

FHWA ENTERPRISE DATA ARCHITECTURE

FHWA Enterprise Data Architecture

February 2017



U.S. Department of Transportation
Federal Highway Administration

Table of Contents

- Executive Summary..... 4
- Introduction 5
 - Strategic Goals 5
 - Justification 5
 - Objectives 5
- Current Data Environment..... 6
- Transition Plan 7
 - Phase 1: Planning dataset consolidations & logical groupings 7
 - Phase 2: Planning and Acquisition Alignment with FHWA Enterprise Architecture (EA) 8
 - Phase 3: Connector, Hub, and Linked Services Delivery 8
 - Phase 4: Public Web Release 8
 - Phase 5: Non-Mission Release 8
 - Phase 6: Mission Release 8
- Cloud Technologies Selection Guide..... 9
- Target Enterprise Data Environment 10
 - ETL/ELT & Data Domain Hubs on Cloud..... 11
 - Business Analytics and Platform Services 12
 - Data Ownership 13
 - Interoperability within Target Enterprise Data Environment..... 13
 - Technologies and Standards Inventory..... 15
 - Data Architecture Security Model 15
- APPENDIX A – KEY TERMS & ACRONYMS 16

Executive Summary

The Federal Highway Administration (FHWA) is planning and developing the Target Enterprise Data Environment (T-EDE). The goals are to:

- Develop a unified ecosystem for FHWA data;
- Establish standardized interfaces for linking and processing information;
- Offer advanced capabilities such as Big Data Storage/Analytics and Business Intelligence applications as part of T-EDE common services.

The T-EDE will be designed on a cloud platform, and it will be delivered in six (6) increments closely aligned with the four (4) increments of FHWA Cloud Implementation per 2014 [FHWA Cloud Strategy document](#). This strategy is consistent with the Office of Management and Budget's (OMB's) "Cloud First" Policy¹, which requires Federal Agencies to explore the feasibility of implementing Cloud Services for all new Information Technology (IT) investments before deciding to make new investments in alternate technologies. Should FHWA discover viable options for implementing Cloud, OMB expects the agency to proceed with developing a suitable Cloud model.

¹ The "Cloud First" Policy was established in OMB's 25-Point Implementation Plan to Reform Federal Information Technology Management (December 2010) and the subsequent Federal Cloud Computing Strategy (February 2011).

Introduction

This document provides a high level depiction of the current state for Federal Highway Administration (FHWA) Data Architecture and presents the vision and a road map for constructing the FHWA Target Enterprise Data Architecture. It defines the transitional stages from current to target while aligning each phase with those specified within the [FHWA Cloud Strategy document](#). The primary audience of this document is FHWA's executive leadership, division chiefs, program managers, system planners, and solutions architects.

Strategic Goals

- FHWA will adopt cloud architecture to develop a standardized and unified data environment with a common set of internal and public interfaces which provide interoperability through automated data collection, linking, and processing.
- FHWA will have an enterprise data environment to support emerging technologies such as Big Data and Business Intelligence (BI) analytics for advanced systems to work with high volume, velocity, and throughputs of structured, semi-structured, and unstructured data.

Justification

The FHWA's data environment is comprised of tightly coupled data systems that are utilized by a close community of users and transactional systems. These data systems have often been developed and deployed in silos where they are stored and maintained by different offices, making them prone to data duplications and discrepancies. Although some of the FHWA systems currently communicate and/or exchange information with other systems, these data sources are not linked or readily discoverable as they often do not share the same structure. This ultimately results in data duplication across multiple systems without proper mapping.

FHWA should stay current with technology standards, and tailor the best practices for the administration's Target Enterprise Data Architecture. Closely aligning the target with industry's technology standards is prudent in maintaining positive control over FHWA IT resources and gaining more awareness of enterprise data architecture during planning, upgrading, migration and integration activities.

Objectives

The FHWA Enterprise Data Architecture will prepare for technology advancement by properly modeling data, and designing and allocating information exchange between systems. The overall objectives are:

- Proper categorization, inventory or cataloging of FHWA data containers;
- Identification and proper management of duplicate "overlapping" data/ data sources;
- Linking data containers into the target data environment for better business intelligence analytics;
- Maintaining legacy data environment until full transition into FHWA target data environment;
- Preparing and providing Application Program Interfaces (API's) for public consumers, and a separate set of API's for trusted data exchanges with internal sources².

² <https://www.whitehouse.gov/sites/default/files/omb/memoranda/2013/m-13-13.pdf>

- Adoption of cloud architecture in accordance with FHWA Cloud Strategy.

Current Data Environment

The current FHWA data environment is segmented into data sources by functional units. These data sets are hosted and managed by FHWA Headquarters, Federal Lands, Turner Fairbanks, Resource Centers, Federal Aid Offices, National Highway Institute and other research centers. Each unit hosts their data in one or several locations as either primitive structures or silos without common data formats. The data exchange mechanics are divided into the following four activity tiers defined below.

Bulk Data – Sets of data collected or purchased from a wide variety of providers where majority of the data is collected in either plain text or XML format. Examples of bulk data providers and consumers are Delphi, State Partners, and Federal Partners.

Data Ingestion – Several Extract Transform Load (ETL) tools are used to extract data sent by bulk data providers to perform complex data transformations then loading data into the environment. Job scheduling services are available to perform ETL loads at predetermined intervals as they are incorporated within custom applications. Bulk data providers upload/download data from managed file transfer servers. Optimized Data Ingestion is performed by specialized ETL tools to directly pull data into staging databases before performing transformation activities i.e. Extract Load Transform (ELT).

Data Processing – Data management tasks are performed to ensure quality and availability for conventional data warehousing. Structured data is typically loaded into relational databases, while unstructured data is kept at file servers or content management systems. Silo data dictionaries contain metadata catalogue views, set standards, and enforce policies. Data can be shared across multiple domains during this stage as needed.

Information Delivery – Information is provided through Extractions, Reports, Websites and other application services (e.g. limited Business Intelligence [BI] services) to a variety of consumers. The information is read-only by default, and special permissions and approval process for full data access may be granted.

Transition Plan

FHWA will take an incremental approach in transitioning into the Target Enterprise Data Environment (T-EDE), to be planned in parallel with a near-term FHWA Cloud Implementation Plan development. This approach minimizes risks since data operations will be gradually released into T-EDE. The transition increments will start with each FHWA data domain pre-planning their Cloud migration.

The transition activities will be carried out in two major classifications: the physical infrastructure, and the datasets. Physical infrastructure and related services acquisition will be considered after the development of the datasets migration plan. While the datasets are being grouped and consolidated, FHWA Cloud vendor selection process will take place, providing the baseline for proceeding into physical infrastructure acquisition planning. Figure 1 aligns activities associated with FHWA Cloud Implementation and the Target Data Architecture development phases. It also depicts the Target Data Architecture's notional timelines with respect to FHWA Cloud Implementation Plan.

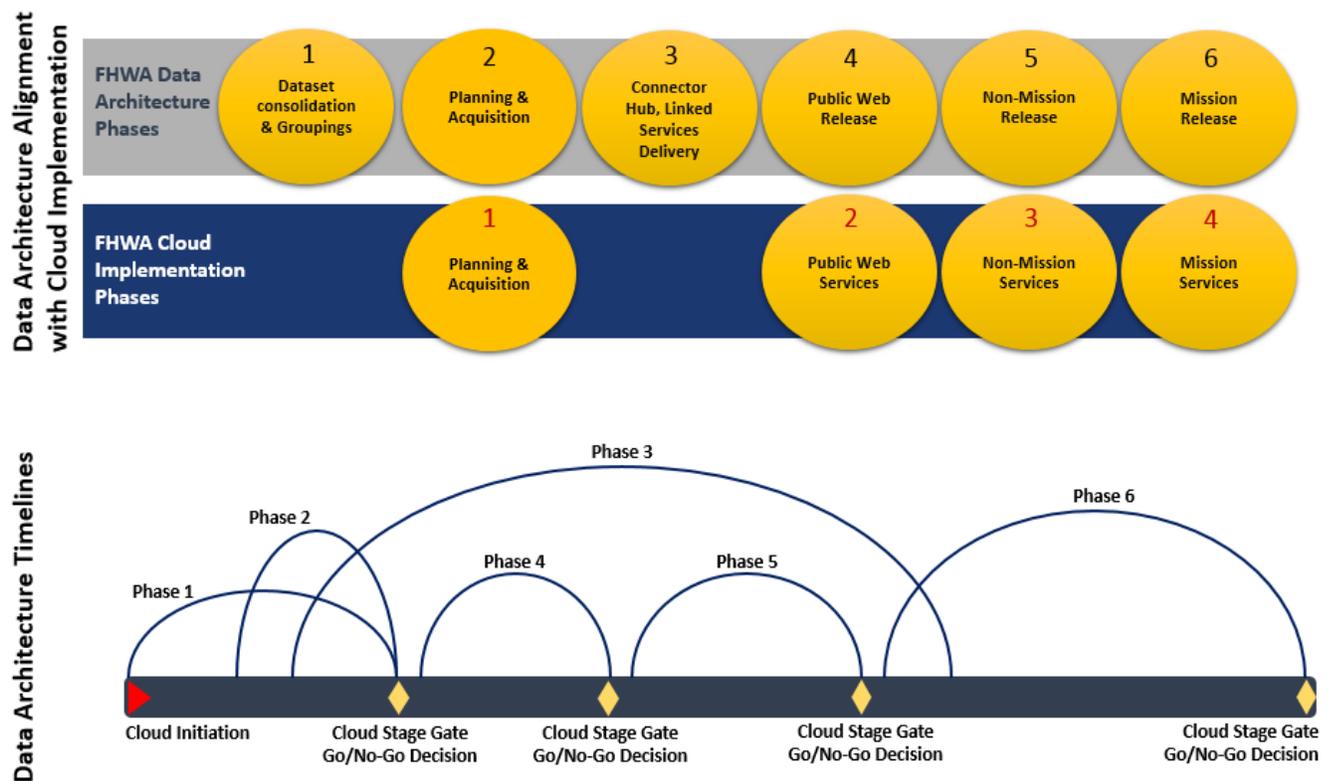


Figure 1 – Target Data Architecture Implementation Alignment & Timelines

Phase 1: Planning dataset consolidations & logical groupings

Approval of the program by the FHWA Investment Review Board (IRB) will be required prior to proceeding into the Phase 1. Data Governance Technical Advisors will assist in aligning T-EDE with FHWA Target EA, resolve any major concerns, and minimize overall risk associated with cloud-based data architecture. After completion of this Phase 1, a milestone meeting will be conducted by the Data Governance Advisory Council (DGAC) and a

decision will be made to either consider this phase as complete or not complete. In case of an incomplete decision, this phase will be reviewed to either address those concerns or take alternate approaches.

Phase 2: Planning and Acquisition Alignment with FHWA Enterprise Architecture (EA)

During this phase most dataset consolidation and logical groupings of FHWA data are assumed to have taken place, and that all data components identified with clearly revealed schemas. This phase is a joint effort connecting the development of T-EDE to FHWA Cloud Implementation's Acquisition Planning (Phase 1). Since the T-EDE is essentially a cloud environment, Phase 2 will start a Cloud Provider selection process to acquire necessary infrastructure, platforms, utilities and software components for cloud migration. This phase will be considered complete once the FHWA Cloud Implementation Phase 1 Stage Gate review has been concluded as successful, in addition to the DGAC consent to proceed into the next phase.

Phase 3: Connector, Hub, and Linked Services Delivery

This phase consists of FHWA Cloud implementation and migration of current environment into the T-EDE for public web, non-mission, and mission services. Logical groupings of working data must be designed for implementation on Cloud platform. The Target Data Environment section will further elaborate on the definition as well as implementation of those components. This phase is a continuous endeavor starting from Phases 1 and continuing to phase 6.

Phase 4: Public Web Release

During this phase, publically available data groupings are moved to T-EDE and prepared for the FHWA Cloud's Public Web Services migration. Applications and resources will migrate to the cloud environment once the data has been released. This phase will be considered complete after a successful milestone review resulting in DGAC consent and approval of the overarching IRB.

Phase 5: Non-Mission Release

Datasets and services for non-mission critical applications and services will be migrated to the FHWA Cloud platform and configured before migrating non-mission critical applications. The security configuration will follow FHWA Cloud's security standards, adding data level security to the resting, transmitting, and working data. Non-Mission Critical Services and Applications will migrate to the cloud environment once the corresponding data hubs have been properly configured and secured. This phase will be considered complete after a successful milestone review resulting in DGAC consent and approval of the overarching IRB.

Phase 6: Mission Release

This phase will migrate and configure datasets and data analytics for the mission critical applications/services over to the cloud platform. The security configuration will follow FHWA Cloud's security standards, adding data level security to the resting, transmitting, and working data. Mission Critical Services and Applications will migrate to the cloud environment once the corresponding data hubs have been properly configured and secured. This phase will be considered complete after a successful milestone review resulting in DGAC consent and approval of the overarching IRB.

Cloud Technologies Selection Guide

A Cloud Technologies Selection Guide will be prepared to serve as guidance or data migration onto a cloud platform. It will accompany the FHWA Cloud Strategy and FHWA Cloud Implementation Plan to provide the Data Stewards with a well-defined set of standards and aid them during the technology selection process. The Cloud Technologies Selection Guide will be prepared in accordance with existing policies while closely resembling the technologies used to implement the Pilot Cloud project. Data Stewards, System Owners and Technical Leads from each data domain will conduct a review of their current environment and verify major components for cloud migration. The Cloud Technologies Selection Guide will include a full listing of technical references to products, service and vendors which are approved by FHWA and DOT OST.

Target Enterprise Data Environment

The FHWA Target IT Infrastructure envisions a services-based environment where physical databases and file servers reside on a Cloud environment and have special provisioning for improved network capacity. Conversely, retention of legacy systems within the current environment will require a proper assessment of those physical IT assets and business justification. Below describes T-EDE major components, and Figure 2 depicts a holistic view of T-EDE.

Data Sources:	These are comprised of data repositories hosted in clustered or geographically dispersed locations, on different technology platforms and in varying formats. The data sources may be structured, semi-structured or unstructured depending upon the collection method or storage medium. The data sources may or may not be interconnected; the existing relationships between data entities are important as they are included as part of the enterprise data environment.
Extract & Commit:	The data extraction and commitment activities are event-driven manner where data is extracted from its source as needed. The extraction and commitment is a two-way channel marshalling raw data into designated environments for further transformation followed by loading into the operating environments.
Data Access & Transformation API:	Provides seamless data transformation from various input nodes prior to entering data zones where systems and applications are ready to consume the requested data.
Public Zone:	Working data environment available to the public or otherwise open/unrestricted data sources.
Trusted Zone:	Working data environment available to the sensitive or otherwise restricted data hubs.
Data Hubs:	Logical working segments where data is made available for applications and/or systems' consumption. Data Hubs are also used to prepare data for storage or transfer into other applications or systems.
Communications Connectors:	The ability for systems, applications, or functions to bring in or submit readily available data, in the expected format.
Business Applications & Platform Services:	These are the data consumers residing at the application tier. There are either self-contained applications loosely coupled with the working data, or rely on additional logical components or metadata in order to function properly. For example, most Business Intelligence (BI) applications require contextual information.

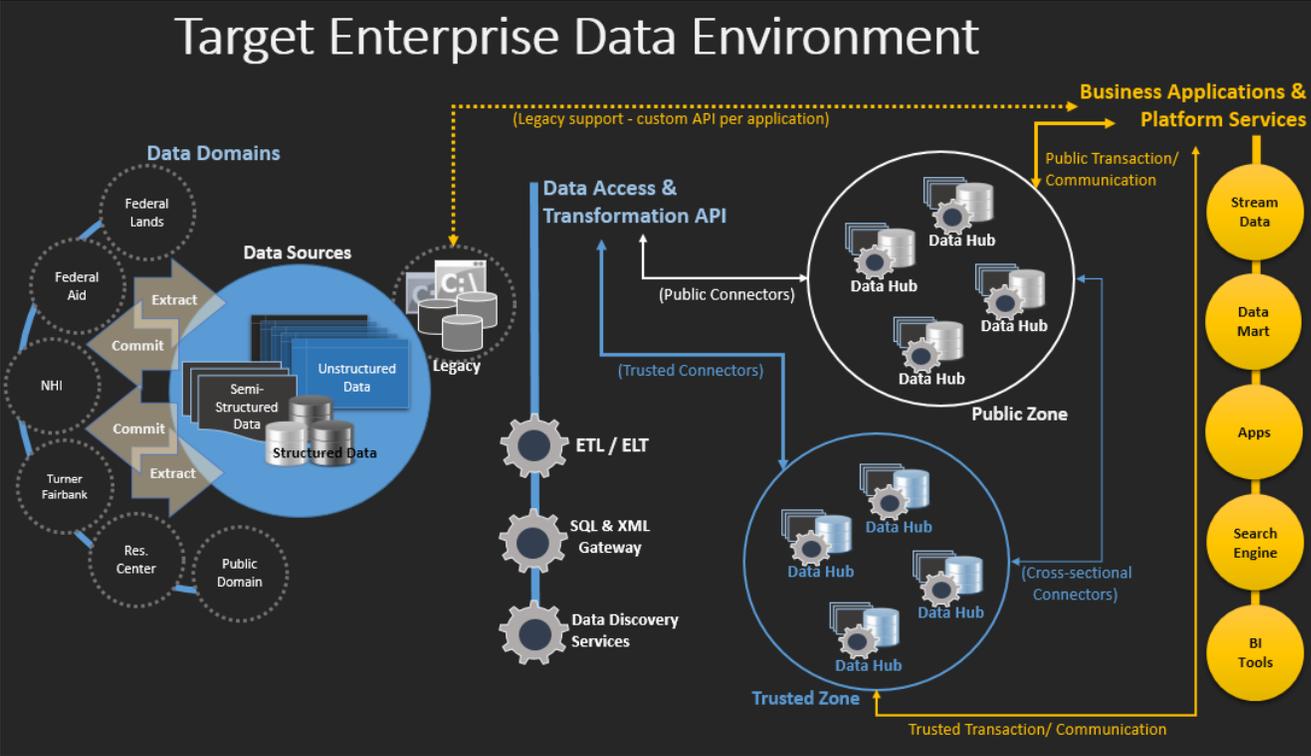


Figure 2 – FHWA Target Enterprise Data Environment

ETL/ELT & Data Domain Hubs on Cloud

To remove the complexities of data processing, the Business Analytics and Applications tier will pull data from internal and external sources via two methods: Extract Transform Load (ETL), and Extract Load Transform (ELT). These two methods are used for batch processing and in-stream/ in-memory analytics. ETL/ELT will also aid in preserving the original state and integrity of the restful-data by decoupling data storage from data preparation environment. Figure 3 demonstrates ETL/ELT on Cloud environment.

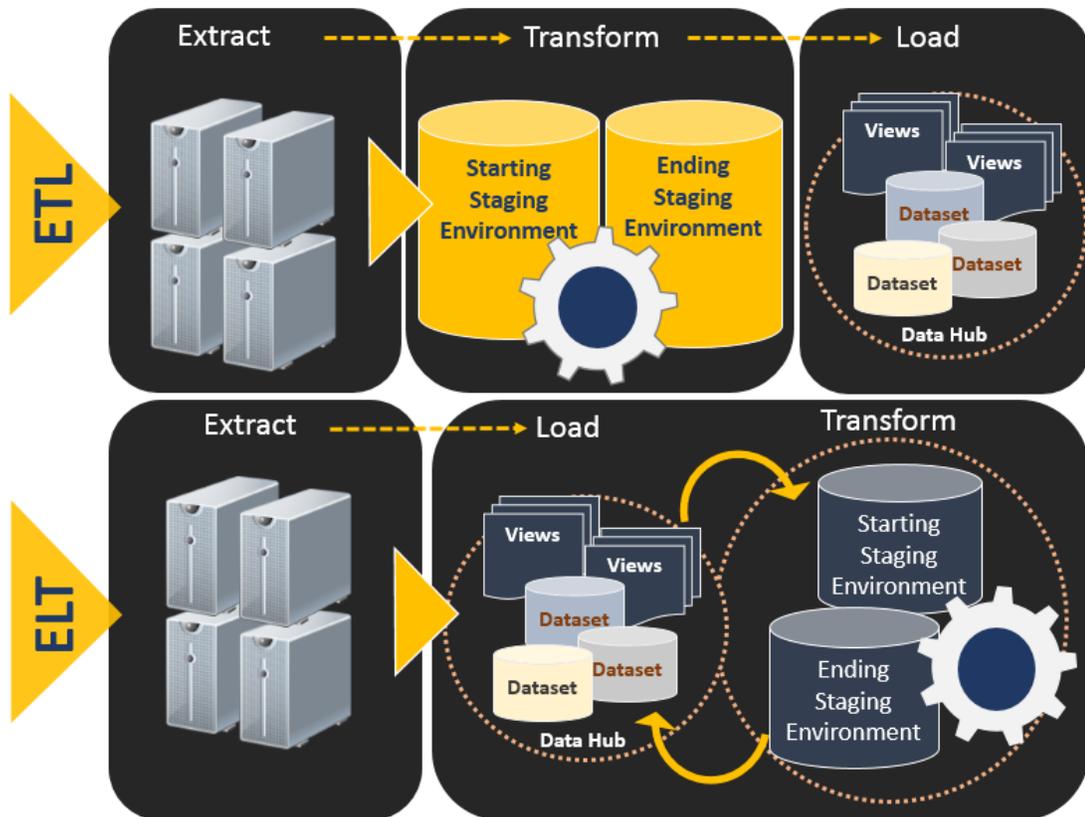


Figure 3 – ETL/ELT depiction

The Data Domain Hubs are in-memory staging areas for the information content and are comprised of variety of data types ready for consumption by the receiving nodes. These are self-contained components connected to the applications, search engines and tools on the Cloud to enhance communication between hubs and the receiving nodes (e.g. data units, applications, tools). Data Domains also manage workload and modulate data transfer between transmitting and receiving nodes. This model ensures the data gets utilized as Business Analytics and Applications tier demand services, yielding more efficient usage of computing powers and data units.

Business Analytics and Platform Services

The Business Analytics and Applications tier components are comprised of services-based technologies which typical end users expect from the enterprise data architecture; they add value to the data consumers' day to day operations or provide data visualization tools. Future FHWA cloud architecture platform services will adequately support data components within the application tier.

Through an incremental approach as elaborated in earlier sections, the Business Analytics and Data driven applications will be transitioned to the target data platform while the supporting infrastructure, platforms and systems migrate to the FHWA Cloud environment. Future development plans must comply with FHWA Cloud models and technical specifications.

Data Ownership

As discussed within Data Governance Plan Volume 1, the data ownership concept extends beyond FHWA due to the fluid nature of a data environment. The T-EDE will contain domain hubs and connectors as logical data groupings which are dynamic and scalable in order to serve the customers' needs. These additional data capabilities will require changes within the target data platform, data ingestion, and processing by introducing a role based ownership matrix of restful and run-time data, demanding a redefined data ownership.

While restful data will follow the current structure, the run-time datasets will benefit from a joint ownership matrix with integrity being preserved at the "rest" level, and propagated within the staging environment.

The migration plans and activities will preserve current data ownerships. They must take into consideration any existing data consolidations or deduplication efforts prior to migrating into the T-EDE. Data Storage and Management Services will be utilized by the Common Platforms Services to provide several common functions, such as SQL/XML Engines and Unstructured Data Discovery. Data Visualization and Reporting Services also will be included within the common data platform. It is imperative to define data ownership at the restful stage and further apply security restrictions during runtime as deemed necessary.

The primary users of FHWA T-EDE will remain the same as those within the current environment. With the additional capabilities provided by the cloud services, FHWA will be able to serve a broader user community and also ingest a much larger and more diverse set of data. The types and the extent of data usage will be determined by the Business Analytics and Applications tier, further restricted through a layered data security model. The only visible change to the current data user model is the access venues to the data. Below lists the general data user categories:

- Researchers
- Academia
- State & Local Partners
- Federal Partners
- Private Sector
- Administration & Congress
- Advocacy Groups
- Internal DOT Customers/ Modes

Interoperability within Target Enterprise Data Environment

Information interoperability is an important element of achieving a federated computing environment. As demonstrated in Figure 2, the FHWA Target Enterprise Data Environment (T-EDE) will be built as an interoperable platform, requiring a multifaceted focus on the following Federal Enterprise Architecture (FEA)

domains: Data/Information, Application, and Business. Data domain specifies what information needs to be exchanged; Application domain provides guidance on how the information should or can be exchanged; and Business domain justifies why a particular dataset or stream of information should be exchanged between different systems or business entities. The Information Sharing Environment (ISE) Information Interoperability Framework (I2F) can be used to give a practical view on the concepts explained earlier. Table 1 provides a listing of Data/Information domain requirements for interoperability within T-EDE derived from ISE I2F.

Information Sharing Environment (ISE) Information Interoperability Framework (I2F)		
ISE I2F Components	Description	Documents/Artifacts
Operational Capabilities	Provides Mission Context & Mission Needs: building on the operational context and defining why information needs to be exchanged. This also provides grounds for business requirements.	<ul style="list-style-type: none"> • Operational policies & procedures • Requirements definitions • Use cases • Business cases • Implementation guidance • Strategy plans • Inter/intra-agency memorandums of understanding (MOUs) • Memorandums of Agreement (MOAs)
Technical Standards	Technical Standards provide clear set of guidelines for both Operational and Technical Capabilities during information exchange. These guidelines are technical and also foundational in nature. These standards are developed by industry organizations and in cooperation with the government, or in some cases, by the government entities.	<ul style="list-style-type: none"> • IEPD/NIEM • XML/XBRL • UCORE
Technical Capabilities	Abstracted necessities stemming from Operational Capabilities Needs, although mission agnostic. Technical specifications are vaguely defined in order to allow for maximum freedom during implementation. Technical Capabilities also provides the necessary guidance for implementation, and incorporates the Technical Standards set forth by a multitude of sources.	<ul style="list-style-type: none"> • Catalogue current data assets & capabilities. • Determine new or needed data assets & capabilities. • Identify capability gaps. • Recognizing technologies necessary to build interfaces that are aligned with interface standards. • Formulate standardized interfaces between data capabilities.

Information Sharing Environment (ISE) Information Interoperability Framework (I2F)		
ISE I2F Components	Description	Documents/Artifacts
Exchange Patterns	Technical specifications on information exchange methods for one way and two way communication. The patterns may be simple abstractions of commonly accepted patterns, or a more complex combination of different approaches depending on the system design and the customer's needs.	<ul style="list-style-type: none"> • Query/Response (two way) • Broadcast (one-way) • Workflow (one and two way) • Orchestrated (one and two way) • Federated (type of orchestration) • Choreographed (one and two way)
Exchange Specifications	These specifications are arranged between different systems or data nodes according to business needs and also exchange patterns and technical capabilities.	<ul style="list-style-type: none"> • Ties in the mission and the business rules to information exchange and interoperability. • Defines the conceptual data structure and attributes to be modeled for implementation of data exchange technologies. • Describes the steps involved in exchanging of information. • Seeks a mature governance process in place for a solidified change control management.

Table 1 – ISE I2F Components and Requirements

Technologies and Standards Inventory

FHWA Enterprise Architecture (EA) will assist Data Governance Regimes & Coordinators and Data Stewards to implement technical capabilities partnering with Data Governance Technical Advisors. FHWA EA will provide FHWA offices with a catalogue of available technologies through either full acquisition process or as shared services. This catalogue will be available within FHWA Reference Architecture.

Data Architecture Security Model

The FHWA T-EDE security model will be defined within the context of cloud architecture security and in alignment with the FHWA Cloud implementation plan. Data will be secured from multiple levels (e.g. operating system, data storage, infrastructure access point, application, etc.) and by logical groupings of information to restrict user access to those domains. Additionally this model separates data domain hubs to restrict access per Trusted Zone and Public Zone where Trusted Zone will be accessible only to the trusted applications with end users granted access to those applications and privileged to corresponding data. All FHWA Information Systems must comply with the FHWA Cybersecurity Program Handbook and all listed Departmental, Administration, and NIST Policy Guidance.

APPENDIX A – KEY TERMS & ACRONYMS

Term (Acronym)	Definition
Big Data	Large volumes of data in variety formats, groupings, and sources with or without explicit associations or joins.
Broadcast pattern	A one-way data transmission sent from one source to many unknown receiving nodes periodically.
Bulk Data	Large collection of data with clearly established associations or joins between data attributes and content.
Business Intelligence Analytics	Advanced and special purpose application services which assist consumer nodes with business decision making, or specific calculations and requests.
Choreographed pattern	A data transmission method with predefined schedule, data content, source(s) and destination(s).
Cloud Providers	Organizations which provide or otherwise assist in providing or implementing cloud-based computing environments and capabilities.
Cloud Technologies Selection Guide	A near future FHWA produced document to guide different offices and divisions with their unique cloud implementations. The document is intended to promote enterprise alignment and conformity to the DOT Technical Reference Model.
Connectors	Technology interfaces which connect functional nodes at many different levels (e.g. application services, commands, data, etc.). Connectors are especially useful in connecting nodes that reside on different clouds, or legacy systems.
Data Containers	General term used to describe data groupings within one system, or across multiple connected systems.
Data Dictionary	A catalogue of data elements and their corresponding attributes or other relevant metadata, within the context of a specific system, business domain, or an enterprise environment.
Data Discovery	Services specifically designed for finding the right information from a vast pool of data.
Data Silos	Data system with inhibited collaboration or connectivity with other data systems, within an enterprise environment or a consumer community.
Data Visualization	Tools and services which assist in visualizing data within a specific context, or visually create synergy between otherwise unrelated categories of data.
Enterprise Architecture (EA)	Enterprise Architecture (EA) provides an abstracted view of an enterprise at various levels of scope and detail through documentation and information which support the planning and decision-making process within an organization. From a financial investment perspective, EA aligns business needs with Information Technology (IT) services to ensure IT investments improve the organization’s overall performance and mission execution.
Extract Load Transforms (ELT)	Data access and initial processing where the data is first extracted, then loaded into the staging or destination environment before being transformed into useful format.
Extract Transform Load (ETL)	Data access method where the data is transformed or processed within a preliminary area before it is loaded into the destination environment.
Federation pattern	Data transmission between otherwise disjointed sender/receiver nodes.

Term (Acronym)	Definition
Hubs	Logical working segments where data is made available for applications and/or systems' consumption. Data Hubs also are used to prepare data for storage or transfer into other applications or systems.
In-memory Data Analytics	Business Intelligence or other analytics services that utilize runtime or in-memory, and/or determine its usage. These services may determine source or destination nodes or any other calculation metrics at runtime and work with all types of data (e.g. structures, unstructured, semi-structured).
Integration Service	Data services specifically designed to integrate or merge data which reside on many different locations.
Investment Review Board (IRB)	FHWA investment decision board comprised of the following FHWA Leadership members: Associate Administrator for Administration, Chief Financial Officer, Deputy Chief Counsel, Director of the Office of Acquisition Management, and Associate Administrator for the Office of Federal Lands Highway, Rotating Senior Manager appointed by the permanent members.
Linked Services	Cloud services specifically created to connect resources to other Cloud services.
Metadata	Information catalogued to describe data elements or categories.
Mission Systems	FHWA system which are categorized as vital and critical for the administration's mission and operations.
NIEM/IEPD	National Information Exchange Model (NIEM) Information Exchange Package Documentation (IEPD) which substantiates the rules and standards for information exchange between systems.
Non-Mission Systems	FHWA system which are categorized as important for the administration's mission; however, these systems are not critical for the administration's operations.
NoSQL Database	Non-relational databases which stores and processes unstructured data. NoSQL databases are specifically optimized for managing and serving large volumes of variety data.
Orchestration pattern	Synchronized exchange of information between enterprises or disjointed computing nodes.
Public Web Systems	FHWA system which are available to the public through web access.
Query/Response pattern	On demand access to data through explicit request and in turn receiving data as a result.
Relational Database	Data storage catering to well-defined set of data attributes and associations (i.e. Structured Data).
Restful Data	Classification of data that is retrieved from or stored into a catalogued repository for each operation.
Run-time Data	Classification of data that is consumed or manipulated while loaded in memory, or temporarily held within a staging environment.
Semi-structured Data	Less orderly form of structured data where the data content contains the structure of the data itself. A good example is Extensible Markup Language (XML).
Structured Data	Data with well-defined set of attributes and associations.
Tightly Coupled	IT systems which heavily rely upon one another in preserving data integrity, operating and

Term (Acronym)	Definition
Data Systems	information linking. Tight coupling promotes faster task execution while causing setbacks in system upgrades such as introducing new components, or retiring other connected system(s).
UCORE	Universal Core – a federal information sharing initiative which supports the National Information Sharing Initiative (NSIS) and the associated agency strategies. ³
Unstructured Data	Data captured and/or stored without a set structure. In other words, unstructured data does not have the conventional attributes and associations that are achieved with structured data. A good example is textual or image files, sound bites, streaming content, etc.
Workflow pattern	Information being shared routinely as part of business operations, with a defined starting and ending node(s) as well as decision points.
XML/XBRL	Extensible Markup Language / Extensible Business Reporting Language

³ <https://www.ise.gov/universal-core-ucore>