future decision-making, and allow the organization to anticipate and answer complex questions more effectively.

3.3.2 Concerns

Question 26 asked respondents to identify concerns related to implementing a common method for organizing and structuring data. Governance and ownership emerged as the most prominent concern (67%), followed by technical feasibility (50%) and the effort required for adoption (50%). Less frequently cited concerns included loss of flexibility (29%) and having no concerns (4%). Additionally, 21% of respondents provided "other" feedback, emphasizing risks such as potential data loss during implementation, inconsistent terminology and usage across groups, and the challenge of aligning diverse organizational needs under a standardized approach. These responses also highlighted concerns regarding timeline feasibility, the possibility that standardization may not meet all user needs, and the impact of staff turnover on sustaining long-term ownership and continuity of vision. The results of this question are shown in Figure 3.9.

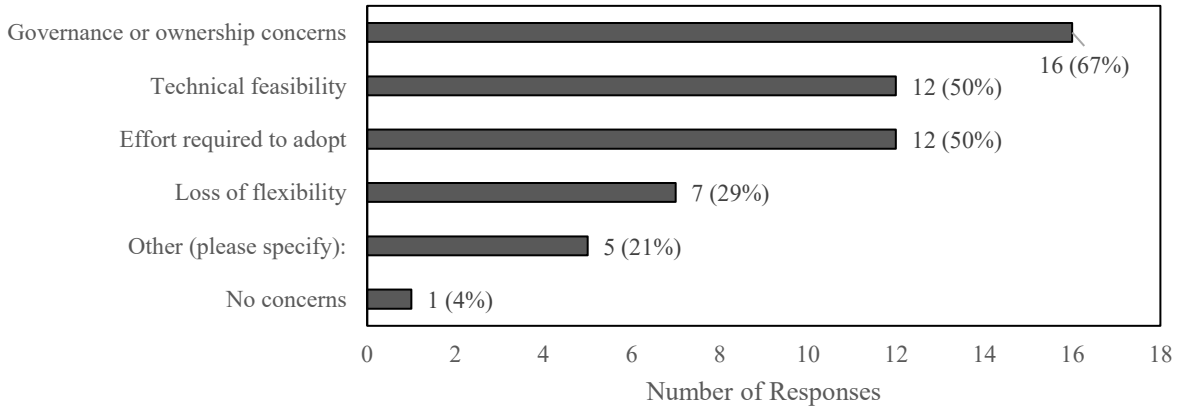


Figure 3.9 Question 26: What concerns would you have about implementing a common method for organizing and structuring data?

3.4 Gaps Between Current and Ideal Practices

The primary gap between current and ideal practices is observable in the absence of a unified framework to support interoperability, governance, and data quality. Current practices lead to siloed systems, inconsistent definitions, manual integration processes, and incomplete metadata. The desired state emphasizes centralized access, standardized schema, clear governance, and linked data. The gaps limit data usability, accessibility, and the organization's ability to support advanced analytics and decision-making. Analysis of stakeholder interviews, survey responses, and existing data management practices revealed four primary areas requiring improvement: data standardization and semantic consistency; data integration; data governance and management practices; and data quality, accessibility, and usability. These themes are highly interconnected, with deficiencies in one area often reinforcing challenges in another. The following sections examine each gap in detail by comparing current practices to the desired future state and identifying opportunities to improve enterprise-wide data management capabilities.

3.4.1 Data Standardization and Semantic Consistency

A lack of unified structure for how data is defined, labeled, and related is a consistent theme across all feedback. Inconsistent terminology and definitions across systems create intergroup discord, leading to significant time spent interpreting data independently or locating subject-matter experts to guide usage. While definitions exist, they are often siloed within individual systems or documents and are neither consistently enforced nor accessible across the organization. Inconsistent schema and data structures further exacerbate these challenges, as differing taxonomies and representations of similar concepts create barriers to quality control, integration, and semantic alignment. Metadata, where available, is often incomplete or misaligned with data structures, limiting the ability to fully understand data context, relationships, and lineage.

In the current environment, datasets referencing the same asset lack a consistent mechanism for linkage, which reinforces fragmentation across systems, an issue that could be addressed through the implementation of a universal asset identifier. Collectively, these inconsistencies hinder interoperability, reduce efficiency, and limit the organization's ability to

scale data integration and analysis. A standardized ontology or schema, supported by consistent definitions, accessible documentation, aligned metadata, and universal identifiers, would enable cross-system understanding, semantic consistency, and more effective data-driven decision-making.

3.4.2 Data Integration

Fragmented systems and inefficient methods for combining data were consistently identified in both stakeholder discussions and survey responses. Data is currently distributed across disparate platforms, including spreadsheets, network drives, cloud environments, and vendor-managed systems, which complicates both data management and integration efforts. Integration practices are largely manual, with a heavy reliance on imports, exports, and user-driven workflows. These approaches are not scalable and introduce risks related to inconsistency, duplication, and data loss. Additionally, the absence of standardized integration mechanisms, such as governed data pipelines or application programming interfaces (APIs), contributes to variability in how data is combined across groups.

Siloed data environments further exacerbate these challenges, as datasets representing similar or identical assets are maintained independently across systems. Without a centralized or logically unified architecture, there is no authoritative reference point for validating or reconciling data across sources. This lack of a common framework limits interoperability and hinders maintaining consistency across integrated datasets. The current state also lacks clear data lineage and traceability, making it difficult to understand how data is transformed and propagated between systems. Variability in update frequency, including a mix of real-time and periodic updates, introduces additional inconsistencies that are not systematically managed.

The ideal state for data integration is a centralized or logically unified architecture supported by standardized integration mechanisms, automated data pipelines, and interoperable systems. This architecture would incorporate consistent data models, enforce linkage through mechanisms such as a universal asset identifier, and provide clear lineage and traceability. Together, these capabilities would enable scalable, reliable integration and establish a single, authoritative source of truth for organizational data.

3.4.3 Data Governance and Management Practices

Inconsistent stewardship, policies, and lifecycle management of data further contribute to inconsistencies and barriers to effective data usage. Unclear or informal data stewardship structures create ambiguity regarding responsibility, accountability, and authority over datasets. In many cases, responsibility is either assumed or distributed without formal designation, which increases the likelihood of mismanagement, duplication, and conflicting data representations across systems. The absence of standardized governance policies exacerbates these challenges. Data sharing, access, and usage policies vary across groups and do not account for other internal policies or practices. Where policies do exist, they are often informal, difficult to locate, challenging to interpret, and are inconsistently applied. Inconsistent governance practices limit an organization's ability to ensure compliance, maintain consistency, and enforce best practices across departments.

Data lifecycle management is similarly fragmented. There is no consistent approach for managing data from creation and collection through storage, maintenance, and eventual archival or deletion. As a result, datasets may persist beyond their intended use, become outdated, or lack proper version control. This contributes to duplicate, incomplete, or obsolete data, further eroding trust in the data and increasing the effort required to validate and use it effectively. Metadata management practices further reflect governance gaps. While metadata exists in many cases, it is often incomplete, outdated, or non-conforming to current data structures. Without effectively maintained metadata, users lack the context to interpret data, understand its lineage, or assess its quality and reliability.

The ideal state for data governance and management practices includes clearly defined, formally assigned data stewardship roles, standardized, well-communicated governance policies, and consistent lifecycle management procedures. A robust governance framework would ensure that data ownership is transparent, policies are enforceable and accessible, and data is managed systematically throughout its lifecycle. Additionally, comprehensive metadata standards, along with defined processes for conflict resolution and quality assurance, would support improved data consistency, reliability, and usability across the organization.

3.4.4 Data Quality, Accessibility, and Usability

Certain data is difficult to trust, access, and effectively use. Data quality issues, including missing values, duplicate records, outdated information, and inaccuracies, contribute to uncertainty regarding data fidelity and reliability. These issues reduce user confidence and increase the time required to validate data before use, thereby limiting overall efficiency. Limited and inconsistent quality control processes further exacerbate these challenges. Validation and verification procedures are not uniformly applied across datasets or systems, allowing errors to persist and propagate. In many cases, quality assurance efforts are reactive rather than systematic, placing additional burden on individual users to identify and correct issues during analysis.

Accessibility and discoverability of data also present significant barriers. Data is often distributed across multiple platforms and storage locations, requiring users to rely on prior knowledge or informal communication to locate relevant datasets. The absence of centralized catalogs or intuitive access points makes it difficult for users to identify what data exists, where it is stored, and how it can be used. Usability is further constrained by inconsistent documentation and varying levels of data literacy across groups. While some users are able to navigate datasets effectively, others require substantial guidance or training to interpret and apply the data. Documentation, where available, may not be sufficiently clear, comprehensive, or accessible to support all users, particularly those with moderate levels of data literacy. This variability creates inefficiencies and limits the broader usability of data across the organization. Additionally, the reliance on manual workflows and tools, such as spreadsheets, increases the risk of errors and reduces the scalability of data use. These approaches often lack built-in validation, version control, and lineage tracking, which further contributes to inconsistencies and limits the ability to maintain high-quality data over time.

The ideal state for data quality, accessibility, and usability includes implementing standardized quality control processes to ensure data is accurate, complete, and consistently maintained. Data should be easily discoverable through centralized catalogs or structured access points, with clear, accessible documentation to support its interpretation and use. Systems should be designed to accommodate varying levels of data literacy, reducing reliance on specialized

knowledge and enabling broader participation in data-driven decision-making. Collectively, these improvements would enhance trust in data, reduce inefficiencies, and support a more effective, scalable use of data across the organization.

3.5 Summary

This chapter examined the current state, the desired future state, and key gaps in UDOT's asset data management practices. While the Transportation Asset Management Plan (TAMP) provides a strong foundation for data-driven decision-making, survey responses and stakeholder input indicate that data management practices remain fragmented across the organization. Data is distributed across disparate systems, integrated through largely manual processes, and governed by inconsistent policies, stewardship structures, and documentation practices. These conditions contribute to challenges in data standardization, interoperability, data quality, accessibility, and usability.

The analysis identified four interconnected areas requiring improvement: data standardization and semantic consistency; data integration; data governance and management practices; and data quality, accessibility, and usability. Stakeholders expressed a strong need for centralized access, standardized data structures, improved metadata and documentation, clearer governance, and mechanisms to support data linkage, such as a universal asset identifier. Collectively, these improvements would support more reliable data integration, increased discoverability, and improved decision-making capabilities. Several underlying factors, including organizational silos, the absence of an enterprise-level data framework, inconsistent governance and stewardship practices, reliance on manual processes, incomplete metadata, and varying levels of data literacy across stakeholder groups, drive the persistence of these gaps. Together, these factors create fragmented data environments that limit the organization's ability to effectively manage, integrate, and utilize asset information.

A summary matrix is provided in Figure 3.10, comparing the current and desired future states of asset data management. These findings establish the foundation for Chapters 4 and 5, which present a proposed ontology framework, schema, and implementation recommendations to

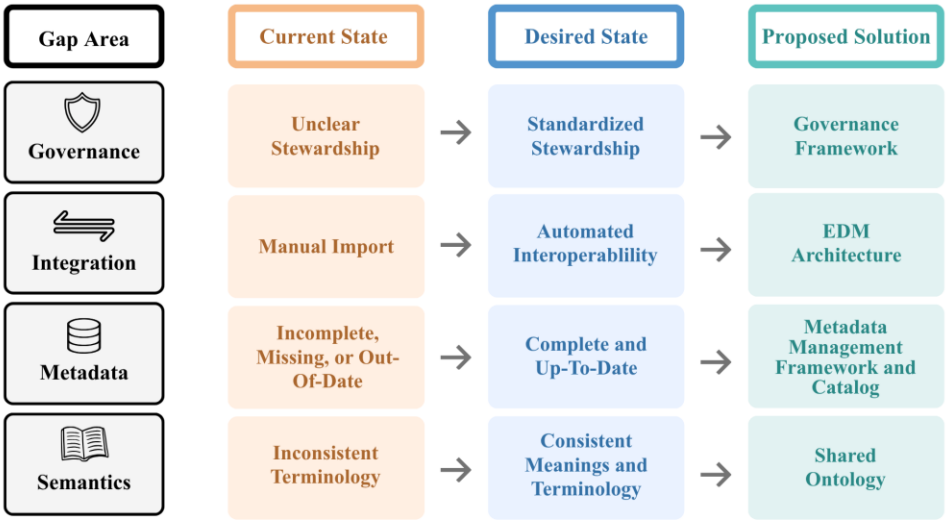


Figure 3.10 A summary matrix of gap analysis findings.

address the identified gaps and support a more integrated approach to enterprise data management.

4.0 SCHEMA DEVELOPMENT

4.1 Overview

This chapter discusses the proof-of-concept schema developed to accommodate the datasets selected for this study (i.e., barrier, pavement, sign face, and sign assembly). The schema development process was iterative, implementing input from the technical advisory committee, UDOT stakeholders, and relevant literature. The role of the schema within the system is to support a more unified taxonomy and to set the stage for the further development of dataset integration and ontology. This schema is not comprehensive, as it has a limited scope and focuses solely on asset inventory. The process for developing the schema and foundational ontology followed the steps outlined in Figure 4.1. These steps align with the data provided by UDOT, the gap analysis, the framework implementation, and the schema implementation.

The management framework utilized in this project is contained within Google Dataplex (*Dataplex Universal Catalog*, 2026). While the principles and methodologies applied throughout this study are platform-agnostic, the terminology, architecture, and tools referenced are based on the Dataplex ecosystem. The framework was developed to support standardized data governance,

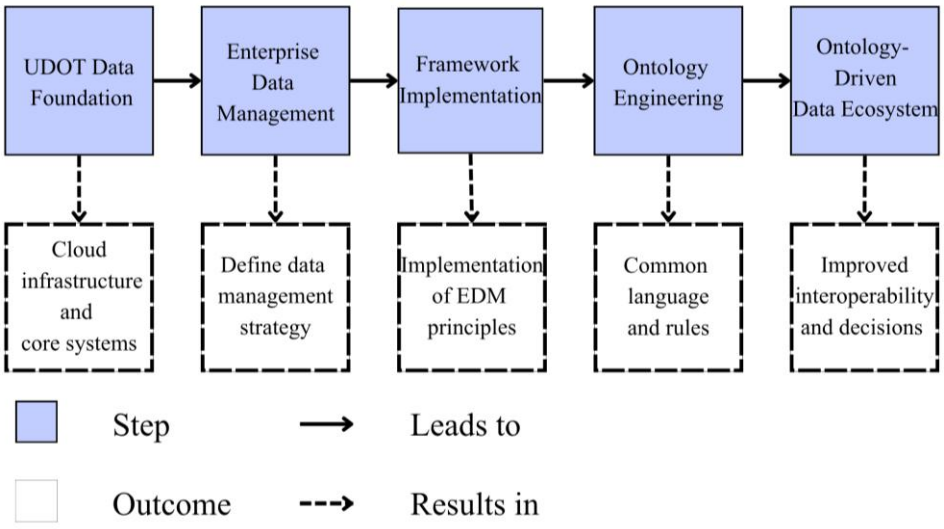


Figure 4.1 Conceptual progression from foundational enterprise data systems toward an ontology-driven transportation data ecosystem.

improved interoperability, and enhanced data lifecycle management across the selected datasets. Development of the framework began with establishing design objectives, followed by defining the governance structure and evaluating available data quality control methods. These components were then integrated into a unified data lifecycle management environment. The following subsections describe the development and implementation of each component of the management framework.

4.1.1 Design Objectives

The design objectives for the management framework included implementing more common, standardized fields and definitions, improving interoperability through semantic unity, increasing scalability, and establishing more well-defined governance practices. Within the Dataplex platform, there are tools and features to support all those objectives. Field or column standardization is managed through the *Knowledge Catalog* in Dataplex. Semantic unification occurs through the management of the *Knowledge Catalog* and lineage tracking. Scalability is ensured through automated metadata cataloging, and governance is improved through permissions management, data profiling, and the ability to attach asset-specific metadata to assets.

4.1.2 Governance Structure

Within the governance structure, clearly defined roles and responsibilities are essential. Formalizing data stewards, responsible for the collection, management, and reporting of data, alongside asset stewards, responsible for the long-term management and sustainability of assets, improves accountability and communication between data and engineering stakeholders. Tracking stewardship assignments and data usage also improves transparency, discoverability, and continuity during personnel transitions.

Assessment of the existing data landscape revealed inconsistent ownership, informal maintenance practices, and siloed management approaches, all of which contributed to variability in data quality and usability. The Dataplex framework addresses these challenges through centralized cataloging, metadata management, and governance capabilities that improve visibility into stewardship responsibilities and managed data assets.

Data policies and standards establish the rules and conventions necessary to support consistency, reliability, and interoperability across datasets. Standardized classifications, naming conventions, metadata requirements, and formatting structures reduce ambiguity, improve semantic alignment, and facilitate integration across heterogeneous systems. Metadata describing dataset ownership, origin, update frequency, and attribute definitions further improves data discoverability and interpretability. The need for these standards became evident through the prevalence of inconsistent naming conventions, incomplete metadata, and differing data structures observed across the existing data environment.

Implementing policies and standards within Dataplex provides a centralized governance solution that promotes consistent management practices across data assets. Through standardized governance structures, metadata requirements, and organizational conventions, the framework establishes a foundation for a more reliable, interoperable, and scalable data management environment.

Data quality and lifecycle management are also critical components of the framework. Quality considerations include completeness, consistency, accuracy, timeliness, and reliability, while lifecycle management encompasses the creation, storage, maintenance, use, and eventual archival of data. Assessment findings revealed challenges, including missing values, duplicate records, inconsistent formatting, outdated information, and fragmented maintenance practices, which reduced confidence in the data and created barriers to integration. The absence of formal lifecycle procedures further contributed to uncertainty regarding ownership, maintenance responsibilities, and data reliability.

Dataplex supports these functions through centralized governance, metadata organization, and improved visibility into managed datasets. Standardized structures and metadata practices improve consistency and interoperability, while validation and monitoring processes help identify common quality issues such as incomplete records, duplicate entries, and inconsistent formatting. Although the implementation developed in this study serves as a proof of concept, it establishes a foundation for more advanced quality monitoring, governance enforcement, and lifecycle management capabilities. Collectively, these practices improve the reliability,

accessibility, and long-term sustainability of organizational data while supporting future integration, analytics, and decision-support applications.

4.1.3 Management Framework Implementation

Dataplex enables cross-organizational communication by integrating two primary functions: a business glossary and a metadata catalog (*Dataplex Universal Catalog*, 2026). The glossary serves as a foundation for non-technical stakeholders in a given domain to interpret terms and concepts that might otherwise be foreign to them. Entries in the glossary clarify the "what" and "why" of data and allow the linking or related terms and synonyms. By standardizing this vocabulary, organizations can move past technical jargon to ensure that business intent is understood by everyone, regardless of their role. In addition to the glossary, the metadata catalog addresses the "how" and "where" of collected data. It offers engineers and analysts a detailed view of data structures, formats, and underlying constraints. Rather than operating in silos, these tools work synchronously. When these business terms and technical definitions are mapped to entries in Dataplex, they create a unified data language that bridges the traditional divide between business logic and technical execution while also automating quality control processes.

The infrastructure supporting this metadata environment relies on *Aspect Types* and *Aspects.* This system replaces traditional, unstructured tagging with a machine-readable framework designed for rigorous governance. An Aspect Type, a predefined metadata template, defines the standards for metadata fields and data types. An Aspect is the functional application of that template to a specific asset. For instance, while an "Asset Governance" Aspect Type establishes the requirement for tracking data cycles, the Aspect itself would store the specific value, such as "bi-annual review." This architectural depth enables highly granular discovery and ensures governance remains consistent across the entire data ecosystem.

4.2 Schema Development

In addition to the management framework, a corresponding asset inventory schema was created to align governance processes, data collection practices, and storage structures. The schema, shown in Figure 4.2, was designed to establish a foundation for the continued advancement of asset management practices at UDOT. The schema development followed an

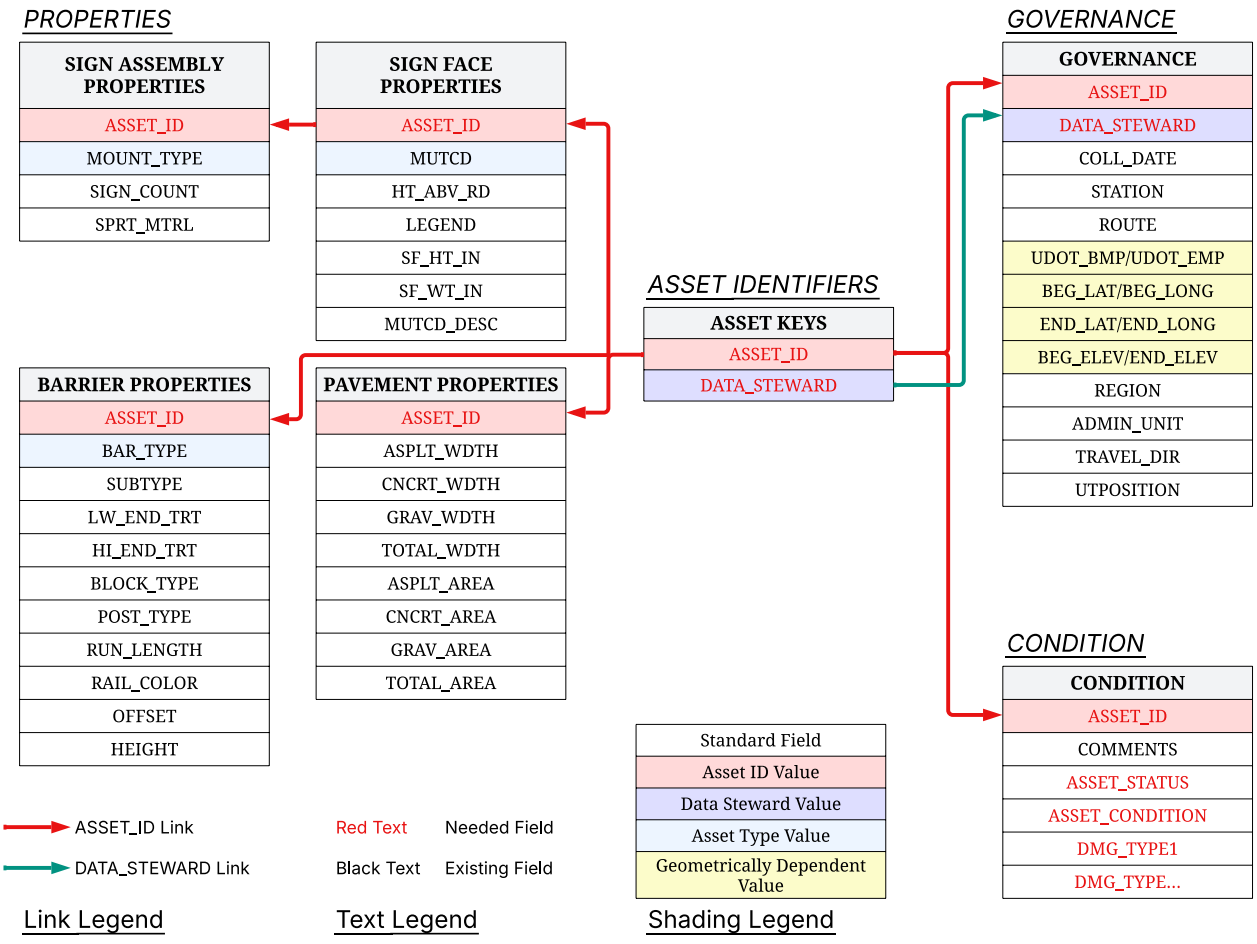


Figure 4.2 Proposed Schema

iterative process informed by reviewed literature, stakeholder feedback from UDOT personnel, modern enterprise data management principles, and the integration of capabilities available within Dataplex. In this section, the source data, design process, validation and testing, and limitations and future developments associated with this schema will be discussed.

4.2.1 Source Data

The initial data considered for schema development were sourced from datasets selected by members of the UDOT technical advisory committee. These datasets contained varying numbers of attributes, some common across all assets and others unique to specific asset types. Preliminary review revealed that all assets included information on location, data collection,

asset properties, and an unstructured comment field used by contractors to communicate additional information not otherwise captured in the existing schema. Further inspection indicated that condition-related information was frequently reported within the comment field rather than through standardized attributes, creating inconsistency, misinterpretation, and incomplete representation of defects. Additionally, assets lacked unique identifiers and instead relied primarily on geographic coordinates and contractor travel direction to determine asset location and route-side association. These observations highlighted several limitations in the existing data structure and informed the development priorities of the proposed schema.

4.2.2 Design Iterations

The organization and structure of the schema underwent multiple iterative revisions throughout development. At the time of initial review, the existing fields were largely uncategorized and inconsistently structured. The groupings proposed in the revised schema were informed by the management framework in Dataplex, where individual columns function as aspects within broader hierarchical aspect types. The development of the framework also considered commonly adopted ontological practices, specifically focusing on the hierarchical class-subclass axioms shown in Figure 4.3.

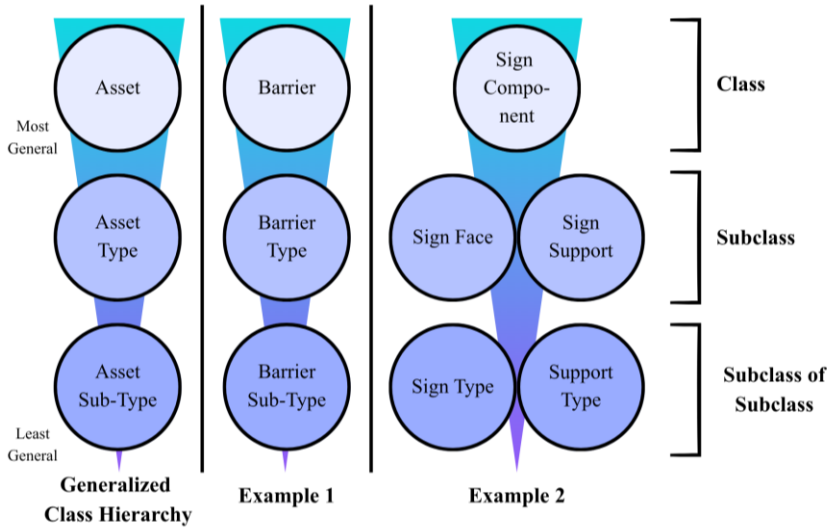


Figure 4.3 Example ontology class hierarchies illustrating increasing semantic specificity across transportation asset domains.

Development of the proof-of-concept schema began by organizing asset information into several categories: governance, geolocation, collection data, properties, and condition. Governance fields captured administrative and stewardship information; geolocation fields stored spatial information; collection data documented acquisition-related details; and properties contained asset-specific characteristics. Among these categories, properties exhibited the greatest variability due to the unique informational requirements of different asset types. Review of existing datasets and stakeholder feedback identified condition data as a significant weakness, as asset conditions were recorded almost exclusively through unstructured comments. While this approach provided flexibility, it limited the ability to consistently query, analyze, and compare condition information across assets.

To address these challenges, several schema concepts were evaluated using both literature-based approaches and stakeholder input. A key finding from this process was the need for a standardized asset identification methodology that can link information across systems and throughout the asset lifecycle. The final framework simplified the schema by consolidating common fields into a unified governance category while retaining separate properties and condition categories. Existing asset information was preserved where possible, and additional fields were developed to improve the structure and usability of condition data. Collectively, these modifications established a more standardized and scalable foundation for asset management, interoperability, and future ontology development.

Because all condition-related information within the original datasets was stored in an unstructured comments field, the existing schema provided limited support for standardized reporting, automated querying, and consistent interpretation of asset conditions. Variability in terminology, reporting methods, and levels of detail created challenges in identifying recurring defects, comparing asset conditions across datasets, and conducting broader lifecycle analysis. To address these limitations, the proposed schema introduces three new structured fields and an additional flexible damage-reporting structure to improve interoperability, semantic consistency, and the long-term usability of condition-related data. A summary of these proposed fields can be found in Figure 4.4.

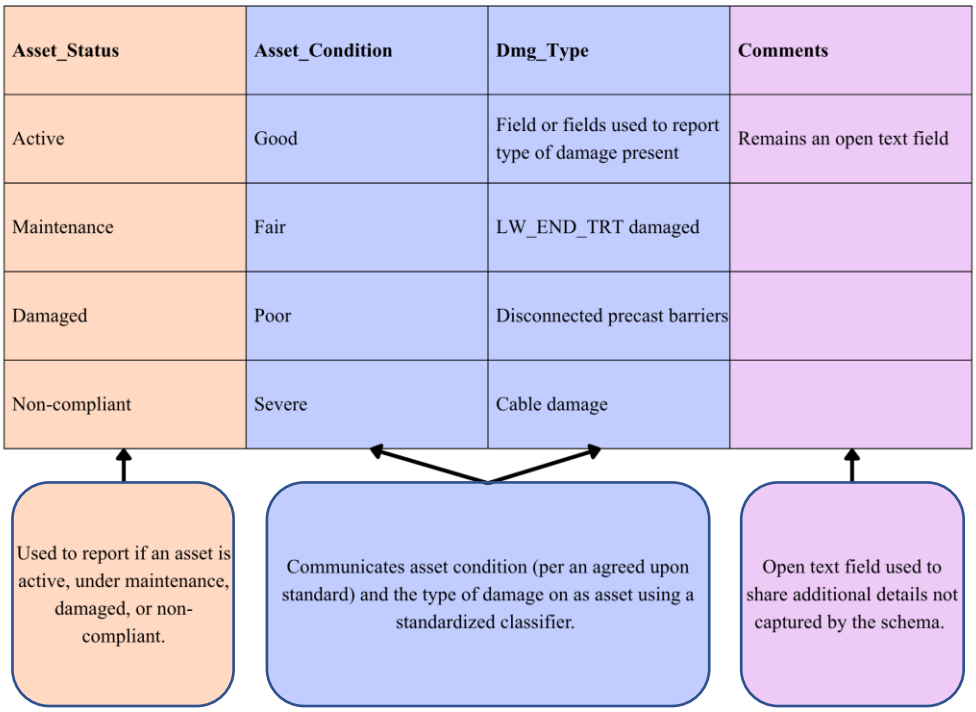


Figure 4.4 Proposed new fields

The newly introduced fields included asset status, asset condition, comments, and damage type. The asset status field was designed to support cross-group interoperability by tracking an asset's operational accessibility or usability. While the physical condition of an asset may remain unchanged between inspection cycles, its operational status may change frequently due to construction activities, temporary closures, maintenance operations, or other external project impacts. Incorporating asset status into the schema, therefore, supports improved coordination among organizational groups and provides opportunities to optimize inspection and collection planning processes.

The asset condition field was designed to provide a standardized, generalized assessment of asset health. Although asset condition evaluations may vary depending on the standards, thresholds, or criteria used for assessment, implementing a consistent condition classification structure provides a simplified way to track deterioration trends over time. Standardizing this information also improves the ability to compare asset performance across datasets and supports

future analytical applications such as deterioration modeling, trend analysis, and automated condition assessment workflows.

The damage type structure was developed to provide more granular information regarding specific forms of asset deterioration or defect occurrence. Unlike the generalized condition field, damage types were designed to accommodate varying numbers of damage descriptors depending on asset type and reporting requirements. Common forms of reported damage observed within each dataset informed the predefined damage classifications included within the proposed structure. Because different asset types experience different forms of deterioration, the available damage descriptors vary accordingly across datasets. Structuring these classifications through predefined damage categories reduces ambiguity in reporting while improving consistency, interoperability, and the ability to conduct more reliable querying and comparative analysis across datasets.

A free-text comments field was retained within the revised schema to preserve flexibility in contractor and steward communication. This field provides a mechanism for reporting unique observations or contextual information that is not otherwise captured by standardized attributes. However, by relocating structured condition and damage information into dedicated fields, the comments field becomes significantly more focused and easier to interpret. Rather than serving as the primary repository for all condition-related reporting, the comments field instead functions as a supplementary communication mechanism for exceptional or non-standard observations. This restructuring substantially improves the ability to process, interpret, and analyze both structured condition data and unstructured supporting commentary. Collectively, these fields provide structure to condition reporting, resulting in a more standardized and interoperable framework capable of supporting scalable asset management, improved lifecycle tracking, and future intelligent analytics applications.

4.2.3 Validation and Testing

The validation and testing of the developed schema took place in the Dataplex environment. This process was evaluated by whether the proposed schema categories, aspect types, aspects, and business glossary entries could be accurately represented in the management environment. The schema categories were represented as aspect types, and the individual fields

were represented as aspects. This process demonstrated that the proposed schema could be operationalized within a metadata catalog rather than remaining only a concept. The implemented schema also included an evaluation of its ability to support automated quality control. Implementing example rules, such as the enumerated values in the Region column, ensured the schema could support future automated data quality scans, improved error detection, and more consistent reporting practices. A dataset with columns from the original schema and an additional, fabricated ID number column, using 50 rows of data, was passed through the framework, applying simple rules, business terms, and column-level and dataset-wide aspects. The Dataplex platform enabled effective data parsing using these rules, automating quality control processes rather than leaving them manual.

An initial, poor-quality dataset was passed through the dataplex framework. This dataset contained duplicate IDs, incorrect regions, and erroneous barrier types. The platform flagged all errors for review. Upon correcting these errors to conform with the rules, a simple query was run on the dataset to extract all rows with an asset ID between 1999 and 2012, in any region, with a barrier type of cable, precast, or guardrail. As a final step to this validation process, the business dictionary and metadata catalog were reviewed to determine whether schema terminology could be documented and discovered consistently across the platform. This validation confirmed that the proposed framework supports improved interpretability, discoverability, stewardship tracking, and semantic alignment across UDOT asset datasets. Business terms can easily be attached to columns to make the data more interpretable for any user. Attached business terms provide transparency on meaning, data use, and additional stewardship information. A validation figure, including aspects applied, business terms attached, and rules applied, of the query run on the dataset, is shown in Figure 4.5.

4.2.4 Limitations and Future Development

The primary limitation of the designed schema is the dependence on an asset identification method that does not yet exist. An asset identifier provides a simple relational key that connects individual assets to relevant governance information, condition data, asset properties, inspection history, and future datasets. The asset ID also serves to prevent asset duplication: if all other characteristics are identical except for an ID number, it can be safely

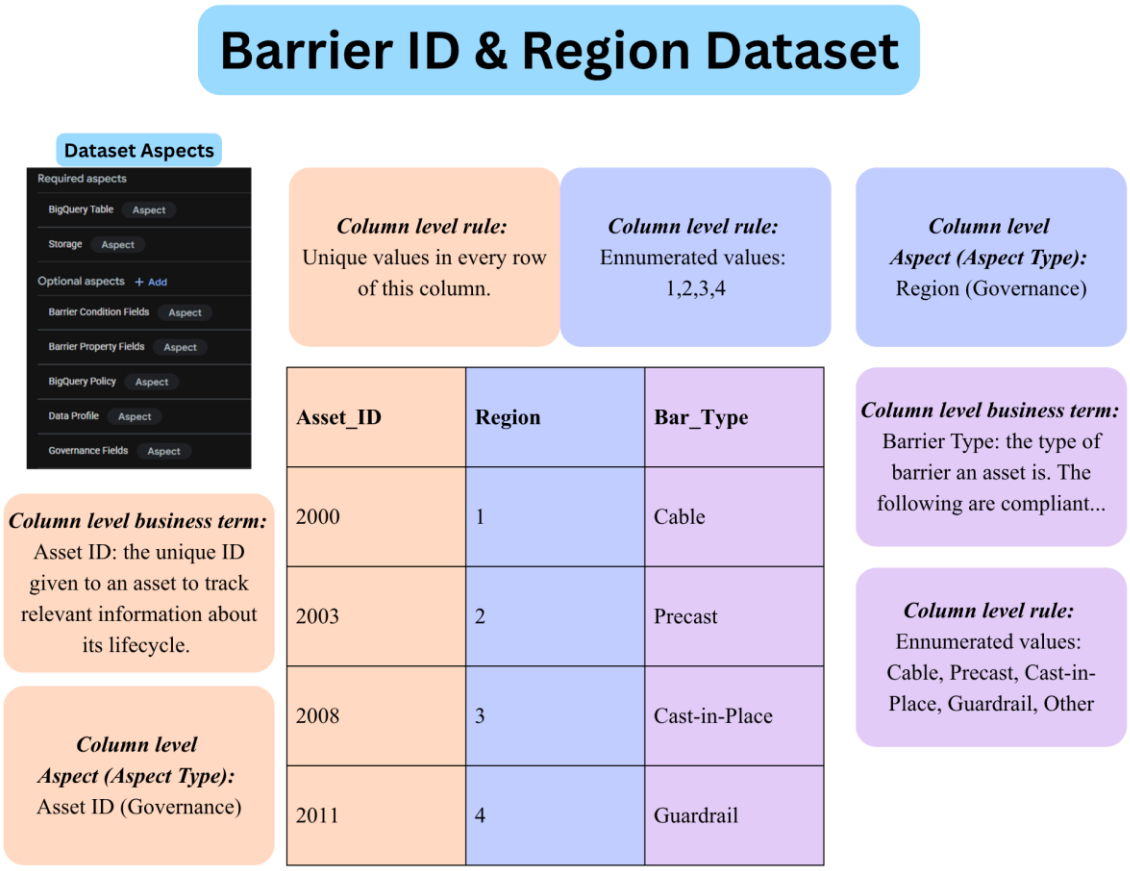


Figure 4.5 The validation query results with attached business terms, rules, aspects, and output of the query.

assumed that a duplicate was accidentally created. An asset ID also serves as a critical component of system interoperability, enabling information from multiple systems to be effectively associated with the same physical asset. Asset identification is also a critical component of governance and stewardship. Implementing a universal asset identification structure enables clearer assignment of stewardship responsibilities, improved accountability, and stronger traceability throughout the data lifecycle. By associating governance information directly with individual assets, stewardship roles become more transparent across systems. Associating this information at the asset level improves the ability to determine who is responsible for maintaining, validating, updating, and utilizing specific asset records. Asset identification also strengthens data lineage by enabling consistent tracking of information associated with an asset across datasets, systems, inspections, and lifecycle stages. Rather than

relying on geographic coordinates, contractor interpretation, or temporary identifiers, a persistent identifier provides a stable reference point that links governance information, condition records, maintenance history, and property data over time. This improves confidence in the continuity and reliability of asset records while reducing ambiguity and duplication across systems.

The inclusion of a universal asset identifier additionally supports more comprehensive lifecycle management practices. Persistent identification enables organizations to track changes in asset condition, operational status, inspections, maintenance activities, rehabilitation efforts, and eventual replacement throughout the existence of an asset. As a result, assets can be managed as continuous entities rather than isolated inspection records. This capability improves long-term monitoring, supports trend analysis and deterioration modeling, and establishes a stronger foundation for future predictive analytics and data-driven decision-making applications.

Within the broader management framework, asset identification also enhances interoperability between organizational groups and data systems. Because the asset identifier functions as a shared relational key across governance, condition, and property datasets, information can be more effectively integrated and queried across previously siloed systems. Establishing standardized asset identification, therefore, represents a foundational step toward scalable enterprise data management, semantic interoperability, and future digital asset management initiatives.

Another limitation of the schema exists in its limited relationship modeling. There are no formally encoded semantic relationships between assets, systems, projects, or environmental factors yet. Dependency, adjacency, causality, and hierarchical inheritance are not explicitly modeled. This schema provides a foundation for formally modeling these relationships in the future.

Currently, there is also no prescribed method for assigning asset conditions and common damage types across all assets. As part of the future development of the schema and metadata catalog, the inspection methods used for each asset should be documented to fully define potential field entries. Consultation with subject matter experts should coincide with the determination of a controlled vocabulary associated with individual asset types. Another consideration is that as asset diversity increases, damage categories expand, and reporting

complexity grows, a damage type field may become inadequate for condition reporting. Ensuring that a relational method, akin to a universal asset identifier, exists to link more complex condition reporting is a priority.

4.4 Summary

This chapter presents the development of a proof-of-concept schema and management framework designed to support a more unified and interoperable transportation asset data environment at UDOT. The framework was implemented within Google Cloud Dataplex and focused on improving governance, interoperability, metadata management, lifecycle management, and data quality practices across selected transportation asset datasets, including barriers, pavement, sign faces, and sign assemblies. A summary of the current schema limitations and the proposed schema benefits is shown in Figure 4.6.

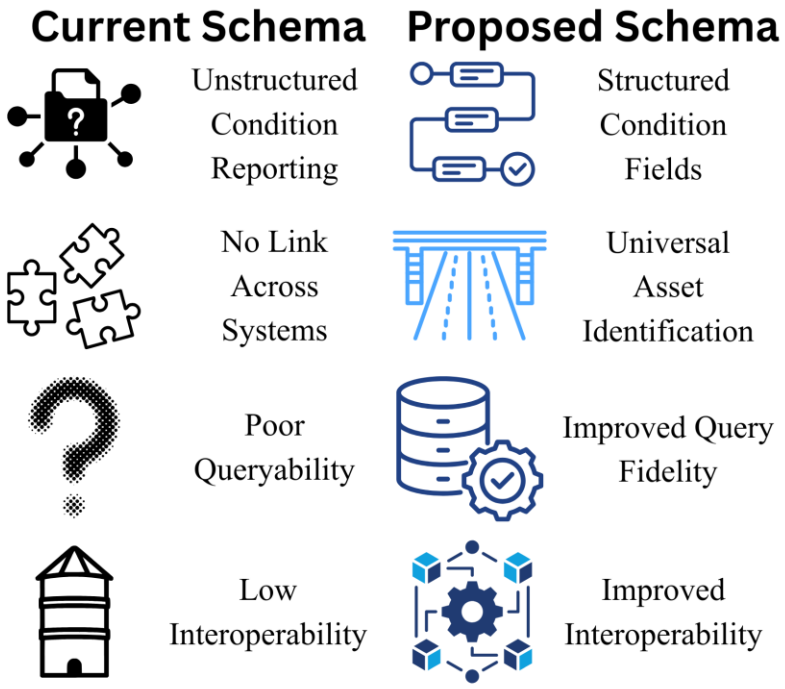


Figure 4.6 Summary of the current schema limitations and benefits of the proposed schema.

The study identified several limitations within the existing data landscape, including inconsistent governance structures, fragmented stewardship responsibilities, incomplete metadata, lack of

standardized terminology, and unstructured condition reporting. To address these issues, the proposed framework incorporates business glossaries, metadata catalogs, aspect types, and aspects within Dataplex to create a more structured and semantically aligned management environment.

Schema development followed an iterative process informed by stakeholder feedback, reviewed literature, and enterprise data management principles. The resulting schema organized data into governance, geolocation, collection, properties, and condition categories while introducing new structured fields such as asset status, asset condition, and damage type to replace reliance on unstructured comment fields for condition reporting. These modifications improved standardization, interoperability, querying capabilities, and support for future analytical applications.

Validation and testing within Dataplex demonstrated that the schema categories, aspect types, aspects, and glossary entries could be operationalized within a metadata catalog environment while also supporting automated data quality rules and governance processes. The chapter concludes by discussing limitations and future development opportunities, emphasizing the importance of universal asset identification, stronger semantic relationship modeling, standardized condition vocabularies, and expanded lifecycle management capabilities to advance the enterprise transportation asset management system.

5.0 RECOMMENDATIONS AND IMPLEMENTATION

5.1 Recommendations

Based on the findings of the gap analysis, stakeholder engagement activities, and the development of the proof-of-concept ontology framework, the following recommendations are proposed:

1. **Establish a Formal Data Governance Structure:** Develop and implement a governance framework that clearly defines data ownership, decision-making authority, approval processes, conflict resolution procedures, and accountability mechanisms. Governance policies should be documented, accessible, and consistently applied across organizational groups.

2. **Designate Asset Stewards and Asset Data Stewards:** Assign asset stewards and asset data stewards to each managed asset class. Asset stewards should be responsible for the long-term management and operational performance of assets, while asset data stewards should oversee data collection, management, governance, quality assurance, and reporting activities.

4. **Develop and Enforce Enterprise Data Standards:** Establish organization-wide standards for naming conventions, controlled vocabularies, metadata requirements, condition assessment methodologies, damage reporting practices, and data validation procedures. These standards should be maintained through governance processes and periodically reviewed for consistency and relevance.

5. **Implement a Universal Asset Identification System:** Develop and deploy a persistent asset identification methodology that uniquely identifies assets throughout their lifecycle. The asset identifier should be incorporated into all asset management systems and datasets to support interoperability, lifecycle tracking, data integration, and data quality improvement efforts.

6. **Standardize Segmentation Methodologies for Linear Assets:** Develop consistent segmentation rules for linear assets, including barriers, guardrails, and pavement.

Standardized asset boundaries, transition criteria, and stewardship responsibilities should be established to improve lifecycle management, maintenance planning, performance evaluation, and cross-system integration.

7. **Expand the Dataplex Metadata Catalog and Business Glossary:** Continue development of the Dataplex metadata catalog and business glossary by documenting dataset ownership, stewardship assignments, update procedures, attribute definitions, business rules, and metadata standards. These resources should also be maintained in an external repository to ensure continuity during future platform transitions.

8. **Implement the Proposed Schema Across Additional Asset Classes:** Expand the proof-of-concept schema beyond the initial asset classes evaluated in this study. This includes descriptive, standardized fields for damage reporting, asset status, and asset condition. Additional transportation assets should be reviewed, mapped to the proposed framework, and incorporated into a standardized enterprise schema to improve consistency and interoperability.

9. **Develop an Enterprise Transportation Ontology:** Build upon the proof-of-concept framework by formally defining transportation asset classes, relationships, business concepts, and stewardship structures within an enterprise ontology. Particular emphasis should be placed on representing relationships between assets, inspections, maintenance activities, projects, organizational units, and geographic features.

10. **Validate the Framework Prior to Enterprise Deployment:** Conduct pilot implementations on selected asset datasets to evaluate governance procedures, metadata management practices, schema effectiveness, asset identification methods, and integration workflows. Lessons learned should be used to refine standards and implementation procedures before broader deployment.

11. **Establish Ongoing Training and Continuous Improvement Processes:** Develop training programs, documentation, and periodic review processes to support the adoption of governance practices, metadata standards, and ontology management procedures.

Continuous evaluation should be used to monitor data quality, stewardship effectiveness, and framework performance over time.

Figure 5.1 summarizes the data benefits associated with these recommendations.

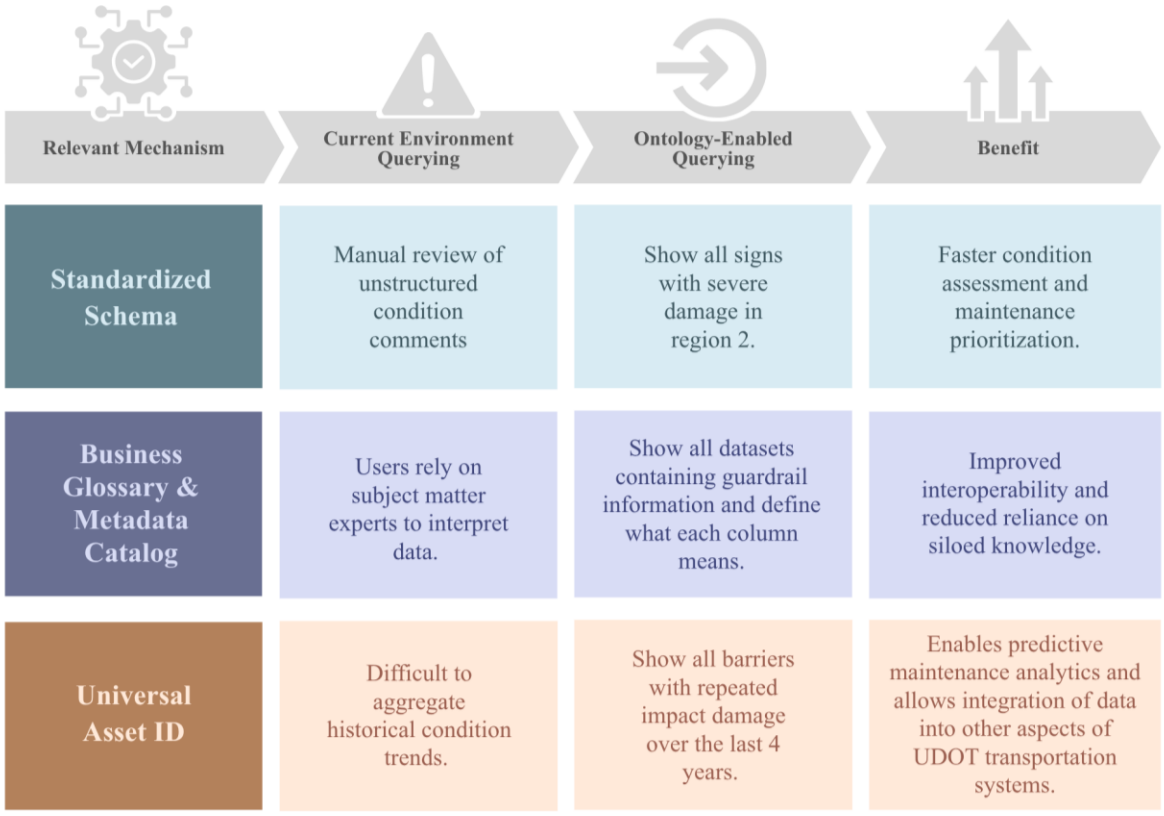


Figure 5.1 The current state of data usage, hypothetical ontology-enabled queries, and the benefit of the ontology as opposed to the current environment.

5.2 Implementation Plan

The first implementation priority is establishing a formal governance structure. This includes assigning asset stewards and asset data stewards, defining the responsibilities of each role, and creating approval, conflict resolution, and escalation procedures. Because governance serves as the foundation for long-term sustainability, clearly defined accountability and decision-making structures are essential. Once governance is established, organization-, group-, and asset-

specific standards should be developed. These standards should include naming conventions, controlled vocabularies, metadata requirements, condition assessment methodologies, damage reporting procedures, and data validation processes. Standardized language and assessment methods improve interoperability and support more advanced analytical capabilities while allowing stewards flexibility to adapt asset-specific schema elements where necessary. A universal asset identifier should be developed and applied across managed assets to support lifecycle tracking, interoperability, and data integration. Concurrently, standardized segmentation practices should be established for linear assets such as barriers, guardrails, and pavement to ensure consistent asset boundaries and improve maintenance planning, performance evaluation, and governance assignment.

Following the establishment of governance, standards, and asset identification methods, a Dataplex business glossary and metadata catalog should be developed. These tools provide centralized access to definitions, metadata, stewardship information, and business context, improving discoverability, consistency, and data quality management across organizational groups. The proposed schema and management framework should be piloted on a selected dataset of assets. This phase should evaluate governance processes, metadata management procedures, asset identification methods, and integration workflows while providing opportunities to refine standards and implementation approaches.

Following successful pilot implementation, integration efforts can expand to additional asset systems and business domains. Lessons learned during the pilot should inform broader deployment activities and support the continued development of semantic relationships between assets, processes, and organizational concepts. Beyond system integration, future implementation efforts should focus on expanding the ontology by formalizing asset classes, relationships, and business concepts across transportation domains. This process would enable the development of a transportation knowledge graph that links assets, inspections, maintenance activities, projects, and organizational entities through shared semantic relationships. Such an approach would further improve interoperability, support advanced querying and reasoning capabilities, and provide a foundation for future digital twin and decision-support applications.

Long-term implementation should focus on expanding governance practices, metadata management, interoperability mechanisms, and ontology coverage across the enterprise. Continuous monitoring and refinement of standards, stewardship practices, and data quality processes will support a scalable and sustainable asset management environment.

Several limitations should be considered when interpreting the findings and recommendations of this study. First, the ontology framework, schema, and management structure were developed as a proof-of-concept and validated using a limited set of transportation safety assets, including barriers, pavement, sign faces, and sign assemblies. While these assets were selected to represent a range of data structures and management requirements, they do not encompass the full diversity of transportation assets managed by UDOT. Additional asset classes may require modifications to the proposed framework, schema, and ontological relationships. Furthermore, the project focused on establishing a foundational ontology framework rather than developing a complete enterprise ontology, leaving many semantic relationships, business concepts, and cross-domain interactions as opportunities for future development.

The proposed framework was not implemented at an enterprise scale, limiting the ability to evaluate long-term performance, organizational adoption, and quantitative improvements in data quality, interoperability, or decision-making capabilities. The findings were informed by stakeholder interviews, survey responses, and analysis of available datasets, which may not fully represent all organizational perspectives or data management practices. Existing data quality issues, incomplete documentation, and inconsistent metadata further constrained portions of the assessment. As a result, the recommendations presented in this study should be viewed as a foundational step toward enterprise ontology development and data modernization rather than a fully validated organizational solution. Future implementation and expansion efforts will be necessary to evaluate scalability, refine governance processes, and extend the ontology across additional transportation asset domains.

The successful implementation of the proposed framework would require organizational, technical, and governance efforts to be aligned through a phased, iterative process. While the schema and management framework proposed in this study represent a proof of concept, they establish a foundational architecture capable of supporting broader data modernization efforts.

Continued expansion of governance policy, metadata management, asset identification, and interoperability mechanisms would position UDOT to develop a scalable, semantically aligned, and analytically capable asset management environment.

REFERENCES

Adams, B., & Esteghamati, M. Z. (2026). *A Semantic Modeling Approach to Transportation Asset Data*.

Al-Badi, A., Tarhini, A., & Khan, A. I. (2018). Exploring Big Data Governance Frameworks. *Procedia Computer Science*, *141*, 271–277. https://doi.org/10.1016/j.procs.2018.10.181

Bakhtiari, P., Pandey, S. R., & Esteghamati, M. Z. (2026). *A Context-Aware Computer Vision-Based Approach for Post-Earthquake Damage Assessment of Bridge Inventories*.

Bakry, I., Elsawah, H., & Moselhi, O. (2016). Multi-Tiered Database Schema for Integrated Municipal Asset Management. *Construction Research Congress 2016*, 1567–1576. https://doi.org/10.1061/9780784479827.157

Benfeldt, O., Persson, J. S., & Madsen, S. (2020). Data Governance as a Collective Action Problem. *Information Systems Frontiers*, *22*(2), 299–313. https://doi.org/10.1007/s10796-019-09923-z

Bernardo, B. M. V., Mamede, H. S., Barroso, J. M. P., & Dos Santos, V. M. P. D. (2024). Data governance & quality management—Innovation and breakthroughs across different fields. *Journal of Innovation & Knowledge*, *9*(4), 100598. https://doi.org/10.1016/j.jik.2024.100598

Boadi, R., Bruce, C., National Cooperative Highway Research Program, Transportation Research Board, & National Academies of Sciences, Engineering, and Medicine. (2026). *Data Ontologies for Data-Driven Decision-Making: Development and Use* (p. 29373). National Academies Press. https://doi.org/10.17226/29373

Brous, P., Herder, P., & Janssen, M. (2016). Governing Asset Management Data Infrastructures. *Procedia Computer Science*, *95*, 303–310. https://doi.org/10.1016/j.procs.2016.09.339

Bureau, U. C. (2018, December 19). *Census News Release*. Census.Gov. https://www.census.gov/newsroom/press-releases/2018/estimates-national-state.html

Dataplex Universal Catalog. (2026). Google Cloud. https://cloud.google.com/dataplex

Du, H., Wei, L., Dimitrova, V., Magee, D., Clarke, B., Collins, R., Entwisle, D., Eskandari Torbaghan, M., Curioni, G., Stirling, R., Reeves, H., & Cohn, A. G. (2023a). City infrastructure ontologies. *Computers, Environment and Urban Systems*, *104*, 101991. https://doi.org/10.1016/j.compenvurbsys.2023.101991

Du, H., Wei, L., Dimitrova, V., Magee, D., Clarke, B., Collins, R., Entwisle, D., Eskandari Torbaghan, M., Curioni, G., Stirling, R., Reeves, H., & Cohn, A. G. (2023b). City infrastructure ontologies. *Computers, Environment and Urban Systems*, *104*, 101991. https://doi.org/10.1016/j.compenvurbsys.2023.101991

Esteghamati, M. Z., Ahmed, M. H., Young, S., & Halling, M. (2024). A multi-task deep learning-based framework for structural details and corrosion detection in steel bridges. In J. S. Jensen, D. M. Frangopol, & J. W. Schmidt, *Bridge Maintenance, Safety, Management, Digitalization and Sustainability* (1st ed., pp. 3530–3537). CRC Press. https://doi.org/10.1201/9781003483755-417

Esteghamati, M. Z., & Flint, M. M. (2021). Developing data-driven surrogate models for holistic performance-based assessment of mid-rise RC frame buildings at early design. *Engineering Structures*, *245*, 112971. https://doi.org/10.1016/j.engstruct.2021.112971

FAIR Principles. (2016). *GO FAIR*. https://www.go-fair.org/fair-principles/

Farghaly, K., Soman, R. K., & Zhou, S. A. (2023). The evolution of ontology in AEC: A two-decade synthesis, application domains, and future directions. *Journal of Industrial Information Integration*, *36*, 100519. https://doi.org/10.1016/j.jii.2023.100519

Fernandez, M., Gomez-Pearez, A., & Juristo, N. (1997). *Methontology: From Ontological Art Towards Ontological Engineering*.

Fernandez, S., Hadfi, R., Ito, T., Marsa-Maestre, I., & Velasco, J. (2016). Ontology-Based Architecture for Intelligent Transportation Systems Using a Traffic Sensor Network. *Sensors*, *16*(8), 1287. https://doi.org/10.3390/s16081287

Firdaus Law, N. L. L. M., Mahmoud, M. A., Tang, A. Y., Lim, F.-C., Kasim, H., Othman, M., & Yong, C. (2019). A Review of Ontology Development Aspects. *International Journal of Advanced Computer Science and Applications*, *10*(7).

Foxxe, R., Dean, P., Gochnour, N., DiCaro, S., Brandley, A., Crabb, B., Cropper, E., Eisenbacher, T., Gilbert, J., Larsen, J., Lloyd, N., Looney, A., Maloney, T., Mayne, C., Mellott, D., Oritt, M., Parker, M., Pearson, M., Roney, N., … Moss, J. (2026). *2026 Economic Report to the Governor*.

France-Mensah, J., & O'Brien, W. J. (2019). A shared ontology for integrated highway planning. *Advanced Engineering Informatics*, *41*, 100929. https://doi.org/10.1016/j.aei.2019.100929

Freitas, N., Rocha, A. D., & Barata, J. (2026). Data management in industry: Concepts, systematic review and future directions. *Journal of Intelligent Manufacturing*, *37*(2), 799–827. https://doi.org/10.1007/s10845-025-02570-z

Gujjala, P. K. R. (2024). Optimizing ETL Pipelines with Delta Lake and Medallion Architecture: A Scalable Approach for Large-Scale Data. *International Journal for Multidisciplinary Research (IJFMR)*. https://doi.org/10.2139/ssrn.5054029

Hai, R., Koutras, C., Quix, C., & Jarke, M. (2023). Data Lakes: A Survey of Functions and Systems. *IEEE Transactions on Knowledge and Data Engineering*, *35*(12), 12571–12590. https://doi.org/10.1109/TKDE.2023.3270101

Harby, A. A., & Zulkernine, F. (2025). Data Lakehouse: A survey and experimental study. *Information Systems*, *127*, 102460. https://doi.org/10.1016/j.is.2024.102460

Infrastructure Report Card | utah section. (2025, May 29). https://sections.asce.org/utah-section/infrastructure-report-card

Janssen, M., Brous, P., Estevez, E., Barbosa, L. S., & Janowski, T. (2020). Data governance: Organizing data for trustworthy Artificial Intelligence. *Government Information Quarterly*, *37*(3), 101493. https://doi.org/10.1016/j.giq.2020.101493

Katsumi, M., Huang, T., & Fox, M. S. (2022). *Toward Requirements for an Ontology of Asset Management*.

Labadie, C., Eurich, M., & Legner, C. (2020). Data democratization in practice: Fostering data usage with data catalogs. *Proceedings of the 20th Symposium of the Association Information et Management*.

Labadie, C., Legner, C., Eurich, M., & Fadler, M. (2020). FAIR Enough? Enhancing the Usage of Enterprise Data with Data Catalogs. *2020 IEEE 22nd Conference on Business Informatics (CBI)*, 201–210. https://doi.org/10.1109/CBI49978.2020.00029

Le, T., Le, C., Jeong, H. D., & Jahren, C. (2018). Visual Exploration of Large Transportation Asset Data using Ontology-Based Heat Tree. *International Journal of Transportation*, *6*(1), 47–58. https://doi.org/10.14257/ijt.2018.6.1.04

Lei, X. (2023). *Improving data management through automatic information extraction model in ontology for road asset management*.

Lei, X., Wu, P., Zhu, J., & Wang, J. (2022). Ontology-Based Information Integration: A State-of-the-Art Review in Road Asset Management. *Archives of Computational Methods in Engineering*, *29*(5), 2601–2619. https://doi.org/10.1007/s11831-021-09668-6

Mohna, H. A., Barua, T., Manager, Facilities and Administration, MetLife, Bangladesh, Mohiuddin, M., Data Engineer, NCC Bank PLC, Dhaka, Bangladesh, Rahman, M. M., & Assistant Manager, Teletalk Bangladesh Ltd, Dhaka, Bangladesh. (2022). AI-READY DATA ENGINEERING PIPELINES: A REVIEW OF MEDALLION ARCHITECTURE AND CLOUD-BASED INTEGRATION MODELS. *American Journal of Scholarly Research and Innovation*, *01*(01), 319–350. https://doi.org/10.63125/51kxtf08

Nambiar, A., & Mundra, D. (2022). An Overview of Data Warehouse and Data Lake in Modern Enterprise Data Management. *Big Data and Cognitive Computing*, *6*(4), 132. https://doi.org/10.3390/bdcc6040132

Nandini, D., & Shahi, G. K. (2018). *An Ontology for Transportation System*. 32–25. https://doi.org/10.29007/qt2m

Nargesian, F., Zhu, E., Miller, R. J., Pu, K. Q., & Arocena, P. C. (2019). Data lake management: Challenges and opportunities. *Proceedings of the VLDB Endowment*, *12*(12), 1986–1989. https://doi.org/10.14778/3352063.3352116

Noy, N. F., & McGuinness, D. L. (2001). *Ontology Development 101: A Guide to Creating Your First Ontology*.

Office of Management and Budget. (2020, July). *Federal Data Strategy Data Governance Playbook*.

Öhgren, A. (2004). *Ontology Development and Evolution: Selected Approaches for Small-Scale Application Contexts*. Jönköping University, School of Engineering.

Otto, B., Grafen, D., Krammer, L., Groger, C., & Lutsch, A. (2024). *Enterprise Data Management*.

Pamula, V. K. (2026). The Medallion Architecture in Practice: A Framework for Building Scalable and Governed Data Lakehouses on Microsoft Fabric. *2026 IEEE 5th International Conference on AI in Cybersecurity (ICAIC)*, 1–5. https://doi.org/10.1109/ICAIC67076.2026.11395677

Pandey, S. R., & Esteghamati, M. Z. (2026). *Benchmarking Computer Vision-Based Approaches To Derive Engineering-Oriented Condition From Existing UDOT Assets Data*. Utah Deparment of Transportation.

Pokhrel, A., Sorensen, A. D., & Esteghamati, M. Z. (2026). Evaluating the Performance of Concrete Barriers in Protecting Bridge Piers under Vehicular Impact in Space-Constrained Settings. *Journal of Performance of Constructed Facilities*, *40*(2), 04026004. https://doi.org/10.1061/JPCFEV.CFENG-5340

Pokhrel, A., Sorensen, A. D., & Zaker Esteghamati, M. (2025). A Numerical Investigation of the Performance of Damaged Concrete Barriers Under Sequential Vehicular Impacts. *Buildings*, *15*(8), 1271. https://doi.org/10.3390/buildings15081271

Rashid, S. M., McCusker, J. P., Pinheiro, P., Bax, M. P., Santos, H. O., Stingone, J. A., Das, A. K., & McGuinness, D. L. (2020). The Semantic Data Dictionary – An Approach for Describing and Annotating Data. *Data Intelligence*, *2*(4), 443–486. https://doi.org/10.1162/dint_a_00058

Ravindran, G. S. (2025). *Next-Generation Data Lakes: Innovations in Real-Time Analytics*. https://doi.org/10.32996/jcts.2025.7.5.90

Ren, G., Ding, R., & Li, H. (2019). Building an ontological knowledgebase for bridge maintenance. *Advances in Engineering Software*, *130*, 24–40. https://doi.org/10.1016/j.advengsoft.2019.02.001

Russom, B. P. (2020, May 29). *The Evolution of Data Lake Architectures*. TDWI. https://tdwi.org/articles/2020/05/29/arch-all-evolution-of-data-lake-architectures.aspx

Sawadogo, P., & Darmont, J. (2021). On data lake architectures and metadata management. *Journal of Intelligent Information Systems*, *56*(1), 97–120. https://doi.org/10.1007/s10844-020-00608-7

Shrestha, L., & Sheikh, N. J. (2022). Multiperspective Assessment of Enterprise Data Storage Systems: Literature Review. *2022 Portland International Conference on Management of Engineering and Technology (PICMET)*. https://doi.org/10.23919

Stadnicki, A., Filip Pietroń, F., & Burek, P. (2020). Towards a Modern Ontology Development Environment. *Procedia Computer Science*, *176*, 753–762. https://doi.org/10.1016/j.procs.2020.09.070

Sudre, G., Kritikos, W., Matamoros, A. B., & Graham, S. (2008). *Ontology engineering for management of data in the transportation domain*. https://rosap.ntl.bts.gov/view/dot/5651

Transportation Asset Management Plan (TAMP)—2026. (2025, June). https://experience.arcgis.com/experience/01a2d3869d6c4c6bb20d77ff9362d7e4

UDOT announces new and ongoing 2025 construction. (2025, April 16). https://connect.udot.utah.gov/2025/04/16/udot-announces-new-and-ongoing-2025-construction/

United Nations Conference Trade and Development. (2025, August 5). *Trade and Development Report 2025: Overview of data principles*. United Nations.

U.S. Department of Transportation. (2026). *United States Department of Transportation Strategic Plan FY 2026-2030*. https://www.transportation.gov/mission/dot-strategic-plan-fy2026-2030

Uschold, M., & King, M. (1995). Mike Uschold and Martin King. *Proceedings of the Workshop on Basic Ontological Issues in Knowledge Sharing*.

Wang, J., Liu, Y., Li, P., Lin, Z., Sindakis, S., & Aggarwal, S. (2024). Overview of Data Quality: Examining the Dimensions, Antecedents, and Impacts of Data Quality. *Journal of the Knowledge Economy*, *15*(1), 1159–1178. https://doi.org/10.1007/s13132-022-01096-6

Wu, D., Zheng, A., Yu, W., Cao, H., Ling, Q., Liu, J., & Zhou, D. (2025). Digital Twin Technology in Transportation Infrastructure: A Comprehensive Survey of Current Applications, Challenges, and Future Directions. *Applied Sciences*, *15*(4), 1911. https://doi.org/10.3390/app15041911

Wunder, J., Halbardier, A., & Waltermire, D. (2011). *Specification for asset identification 1.1* (NIST IR 7693; 0 ed., p. NIST IR 7693). National Institute of Standards and Technology. https://doi.org/10.6028/NIST.IR.7693

Xu, X., Yuan, C., Zhang, Y., Cai, H., Abraham, D. M., & Bowman, M. D. (2019). Ontology-Based Knowledge Management System for Digital Highway Construction Inspection. *Transportation Research Record: Journal of the Transportation Research Board*, *2673*(1), 52–65. https://doi.org/10.1177/0361198118823499

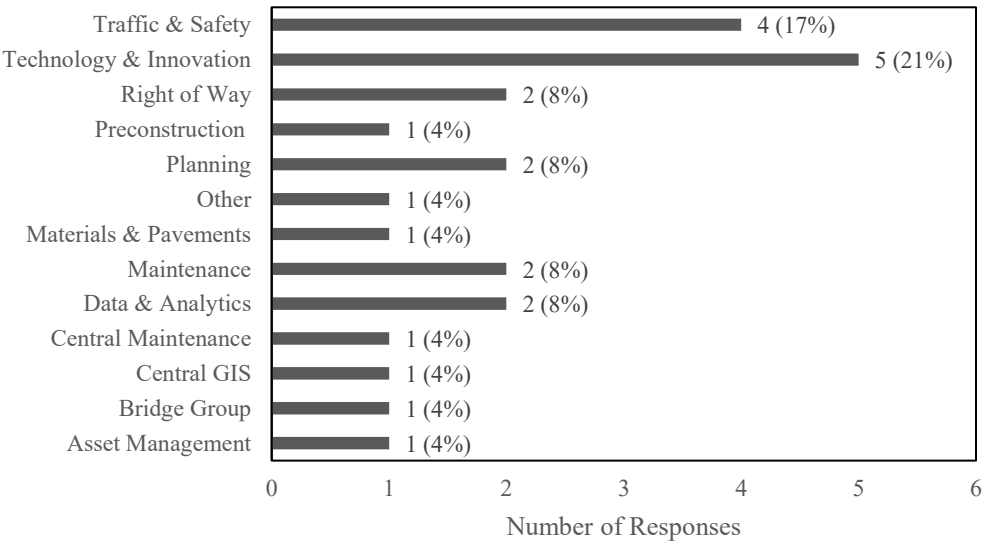
Yan, B., Yang, F., Qiu, S., Wang, J., Cai, B., Wang, S., Zaheer, Q., Wang, W., Chen, Y., & Hu, W. (2023). Digital twin in transportation infrastructure management: A systematic review. *Intelligent Transportation Infrastructure*, *2*, liad024. https://doi.org/10.1093/iti/liad024

Zaker Esteghamati, M., Bean, B., Burton, H. V., & Naser, M. Z. (2025). Beyond Development: Challenges in Deploying Machine-Learning Models for Structural Engineering Applications. *Journal of Structural Engineering*, *151*(6), 04025059. https://doi.org/10.1061/JSENDH.STENG-13301

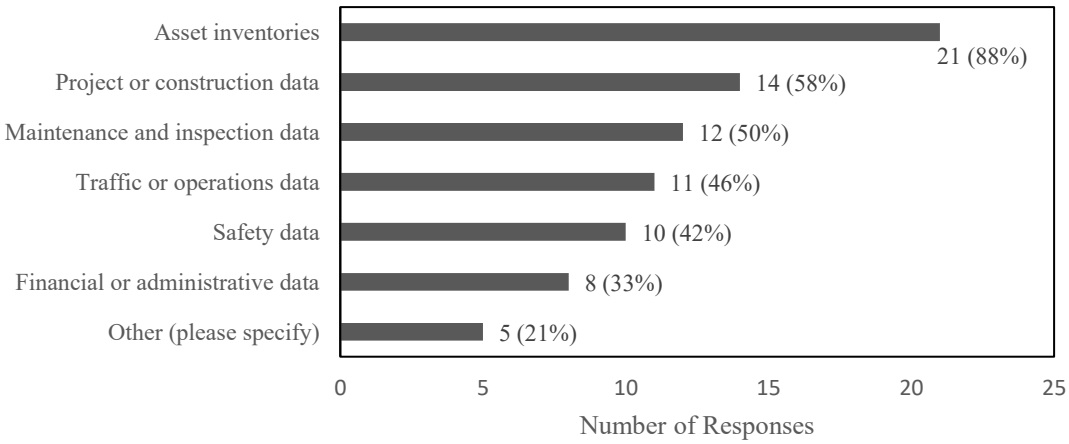
APPENDIX A: Survey Questions and Results

This appendix includes the questions and responses for all survey questions administered to Pathways data users at UDOT. This survey was conducted to gain further insight into how the data is being used, the current state of data management, and feedback for a desired future data management landscape at UDOT. 24 individuals were identified to receive the survey, and all 24 individuals responded.

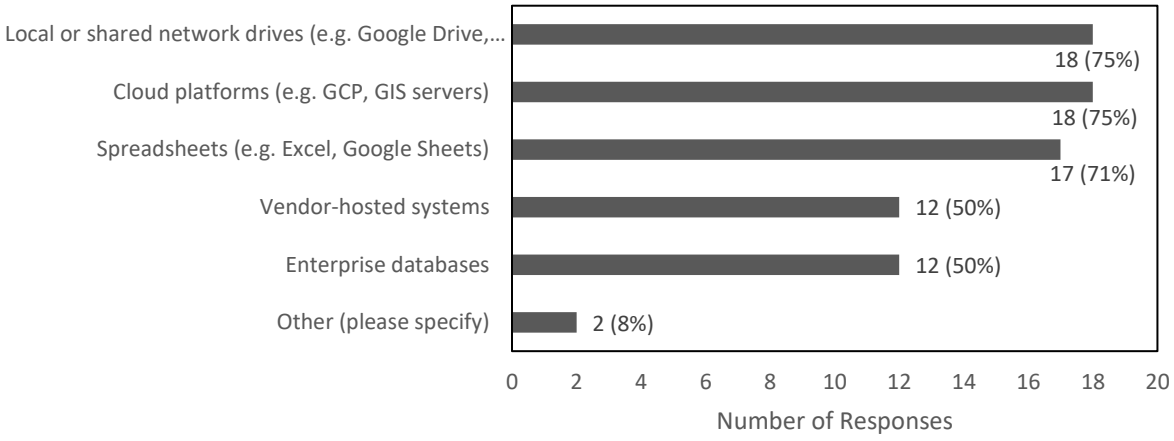
Q1. Which group are you a part of?



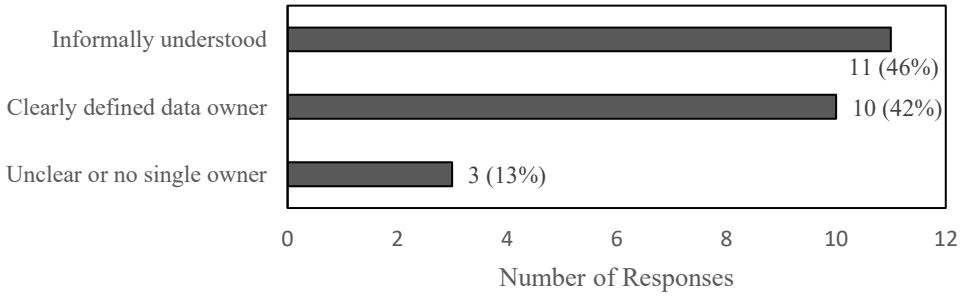
Q2 What types of data does your group manage or regularly use? (Select all that apply)



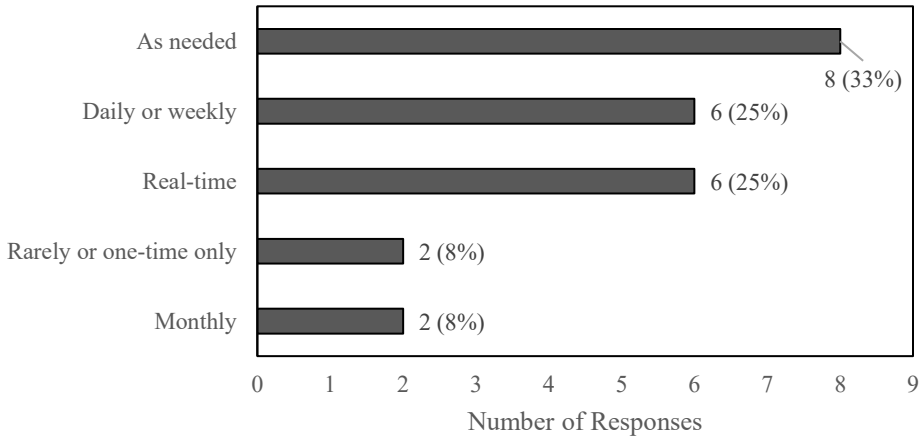
Q3 Where is this data currently stored?



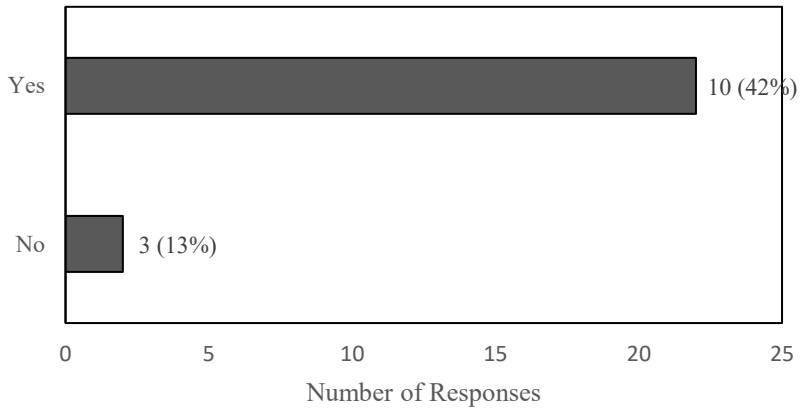
Q4 Is the data owner for your primary datasets clearly defined?



Q5 How often is your data updated?

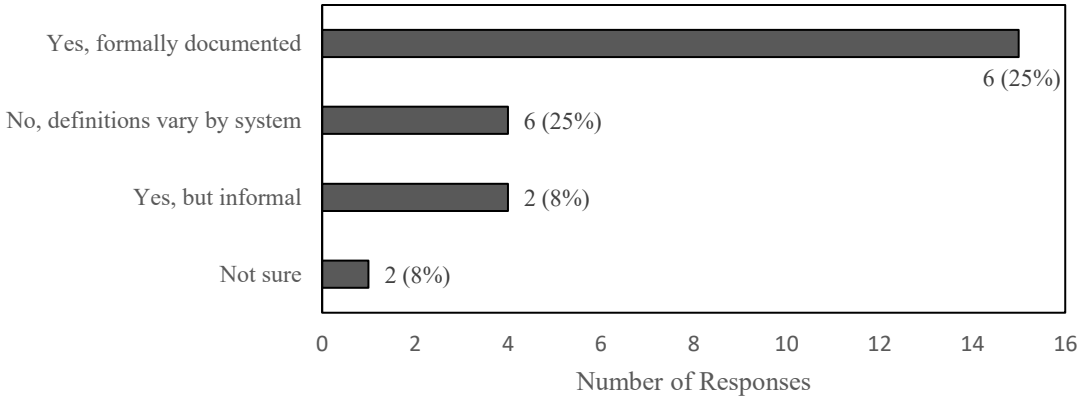


Q6 Have you used UDOT roadway images or LiDAR at https://pathweb.pathwayservices.com/ut?

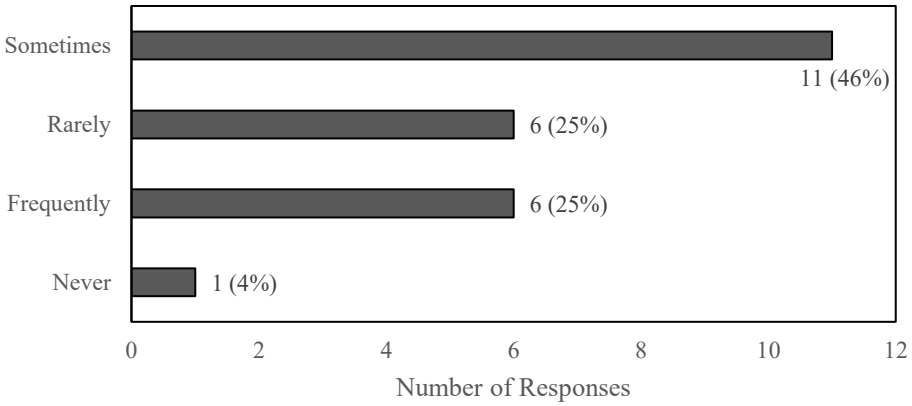


Q7 If so, how are you using it?

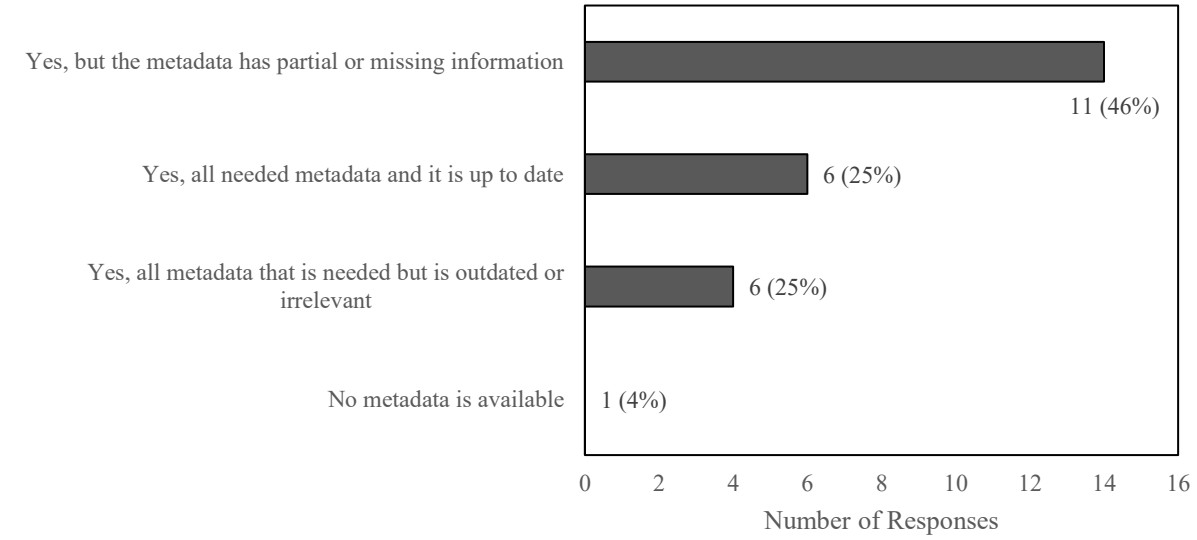
Q8 Are standard definitions used for key terms (e.g., asset type, condition, location)?



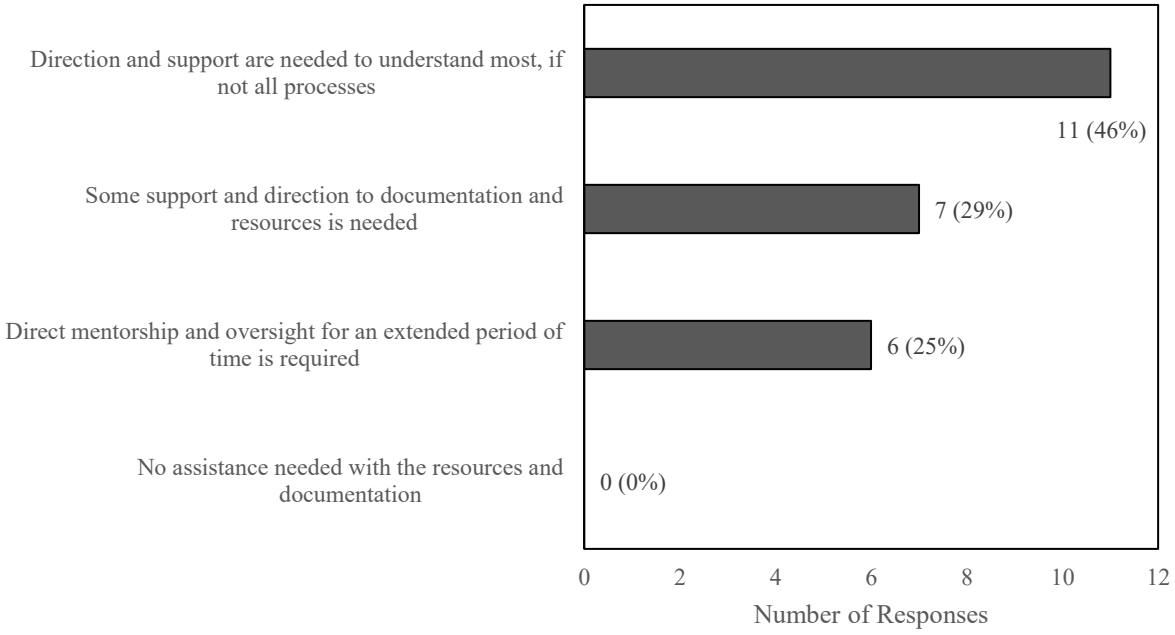
Q9 Do multiple systems use different names or definitions for the same concept?



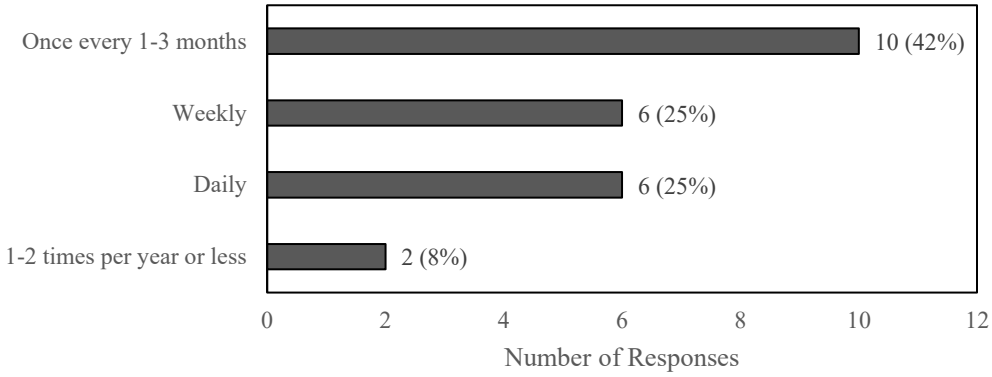
Q10 Is metadata available for your datasets? Metadata is basic information about a file or dataset. Examples include, but are not limited to, who created it, when it was created, what the fields or columns mean, the type of file it is, or how big it is.



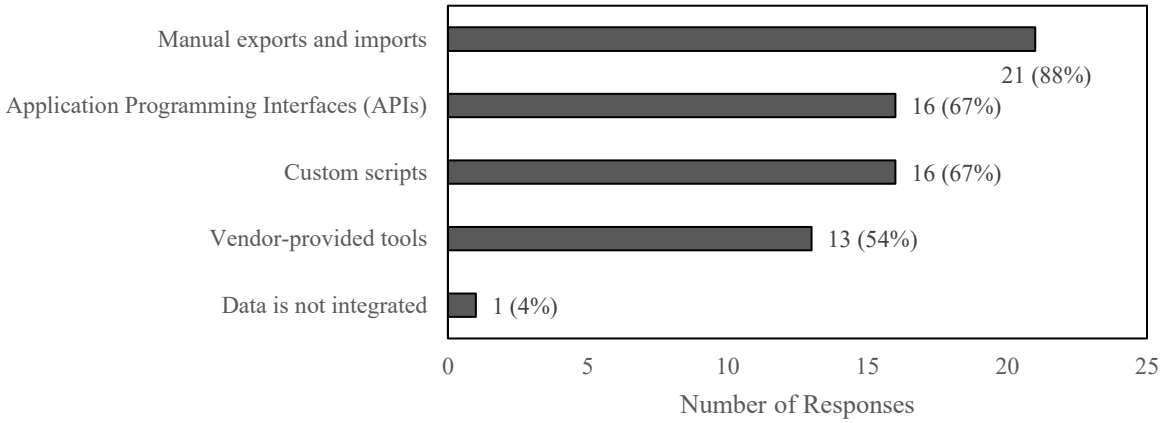
11 How difficult is it for a new staff member to understand your data without direct assistance?



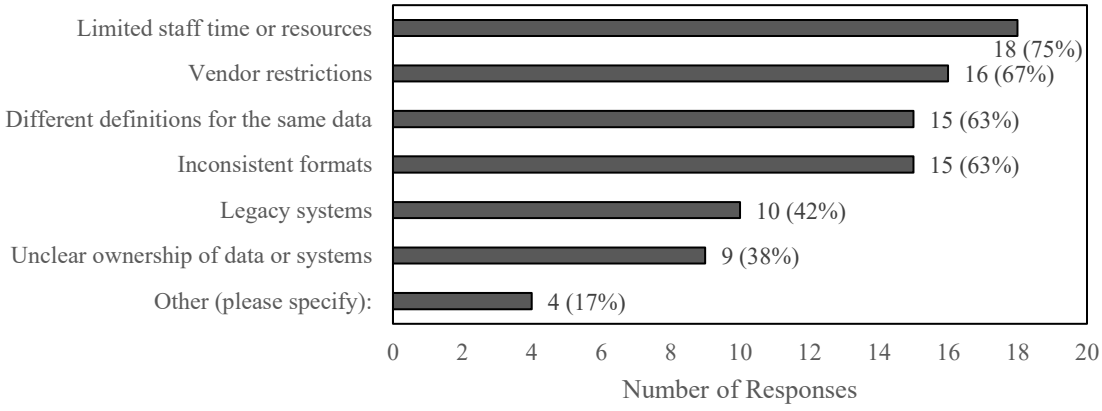
Q12 How often do you need to combine data from multiple systems (e.g., storage locations, softwares)?



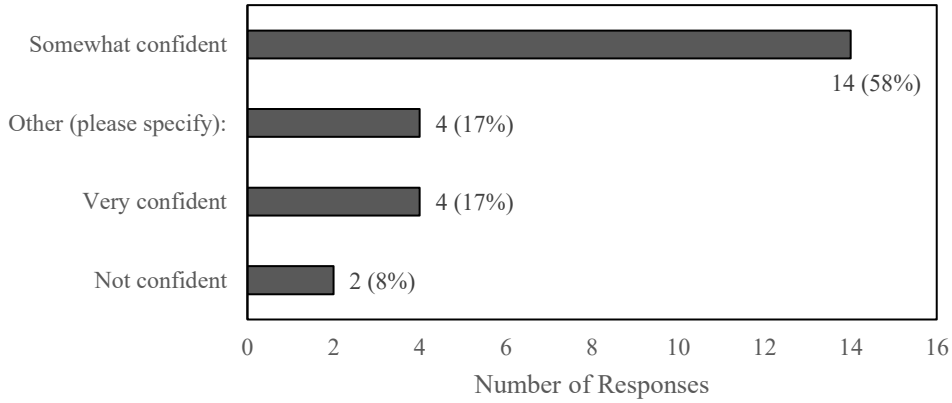
Q13 How is data typically integrated today? (Select all that apply)



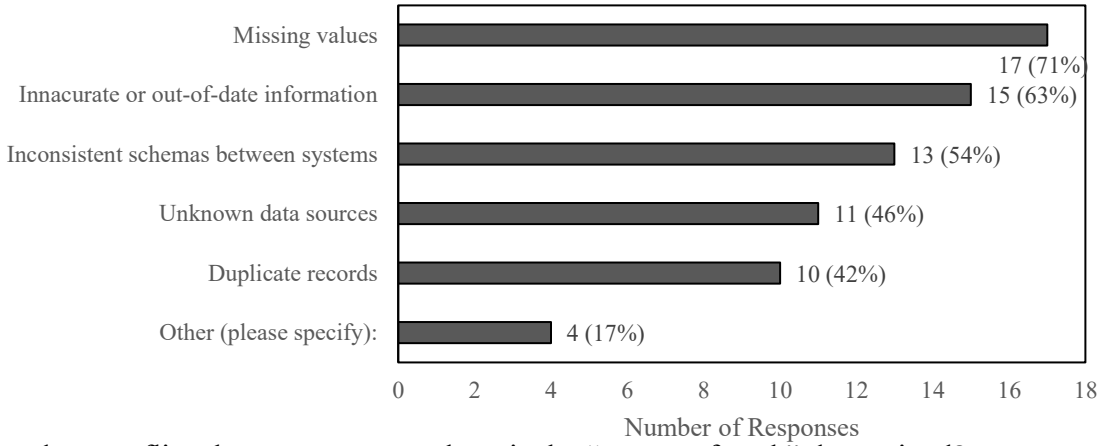
Q14 What are the biggest barriers to integrating data across systems? (Select all that apply)



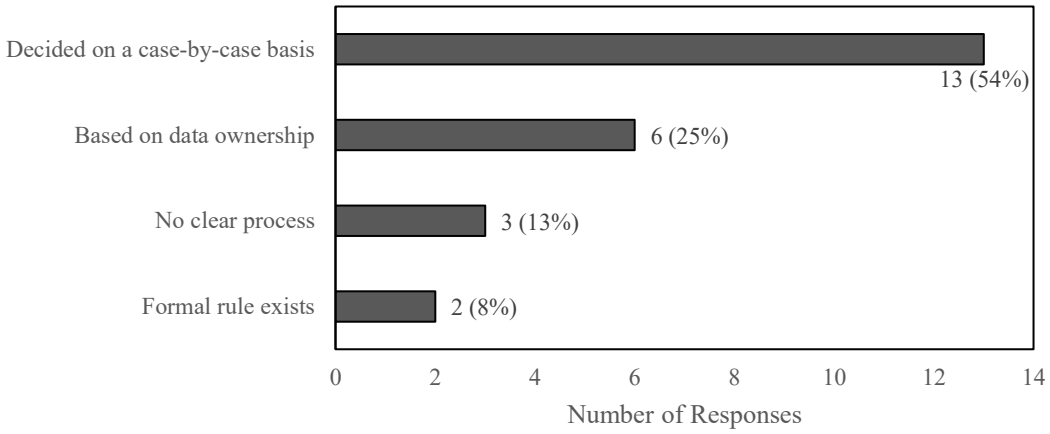
Q15 How confident are you in the accuracy of your data?



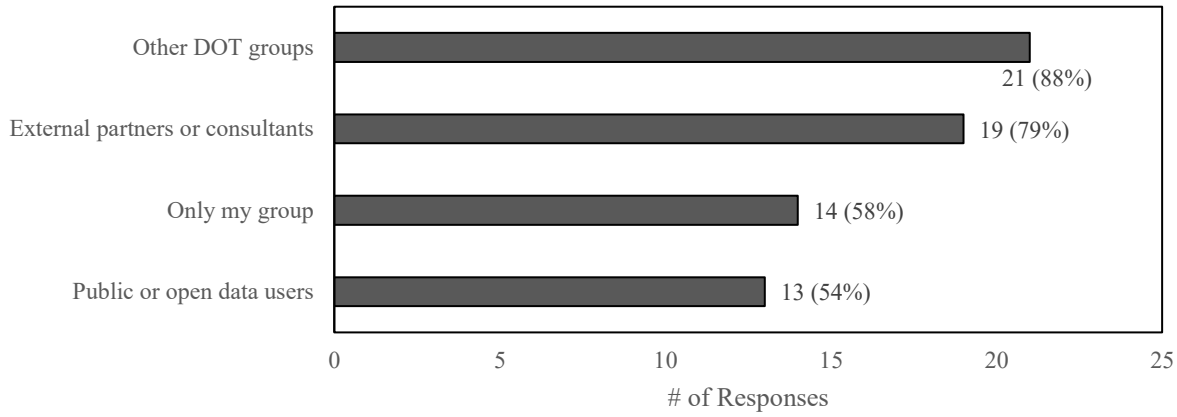
Q16 What data quality issues do you encounter most often? (Select all that apply)



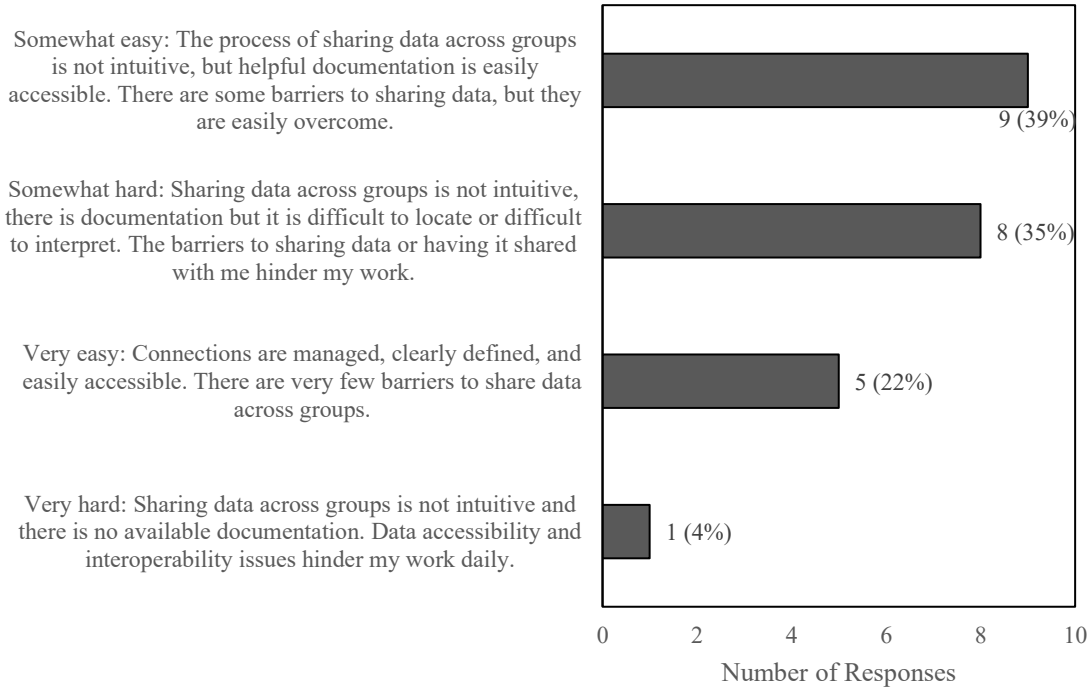
Q17 When data conflicts between systems, how is the “source of truth” determined?



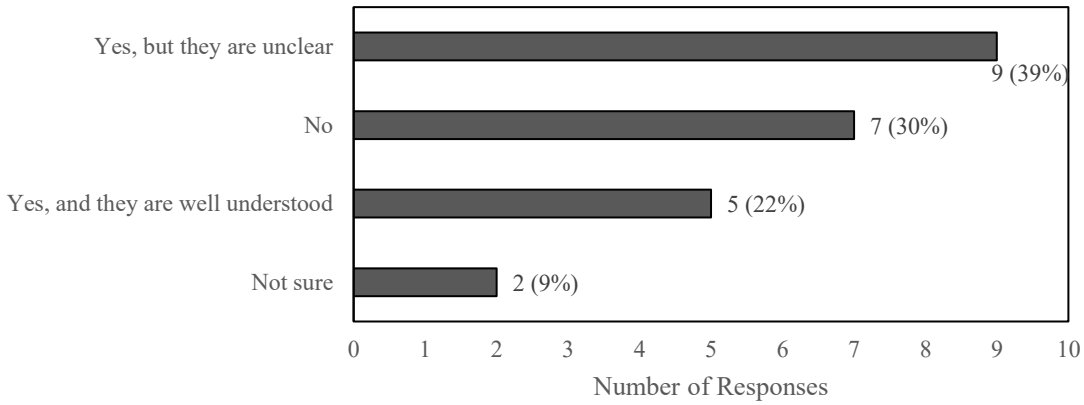
Q18 Who can access the data you are responsible for? (Select all that apply)



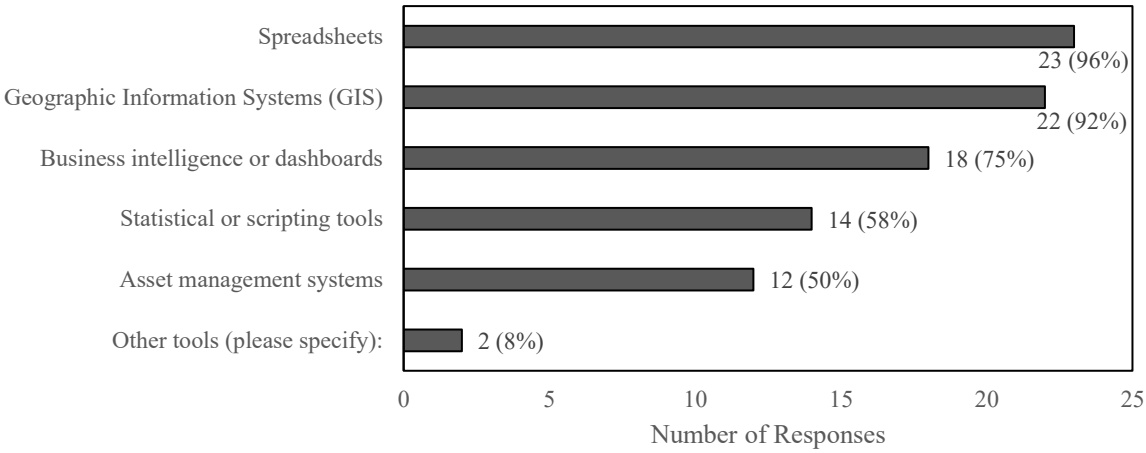
Q19 How easy is it to share data with other DOT groups?



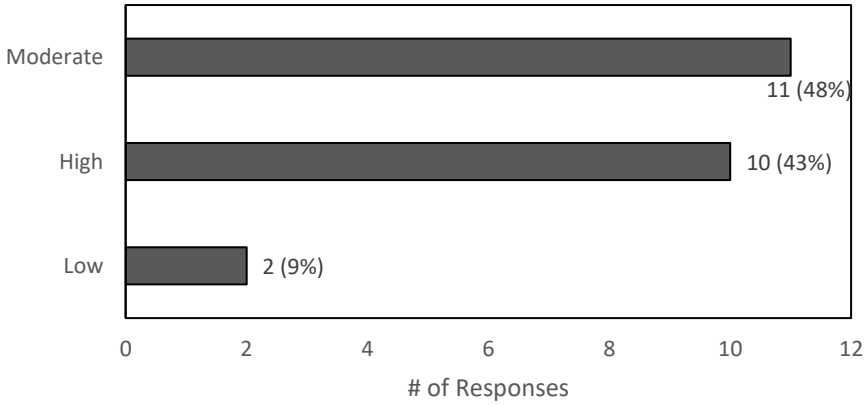
Q20 Are there formal policies governing data sharing and use?



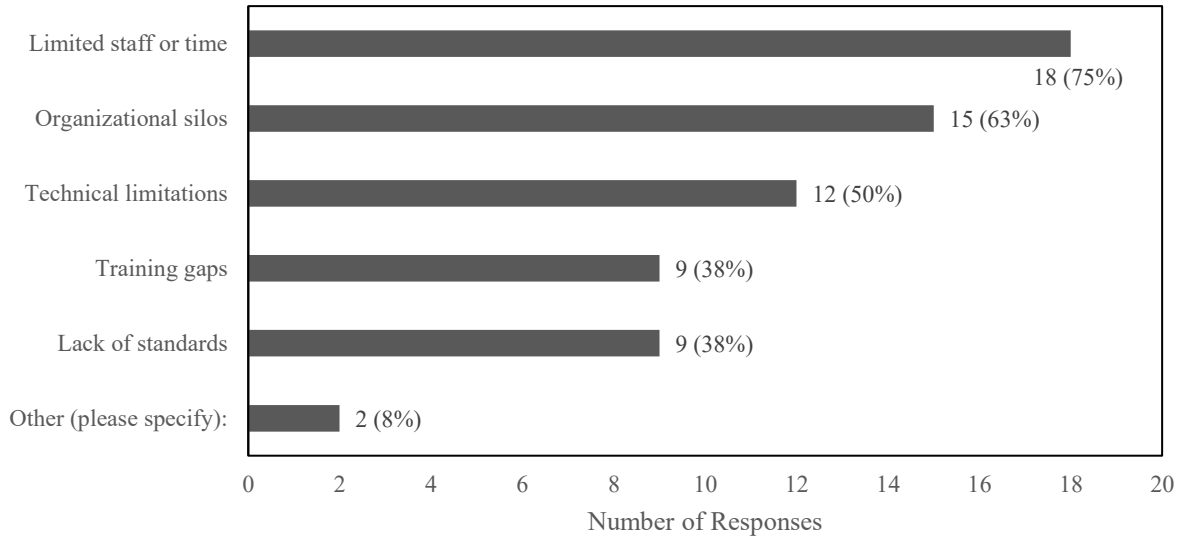
Q21 What tools do you regularly use to work with data? (Select all that apply)



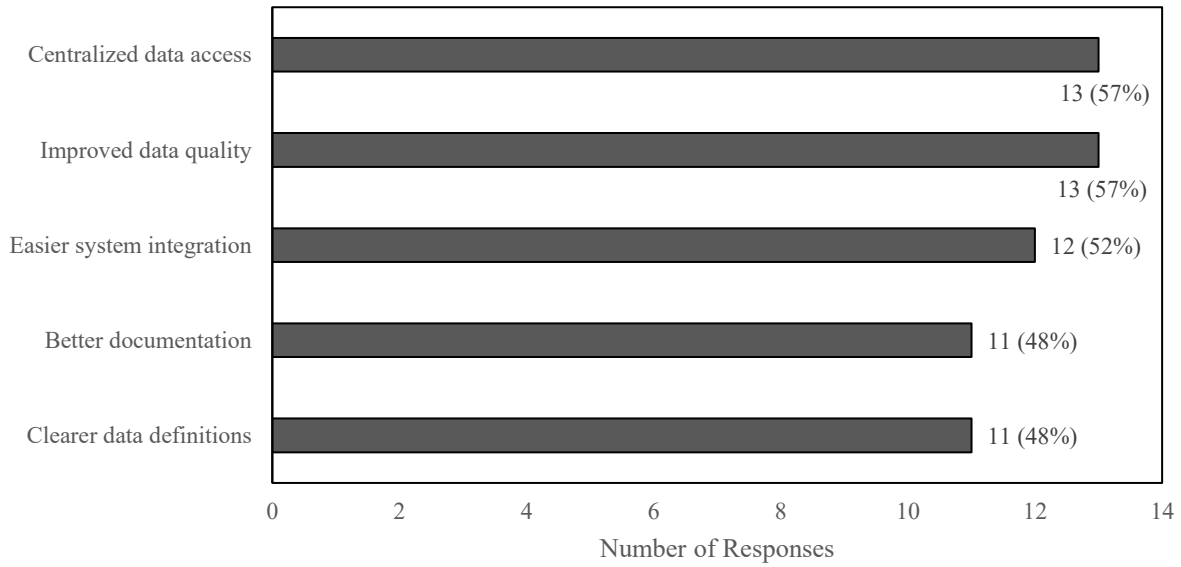
Q22 How would you rate your group's ability to read, understand, use, and communicate with data for better decision-making?



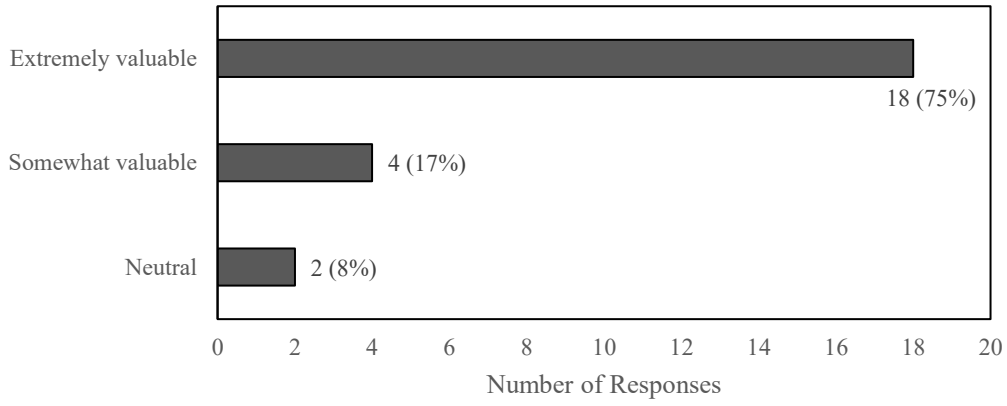
Q23 What limits your ability to better manage or use data? (Select all that apply)



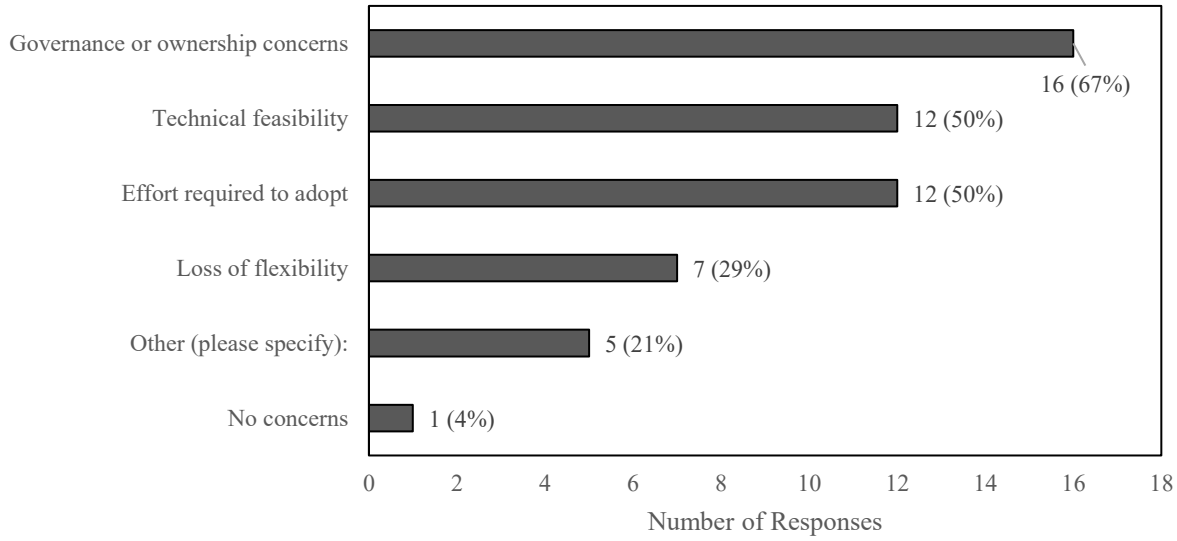
Q24 Which improvements would have the biggest positive impact on your work? (Select up to three)



Q25 How valuable would a shared, DOT-wide data vocabulary or organized and defined data structure be?



Q26 What concerns would you have about implementing a common method for organizing and structuring data? (Select all that apply)



Q27 What should a DOT-wide method for organizing and structuring data include to be most useful for your work?