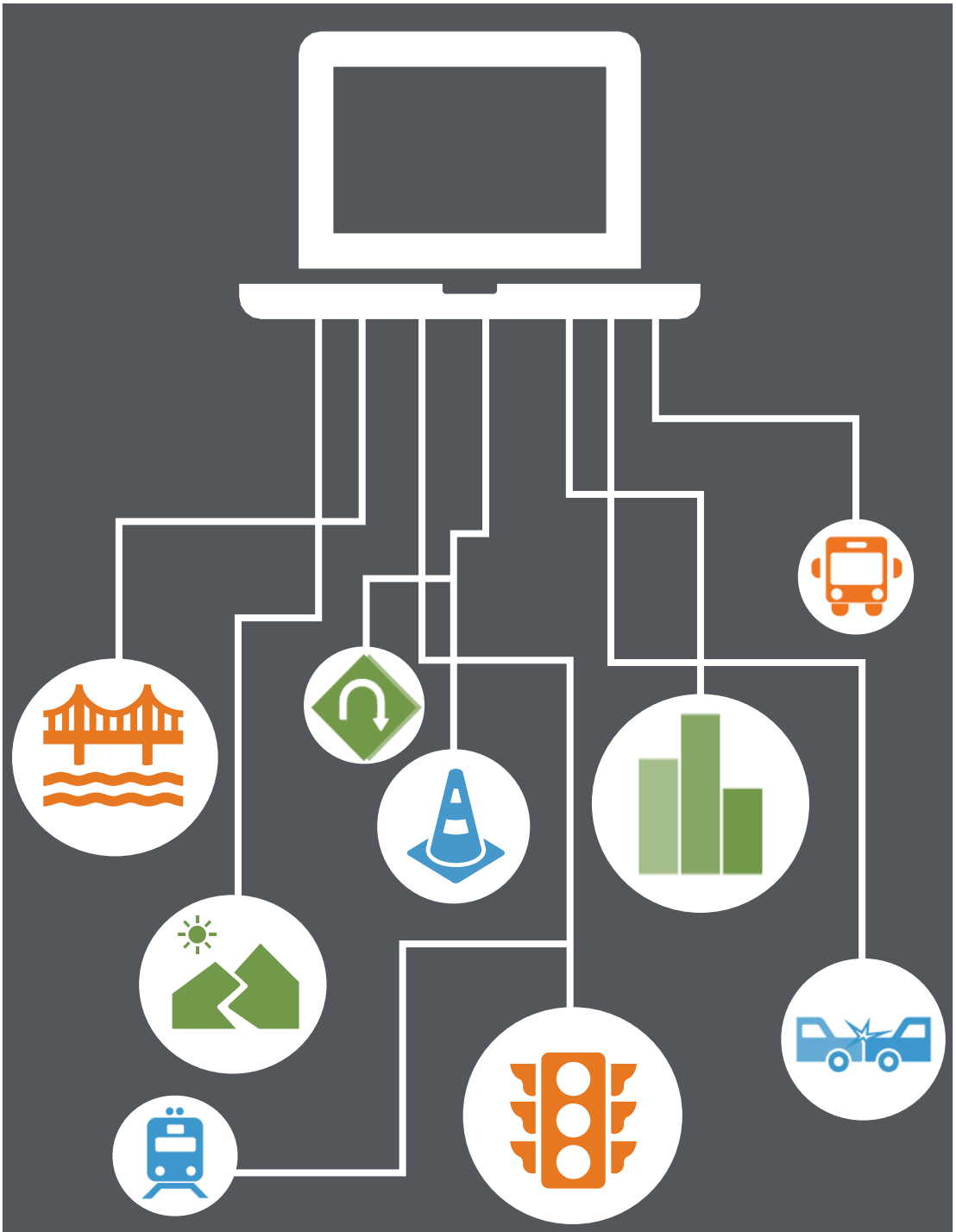


DATA MANAGEMENT
Business Plan

June 2020

Contents

01 Introduction

The Iowa Department of Transportation (DOT) gathers, stores, analyzes, and relies on a wide range of data and information to support business functions across the Agency. The Agency has developed three documents outlining it's strategic, operational, and tactical approach for managing one of the Agency's most critical assets – data.

This document represents the Iowa DOT's *Data Management Business Plan* (**DMBP**), which describes the data management strategies (data standardization and data sharing and integration), business procedures, and processes Iowa DOT will implement to achieve the optimal use of available data resources. The DMBP should be read along with the two supporting plans referenced in **Figure 1**.

The **DMBP** is a resource for Data Domain Trustees within the Agency and external stakeholders. However, the success of this plan depends on sustainable buy-in from across the Agency including the executive level, Information Technology (IT), Data Stewards, and data users.

Iowa DOT Data Management Goals

- Strengthen data governance
- Formalize data life cycle and management
- Improve data architecture and integration
- Improve data collaboration
- Improve data quality

Data Governance Roles

A ***Data Steward*** is an operational-level role that implements data management strategies for a specific data source.

A ***Data Domain Trustee*** is a tactical-level role that helps integrate high-level strategies for data management with day-to-day data activities carried out by data users.

The ***Data Management Committee*** (DMC) is a strategic-level group that ensures continuous buy-in for data management initiatives and activities.

Figure 1. *Iowa DOT Data Management Plans*

Organization of the DMBP

The DMBP includes the following sections:

- ***Data Standards:*** This section provides a framework for accomplishing the goals of the Agency through data standardization. In this section, a procedure for creating or modifying data standards is provided. The section also discusses data standards discovery, data definitions, data naming conventions, and data policies.
- ***Data Sharing and Integration:*** This section provides a tactical overview of how data will be shared and integrated within the Agency. To avoid the dangers of siloed data and data projects, this section discusses the necessary flow of data, safeguards for data storage, and the means in which data can be accessed. Additionally, it discusses a procedure for sharing new data sources and information on data-related projects throughout the Agency. These procedures ensure new data and data-related efforts are well-communicated to avoid duplication of efforts.

02 Data Standards

Data standards provide the opportunity for the Agency to gather, format, define, and share data that meets the Agency's business needs. Through the implementation of data standards, the Agency can establish a process to reduce ambiguity, redundancy, and inconsistency present if data was managed without regard to other datasets. Clear management strategies are especially important as the Agency creates and uses diverse and abundant datasets. Therefore, by developing and implementing data standards, the Agency will be able to understand and reference data in a clear and concise manner through centralized policies and documentation, as well as save time in data processing and provide added value to data users.

This section describes a procedure for creating or updating data standards within the Agency. The guiding framework is intended to assist the Agency in better understanding where data standards are needed and how existing standards can be improved. **Figure 2** describes the framework and the following paragraphs explain each task.

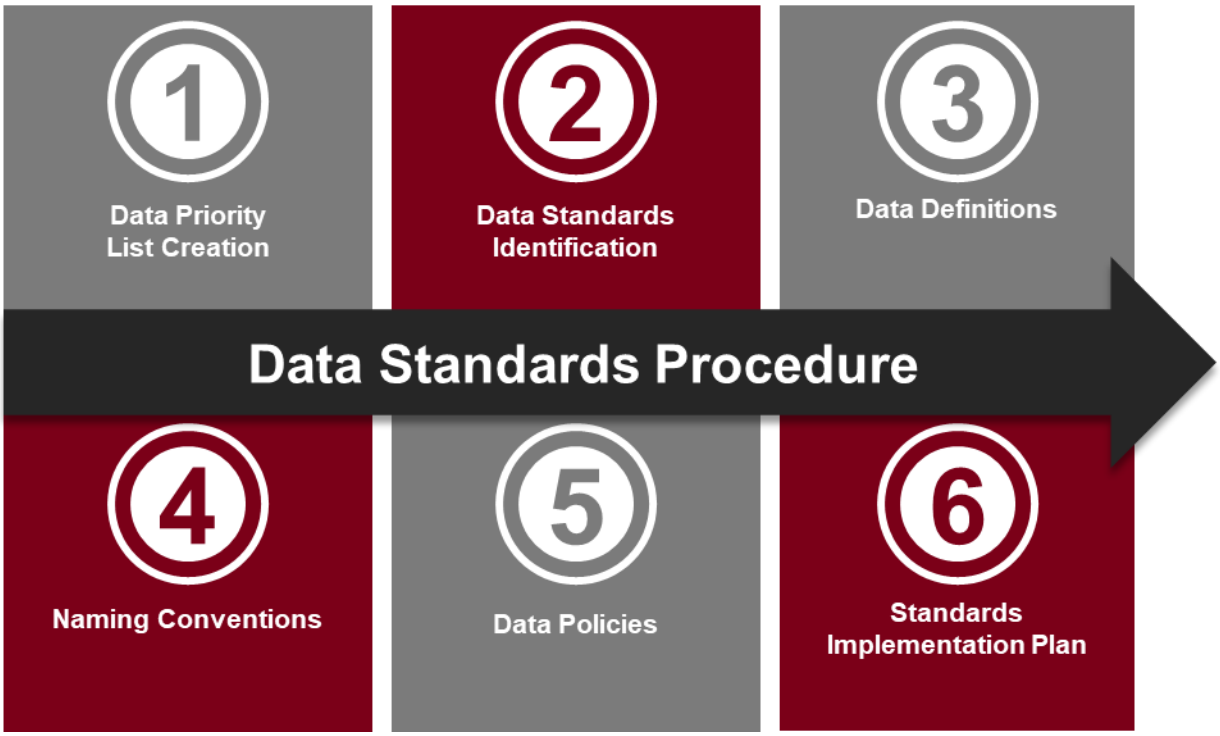


Figure 2. *Data standards framework*

Data Priority List Creation

The Agency's *Data Management Strategic Plan* (**DMSP**) contains a preliminary list of data sources and data activities used to support pavement and bridge asset management. The identified data sources provide an overview of the types of data used to support a business process—in this case, asset management. In the development of data standards, the Agency will build on on-going work to better understand where data standards need to be implemented and where standards currently exist. However, the Agency will also organize identified data sources by priority.

The priority of a data source corresponds to the data source's importance in supporting data activities within the business area. Specifically, the Agency will evaluate the priority of each data source using a combination of the defined importance of each data source to each data activity and engineering judgment. Prioritization of the data sources will enable the Agency to effectively utilize Agency resources on data sources that are of high priority. While the goal is to establish or refine standards for each data source, the Agency will develop standards for data sources in order of priority.

Data Source Priority

The importance of a data source in supporting data activities within the business area.

Example—Data Priority List Creation

In July 2019, the Agency conducted a data maturity assessment and created a list of data sources crucial to pavement and bridge asset management. In this example, four of the identified bridge asset management data sources are explored; the data sources include bridge condition data, traffic volume data, drainage/roadside assets data, and geospatial foundations. In order to prioritize this data, the Agency will create a table that summarizes the number of activities dependent on each of these data sources (denoted by X). The priority table for these five data sources is provided below.

Activity Type	Data Source of High Importance to Activity			
	Bridge Condition	Traffic Volume	Drainage/ Roadside Assets	Geospatial Foundations
Current Conditions Assessment	X			
Resource Allocation and Treatment Selection	X	X	X	
Bridge Needs and Risk Assessment	X			
Inspection Management	X	X		
Scope Development	X			
Project Prioritization	X			
Strategy Prioritization	X			
Bridge Plan Integration	X	X		X
Local Bridge Funding Selection	X	X	X	
Load Rating and Heavy Load Permits	X	X		X
Oversize Permits	X	X		X
Performance Management and Target Setting	X	X		X
Cross-Asset Tradeoff	X	X	X	
Bridge Equipment Management	X		X	
Heavy Load Routing	X	X		X
Total Number of Activities	**15**	**9**	**4**	**5**

Assuming that the total number of activities where the data sources are of high importance is aligned with Agency's expectations, the order of priority of these data sources —from highest priority to lowest priority—is bridge condition data, traffic volume data, geospatial foundations, and drainage/roadside assets data.

Data Standards Identification

Prior to creating or modifying any data standard for the prioritized list of data, existing data standards within a business area need to be identified. As stated previously, data standards are rules used to describe data and to make data more consistent. Therefore, identifying existing standards within the Agency should focus on determining if datasets comply with a set of rules, both formal and informal. In discerning data standards that already exist, the duplication of data standards or policies can be avoided.

> **Data Standard**
>
> Rules used to describe data and to make data more consistent.

Standards identification is further divided into three key tasks that provide insight on the current state of data standards within a business area. The tasks include the development of a master list or catalog of current standards (both formal and informal), a high-level summary of each standard, and the identification of gaps in existing standards or data sources requiring additional standards. **Figure 3** provides an overview of the key actions that will be taken during the data standard identification process. The identification of standards should largely be an effort of Data Domain Trustees, Data Stewards, and data users who have extensive knowledge of the data sources, data activities, and data standards within the business area being explored.

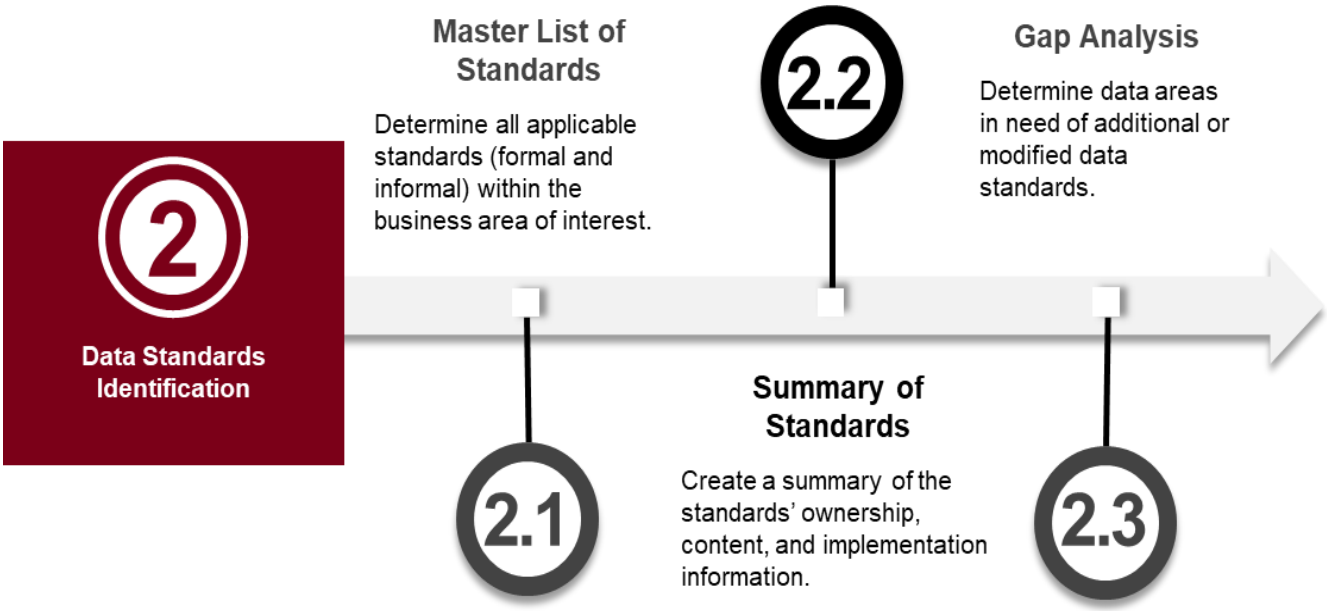


***Figure 3**. Data standards identification process*

2.1 Master List of Standards

First, the Agency will revisit the data sources and data activities pertinent to the business area being explored. The target of the team is to identify specific standards and practices governing data sources within the business area. Standards can be federal- or state-created, formal (documented) or informal, and widely adopted by the Agency or only adopted by a few data users. Formal standards should be recorded and assessed separately from informal standards, as informal standards require additional information gathering during the data standards identification process. An informal standard can include a practice adopted by a Bureau that is not documented.

The team will identify data sources that currently are not governed by any data standards. Further discussion of these data sources will occur during the gap analysis.

Example—Master List of Standards

To create a master list of data standards, the Agency will identify any standards that may exist for each data source identified. For example, in pavement asset management, the data sources that can be assessed include (but are not limited to) pavement condition data, vehicle volumes, crash records, and Linear Referencing System data. Both formal and informal standards for each of these data sources are listed below.

Data Source	Formal Standards	Informal Standards
Pavement Condition	Distress Manual, Data Quality Management Plan (DQMP), Fast Act (Final Rule (§490.311))	State-level pavement condition is categorized as “Good”, “Fair”, or “Poor” based on the IRI alone.
Vehicle Volumes	Traffic Monitoring Guide (TMG)	N/A
Crash Records	Model Minimum Uniform Crash Criteria (MMUCC)	N/A
Linear Referencing System	All Road Network of Linear Reference Data (ARNOLD)	N/A

2.2 Summary of Standards

Using the comprehensive list of existing data standards created in the previous step, the Agency will focus on summarizing the high-level details of each standard. In addition to identifying whether the standards are formal or informal, additional information useful for data users and data managers will be collected by assessing the standards documents or in the case of informal standards, by interviewing technical area experts on practices currently used for the key data sources identified.

The following is a list of key questions to answer through the summary process. While this list provides areas of interest for the data management practice, the Agency will develop additional questions based on the specifics of the business area being assessed. For ease of access and the consistency of the summary process, the Agency will document and store the final list of questions appropriate for the exercise in a central location.

Standards Ownership

- Who created the current set of standards?
- Who owns the current set of standards?
- How often do the owners update these standards?
- Where can the standards be found (i.e. online, in a central drive, or hard copy)?

Standards Content

- What data sources are governed by these standards?
- What activities are governed by these standards?
- What is the overall role of these standards in terms of data life-cycle and data management?

Standards Implementation

- Are the standards federally required?
- How are the standards implemented? How many data users adhere to the standards?
- Are the standards related to or overridden by other standards within the business area?
- How are the standards enforced?

Informal Standards Only

- Why are the standards undocumented?
- Is there a plan to document these standards? If so, who is responsible?
- How were the standards created?
- Has any authority acknowledged or vetted the standards?

2.3 Gap Analysis

The final task in the data standards identification process is to perform a gap analysis. The Agency will work with other stakeholders to assess and identify areas of weakness and opportunities. The assessment will consider all the data standards collectively to discern whether existing data standards a) adequately govern all the data sources within the data practice, and b) provide comprehensive guidelines for collecting, analyzing, and storing data used by the Agency.

The data sources without governing data standards will be used to determine the specific needs for additional standards. Details of the proposed standards will be developed further through Task 3, Task 4, and Task 5.

3 *Data Definitions*

The primary purpose of this task is to ensure that stakeholders have a common understanding of data collected and created across the Agency. Defining data allows the Agency to avoid ambiguity and to provide a shared understanding of the Agency's data. Through this process, the Agency will be able to define the relationships between different data sources.

The creation of metadata is one means the Agency uses to define data. Metadata is information about data necessary to identify and organize data. Therefore, the creation of metadata makes data activities such as data integration and data sharing easier to conduct.

Metadata

Information about data that is necessary to identify and organize data.

The following subsections describe an approach to establishing good metadata practices for existing and new data sources the Agency relies on. **Figure 4** provides an overview of the key activities that will take place during the data definition process.

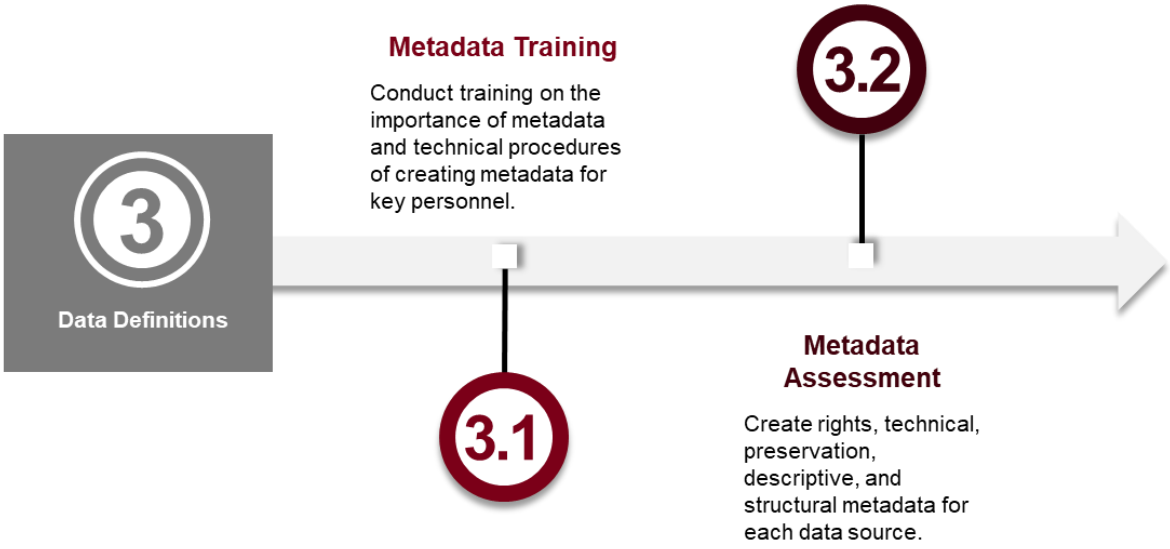


Figure 4. *Data definition process*

3.1 Metadata Training

Prior to creating metadata for data sources within the Agency, key personnel, such as Data Stewards, will require training on what metadata is and why metadata is important to the Agency. The Agency will consider different categories of metadata, such as the five predominantly used for digital libraries[1]. The subsections below provide a preliminary overview of the five metadata types that will be the

[1] UC Santa Cruz University Library. *Metadata Creation*. Retrieved from https://guides.library.ucsc.edu/c.php?g=618773

focus of the training. The training may include in-person presentations and workshops as well as informational videos.

Rights Metadata

Rights metadata is focused on the usability or accessibility of a data source or database. In creating rights metadata, the Agency will ensure that the data sources identified within a business area are being utilized and accessed properly. Specifically, the Agency will use rights metadata to identify the creator of the data, the year the data was created or updated, the copyright status of the data, publication status (whether it is available to the public), and the date the rights metadata was created or updated.[2]

Rights Metadata

Information on the usability or accessibility of a data source or database.

Technical Metadata

Technical metadata details information about the type of data and technical needs for accessing a data source. In an Agency where data often requires additional processing or analysis tools to be useful, technical metadata is key. During the creation of technical metadata, the file type, the file size, software or tools necessary to access or process data, and hardware specifications for utilizing the data will be identified. If additional software or tools are necessary for a data source, the type of software, version of software, and operating system and type will be specified in the metadata document.[3]

Technical Metadata

Information about the type of data and technical needs for accessing a data source.

Preservation Metadata

Raw data is not often directly utilized or managed by the Agency. Therefore, an understanding of the pre-processing procedures or operations involved in preparing a data source will be detailed in preservation metadata. Information collected for the purpose of the metadata documentation will include equipment or software used to transform data into a usable form, any quality checks or data authentication, operations carried out including field name changes, metric conversions, or field calculations to prepare the data for use, and any secondary data sources the data relies on for field population or field calculation.[4] By creating preservation metadata, the Agency can effectively monitor where data comes from and how it is assessed for quality purposes.

Preservation Metadata

Information on processing procedures or operations involved in preparing a data source.

Descriptive Metadata

Descriptive metadata provides information about the attributes collected and stored within a database. Specifically, this metadata contains a dictionary of attributes reported within the database. During this process development, the data owner will define the name, purpose or summary, and field type (i.e., numeric, text, binary) of each data attribute. For a robust system, each field will be thoroughly described, and possible field values will be determined. Due to the level of detail required for descriptive metadata, this process can be costly in terms of time and resources required for establishment. However, descriptive metadata provides long-term value as it enables databases to be more easily queried and efficiently used by individuals less familiar with the datasets.

Descriptive Metadata

Information about the attributes collected and stored within a database.

[2] Whalen, Maureen. (2008). *Rights Metadata Made Simple.* Introduction to Metadata. Paul Getty Trust.
[3] University of Warwick. (2019). *Technical Metadata.* Retrieved from https://warwick.ac.uk/services/library/mrc/digital-preservation/describing/technical-metadata/
[4] International Association of Sound and Audiovisual Archives. *Preservation Metadata*. Retrieved from https://www.iasa-web.org/tc03/14-solutions-dmsss-small-scale-manual-approaches-digital-storage

Structural Metadata

Structural metadata describes how data is organized or related to other databases. While preservation metadata defines secondary data sources a dataset is dependent on for field population, structural metadata expands on the data dependencies. At a foundational level, structural metadata will provide information on how elements are combined or organized for use within the Agency. However, structural metadata in this process refers to establishing the interconnectivity between data sources. In doing so, activities related to data sharing and integration are more easily carried out.

> **Structural Metadata**
>
> Information about how data is organized or related to other databases.

3.2 Metadata Assessment

With an understanding of data definitions, specifically metadata, the Agency will next assess existing data definitions and create metadata for each data source identified within the business area. Table 1 provides a summary of the different metadata types and the subsequent data elements that will be defined during this step.

***Table 1**. Metadata types summary*

Metadata Type	Contents	Data Elements Defined
Rights	Information about the usability or accessibility of a database. The information collected may include when or if the database is made available to the public.	• Creator of the data • Year the data was created • Copyright status of the data • Publication status (whether it is available to the public • Date the rights metadata was created or updated
Technical	Information about the properties of the data such as the format, data size, and data type.	• File type • File size • Software or tools needed to access or process data • Hardware specifications for utilizing the data
Preservation	Information on actions carried out on a dataset to ensure the data is usable going forward. This process includes the renaming of data attributes or the addition of newly collected data.	• Equipment or software used to transform data into a usable form • Quality checks or data authentication • Operations carried out including field name changes • Metric conversions or field calculations to prepare the data for use • Secondary data sources the data relies on for field population or field calculation
Descriptive	Descriptions of the resource and subsequent attributes within the database. For clarity, it is important to ensure that each attribute is adequately defined.	• Name of data fields • Purpose or summary of data fields • Field type (i.e. numeric, text, binary) of each data attribute

Metadata Type	Contents	Data Elements Defined
Structural	Information on the relationship between objects within the database and the relationship between separate databases.	• Information on how elements are combined or organized for use

The Agency will use different methods to create metadata, including Microsoft Access tables or more advanced/integrated software within existing systems. The metadata creation task will involve Data Stewards from the respective data domains considered to be a part of the agency-wide data management initiative. Through this exercise, users will better understand how the data is organized and what each data source contains. Task 8 provides more information on how the Agency will create structural metadata.

Example—Data Definitions

In this example, assume the Agency wants to create data definitions for Crash Vehicle Data[1]. This crash data contains attributes related to driver characteristics and the causes of the crash and can be used to recommend pavement and bridge treatments or safety improvements. For the purpose of this example, assume this data is directly related to existing pavement condition and bridge condition data. The table below describes each of the metadata definitions for the data source.

Metadata Type	Example information
Rights	Data is available to the public via the Iowa DOT Open Data Portal.
Technical	Data is available as a .csv (759 MB), .gdb (353 MB), or a .kml (3.02 GB).
Preservation	Data was compiled from the Traffic Safety Data and Analysis website.
Descriptive	*Example: Driver Age* The Driver Age is age of the driver at the time of the crash record.
Structural	The data is related and linked to pavement and bridge condition data through the GlobalID.

1 https://public-iowadot.opendata.arcgis.com/datasets/crash-vehicle-data/data

Naming Conventions

Another part of creating data standards is establishing and updating naming conventions to suit the needs of the Agency and its diverse data users. This process will establish or update a naming standard the Agency will use to develop, define, and name databases and data elements or attributes. The naming standards will adhere to IT policies and preferences while serving the needs of data users. The overarching goal of the Agency is to ensure that the approach to standard naming conventions aligns with good industry practice and standards such as those proposed by the International Organization for Standardization (ISO). The Agency, with oversight from the Data Management Committee and IT, will implement an effort to reinforce existing naming conventions or identify and recommend methods to standardize data elements or attributes across the Agency.

In this section, the procedure for implementing naming conventions is discussed. The practices are meant to be used on each data source identified within the business area being assessed. **Figure 5** provides an overview of the key activities that will occur during the data naming process.

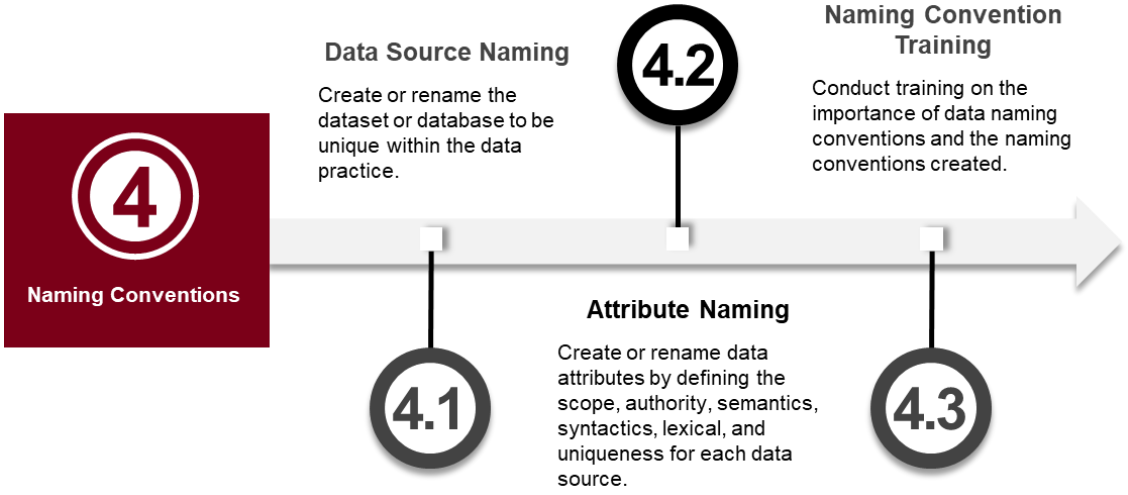


Figure 5. *Naming convention process*

4.1 Data Source Naming

The name of a database or data source is important for data standardization. When databases are properly named, a data user can understand what data the database contains and the version or creation date of the data, which encourages data sharing and facilitates data integration. As such, determining data source naming conventions that best serve the Agency is the first step in the standardization of naming conventions. The process can be further divided into two key actions—selecting identifiers for each data source name and determining how the identifiers will be used to create the final data source name.

Standards, such as ISO 11179 Metadata Registry, suggest data sources utilize three identifiers for naming purposes. The three identifiers suggested are a registration authority identifier (RAI) or a description of the entity that owns the data, a data identifier (DI) or a description of the category or intended use of a data source, and a version identifier (VI) or a version number or creation date of the data source. When considering the data sources owned by the Agency, the RAI can be defined as the name of the Bureau where the data originated (i.e. Analytics), the DI can be defined as the name or abbreviation of the contents of the data source (i.e. PMIS), and the VI can be defined as the date the data was created (i.e. 01122019). Each identifier will need to be determined for each data source within a business area. In addition to determining the three identifiers of each data source, the Agency will also establish how these identifiers will standardized. IT and Data Stewards will need to identify a complete list of all the potential RAIs for the Agency; RAIs selected for each data source should match the syntactics used on the list. In terms of standardizing DIs and the VIs, the Agency will determine how each of these identifiers will be created. The Agency will consider the syntactics and format for these identifiers, i.e. will DIs be limited to a length of 10 letters and will the dates reported for the VI follow a “MMDDYY” naming system?

Data Source Naming Identifiers

Registration Authority Identifier (RAI): Description of the entity that owns the data.

Data Identifier: Description of the category or intended use of a data source.

Version Identifier: Version number or creation date of the data source.

Once each of the individual identifiers has been standardized, the Agency will identify the order and the separators used to create a single data source name. For example, a data source name may consist of

the three identifiers separated by spaces or by underscores. The Agency will document and communicate the final data source naming conventions once established.

Example—Data Source Naming

Assume the Agency wants to create a standardized name for a new data source containing average costs for crack sealing throughout the State. For the purpose of this example, assume the data source was created by the Maintenance Bureau on June 19, 2015. To create a standardized data source name, the Agency creates a table describing each identifier, the standards adopted by the Agency, and the defined identifier for the data source being considered.

Identifier	Standards Description	Example
Registration Authority Identifier	The RAI corresponds to the Bureau where the data originated. The Bureau name is shortened to the first five letters.	MAINT
Data Identifier	The DI is 3-12 letter description of the data source. All letters of the description are capitalized and all spaces necessary are denoted by underscores.	CRKSEAL_COST
Version Identifier	The VI is the date of data collection utilizing the MMDDYYYY format.	06192015
Final Data Source Name	The final name is a concatenation of the three identifiers using underscores as separators. Therefore, all data sources are named as RAI_DI_VI.	MAINT_ CRKSEAL_COST_06192015

4.2 Attribute Naming

In addition to creating standards for naming individual data sources, the Agency will also define standards for naming data attributes or data elements. Data attributes or data elements are analogous to column headers within a data source. Because data elements between data sources are used to link data for data analysis, it is important that these attribute names follow the same set of standards. ISO 11179 provides individual principles that guide the creation of attribute naming conventions for data owners. To promote data sharing and integration, the following principles will be defined.

Each of the naming conventions are to be defined for the attributes of each data source. The exact naming conventions fall under the purview of IT, Data Stewards, and data users within a data domain. The Data Management Committee will ensure that these naming conventions align with federal requirements while meeting business objectives.

Example—Attribute Naming

Assume the Agency wants to define the attribute naming conventions for the Bridgelet Location dataset[1]. The table below describes each of the attribute naming convention principles for the data source.

Naming Convention Principle	Example Information
Scope	Because the naming conventions were established after the data was collected, the naming conventions are considered descriptive.
Authority	The DMC and IT have authority over the naming conventions of this file.
Semantic	All attribute names are delimited by spaces.
Syntactic	No specific syntactic rules were applied to attribute names. However, all attribute names are descriptive.
Lexical	No specific lexical rules were observed.
Uniqueness	Each attribute name in the table is unique to that table but not unique to all datasets/systems (i.e., Location is a field that may exist in multiple tables but is unique within those tables).

1 https://public-iowadot.opendata.arcgis.com/datasets/bridgelet-location

4.3 Naming Convention Training

Once both data source and data attribute naming conventions have been created, the Agency will provide training on these standards to key personnel, such as IT and Data Stewards. Specifically, the training will describe the importance of naming conventions to the data practice and introduce the naming conventions created throughout the previous steps. The training may include in-person presentations and workshops as well as informational videos.

5 *Data Policies*

Data management policies are derived from the Agency's vision, mission, and principles for data management, and are expected to provide specific guidelines for how the Agency conducts data activities to ensure data complies with the data principles of the Agency. As such, the defined policies will address what the Agency does to ensure data is **valuable**, **accessible**, **relevant**, **compliant**, **of high quality**, **uniform**, **secure**, and **efficient**.

Data Principles

The Agency is focused on providing data that is:

- Valuable
- Accessible
- Relevant
- Compliant
- Quality
- Uniform
- Secure
- Efficient

The Agency will accomplish the assessment and creation of data policies through multiple activities. Data Domain Trustees and Data Stewards will lead these efforts with oversight from the Data Management Committee as necessary. **Figure 6** provides an overview of the key activities that will take place during the data policy creation process.

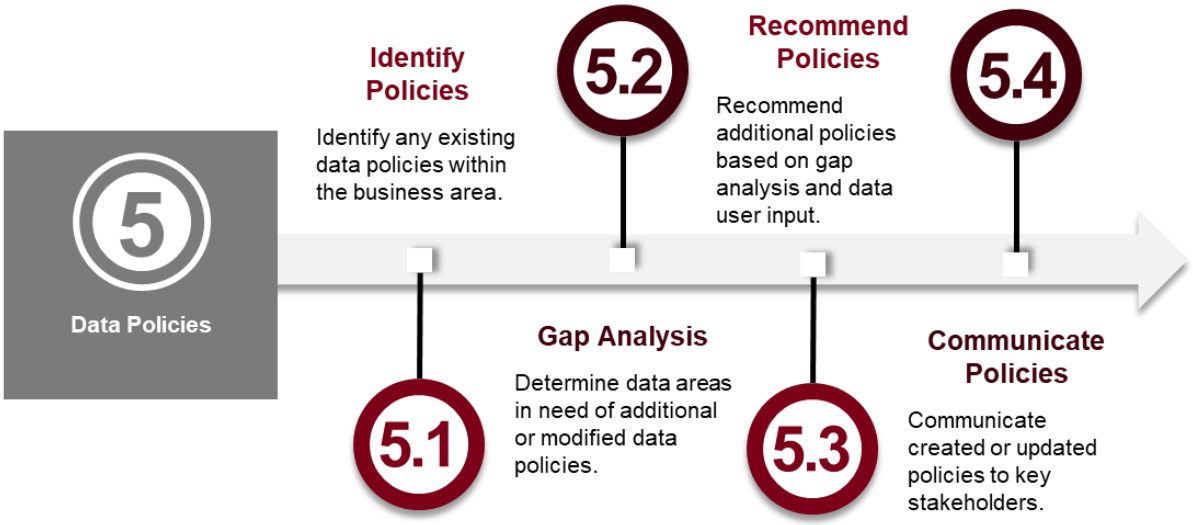


Figure 6. *Data policy creation process*

5.1 Identify Policies

To reduce redundancy and duplication of efforts, the first step in assessing or creating new data policies is to extract and understand existing policies in the data standards identified through Task 2. The Agency will focus the policy identification process on understanding how data operates in the context of a given business area. Within the Agency, policies will be considered through the lens of the data management goals and the Agency's data principles.

Table 2 summarizes examples of typical questions and data policies that can be implemented in each of the policy areas.

Table 2. *Data policy examples*

Policy Area	Typical Questions	Data Principle(s) Addressed	Example Policy for Agency[5]	Responsible Party
Strategy and Governance	Does the current data system comply with federal or state policy?	Compliant	All data that is accessible to the public may not include any personally identifiable information.	Information Technology Division, Data Management Committee, Data Domain Trustees
Life Cycle Management	How long does data need to be maintained?	Valuable	Databases will be reviewed yearly to determine the usefulness of items collected.	Data Management Committee, Data Domain Trustees and Stewards
	When does data lose its usefulness?	Valuable, Relevant	Data retention policies will be reviewed annually.	
	Is that data easily accessed and analyzed by users?	Accessible	All data definitions and descriptions must be documented and stored in a central location.	

[5] Harrison, F. D. (2015). *NCHRP Report 814: Data to Support Transportation Agency Business Needs: A Self-Assessment Guide.* Washington, D.C.: Transportation Research Board of the National Academies.

Policy Area	Typical Questions	Data Principle(s) Addressed	Example Policy for Agency[5]	Responsible Party
Architecture and Integration	Are the proper types of data being collected?	Relevant	Data collected will be reviewed annually to determine the usefulness and value of the data to the user.	Information Technology Division, Data Management Committee, Data Domain Trustees
	Is there enough detail in the data collected for decision-making purposes?	Relevant, Uniform		
Collaboration	Who has access to specific databases?	Accessible, Secure	Data must be classified and restricted based on sensitivity.	Information Technology Division
	What type of access do users have (i.e., who can edit data)?	Accessible, Efficient	Data access must be provided based on standard methods for managing access (i.e., standards book).	
Quality	Is the data collected frequently, accurately, and completely?	Relevant, Quality	Data collection procedures will be reviewed annually to ensure the data is collected frequently, accurately, and completely.	Data Management Committee and Data Domain Trustees

The Data Domain Trustees and Data Stewards conducting the analysis will use these questions as a starting point. Policies for each set of standards will be documented and stored in a central location.

Gap Analysis

Once the existing policies have been identified, the Data Domain Trustees and Data Stewards will determine policy areas with unclear or missing policies. As a minimum, for all the data sources within a business area, the questions developed in

Table 2 will be answered. Policy areas that do not meet this minimum will be noted as an area needing improvement. For data sources without any governing standards, Data Domain Trustees will identify the necessary policies in each of the identified policy areas.

While the gap analysis is the responsibility of the Data Domain Trustees, data users are key to identifying operational or tactical issues with existing policies. During this phase of the process, data user input on the policies identified will be collected via workshops, surveys, or other communication methods.

Recommend Policies

The next step in assessing and developing data policies is to provide recommendations for new policies or suggest revisions to existing policies based on the gap analysis process. Recommendations for new policies will be vetted by the Data Management Committee to ensure the policies align with Agency-wide data management goals. At the operational level, the Data Domain Trustees will provide a reasonable period to allow data users to voice any questions, comments, or amendments to policies proposed. Once buy-in from the strategic-level, tactical-level, and operational-level personnel is established, implementation procedures, enforcement, and monitoring procedures will be planned.

Communicate Policies

The Agency will implement and enforce policies updated or created during the data policy assessment process. While some data policies will affect an entire business area, others may be applicable to a smaller subset of stakeholders. Because the effectiveness of data policies is rooted in implementation and compliance, the identification of Divisions, Bureaus, and key stakeholders affected by

the data policies is crucial to the policies' effectiveness. Through this process, the Agency will consider each of the data policies that have been updated or created and determine individuals that create, manage, or consume data to support analysis, planning, decision-making, or communication with stakeholders within the affected area. Stakeholders or groups whom the policies apply to will be noted in the policy documentation (when applicable) and notified and trained on any changes made to the practice.

Standards Implementation Plan

Once each of the previous tasks have been carried out, the Agency will compile the findings and recommendations into a data standards implementation plan. The plan, which will be focused on identifying key action items necessary to establish or improve the Agency's data standards, will be created by the Data Domain Trustees with insight from the Data Management Committee and Data Stewards. Depending on the strength of existing standards and the number of action items proposed, the plan will prioritize standardization of the high importance data sources identified in the first task of the data standards framework.

03 Data Sharing and Integration

The Agency collects, combines, and shares data and information across Divisions, Bureaus, Districts, and business areas, as well as with external stakeholders. Internally, there is a horizontal and vertical flow of data and information across Divisions and Bureaus to support decision-making. Externally, the Agency shares data and information with the public and stakeholders of the transportation system. These activities are largely supported by the data sharing and integration practices adopted by the Agency. Data sharing describes the process of providing access to data. Data integration is the use of technical and business processes to combine data from disparate sources into meaningful and valuable information. These processes, when correctly implemented, reduce issues related to mismatched management systems and data formats, timely access to information, increased data processing times, data duplication, and data redundancy. Therefore, it is important for the Agency to have efficient data sharing and integration processes in place.

Data Sharing

The process of providing access to data.

Data Integration

Use of technical and business processes to combine data from disparate sources into meaningful and valuable information.

Addressing these challenges, this chapter describes a data sharing and integration framework that leverages existing culture, practices, and strengths of the Agency's data sharing and integration infrastructure, including applications, techniques, technologies, and management services. The framework provides enough flexibility for the Agency to handle future changes in technology and new application development. **Figure 7** describes the key components for improving the data sharing and integration practice.

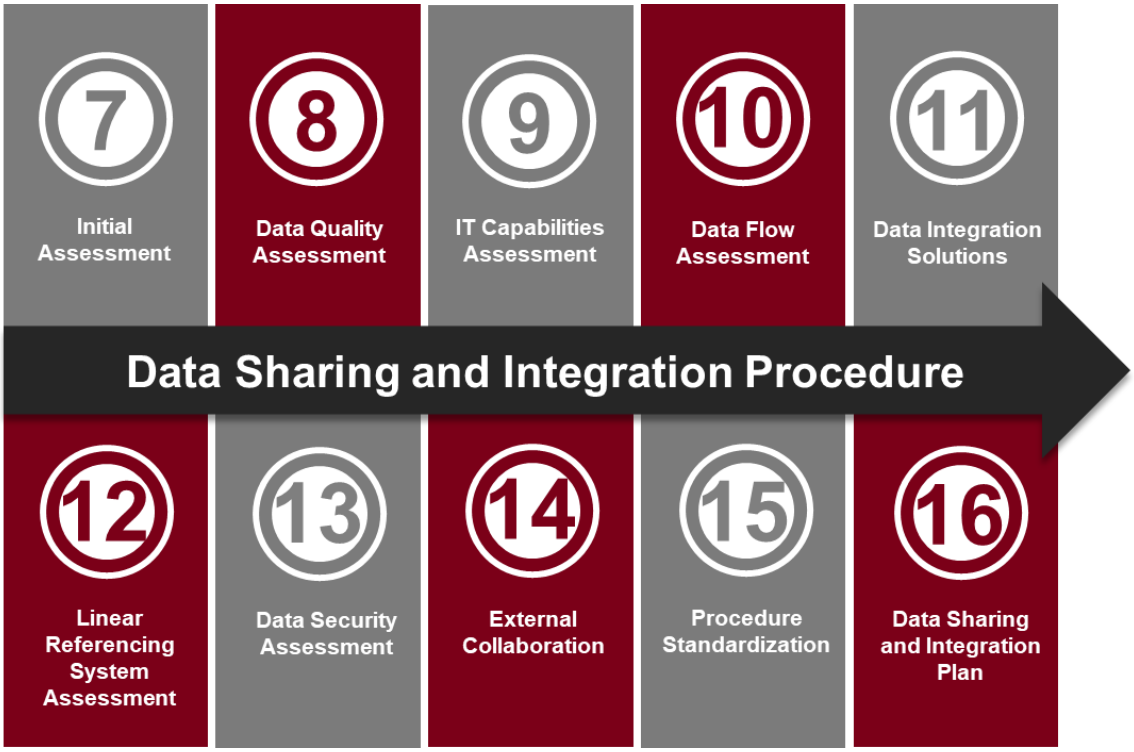


Figure 7. *Data sharing and integration framework*

Initial Assessment

Prior to developing or implementing any data sharing or integration methods, the Agency needs to take stock of and identify the key data sources, data activities, skills, and practices necessary for data sharing and integration within a business area. By identifying what is existing and what is needed in the area of data sharing and integration, the Agency can determine key objectives of the improvement process and better utilize Agency resources. **Figure 8** provides an overview of the tasks necessary to establish a foundation for identifying what purpose the data sharing and integration efforts will serve.

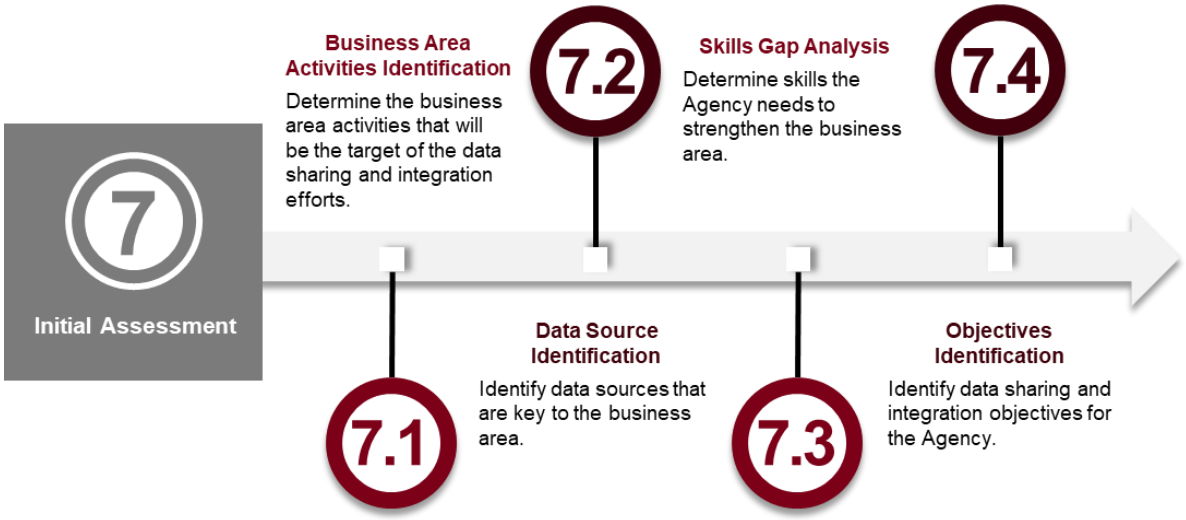


***Figure 8**. Initial assessment of data sharing and integration process*

7.1 Business Area Activities Identification

While the Agency will move towards strengthening the data sharing and integration capabilities of the Agency enterprise-wide, the initial efforts will focus on optimizing data sharing and integration for a specific business area. Therefore, the business area being assessed will be clearly defined to better understand data sharing and integration needs. Specifically, the Data Domain Trustees (with oversight from the Data Management Committee) will need to determine the business area activities where data sharing and integration are important. The Agency will revisit the data sources and data activities identified in the DMSP and in Task 1 of the data standards creation procedure and determine if there are any additional data activities the Agency would like to support in the future. Future activities may include activities the Agency is not currently supporting due to a lack of data or resources but is core to the goals of the business area. In doing so, the Agency will have a better understanding of which data activities need to be supported by the data sharing and integration processes.

7.2 Data Source Identification

Once the data activities the Agency plans on supporting have been identified, the specific data sources necessary to conduct each activity will be documented. The Agency will utilize data sources identified during the data maturity assessment discussed in the DMSP to create a list of future and current data sources for data sharing and integration. The identification or revisiting of previously identified data sources will be conducted by Data Domain Trustees and Data Stewards (defined in the DMSP) within the respective business area being assessed. The Data Domain Trustees and Data Stewards identified for each data source will receive regular communication and training regarding data sharing and integration strategies as changes are implemented by the Agency.

7.3 Skills Gap Analysis

In addition to identifying the data activities and data sources reliant on data sharing and integration practices, the Agency will also assess the overall data skills necessary to support data sharing and integration within the Agency. Skills necessary to support a business area may include extracting

data from raw data sources, loading data into the targeted data system, and transforming data into the final usable form. These processes, which are often supported or conducted by specific personnel, will be key in selecting the data sharing and integration practices for implementation; the goal is to automate or standardize as many of these processes as possible to ensure efficiency and quality within the data practice.

Skills needed may align with existing data and IT roles; **Figure 9** provides example roles and skills that the Agency may consider during the assessment. Once the list of current skills has been created, the Agency will identify the gap between skills necessary to achieve Agency goals and objectives and existing skills within the Agency. The Agency will also consider the need for redundancy in some of the skills of Data Stewards and data analysts so that the Agency reduces any risks related to reliance on one Data Steward or data analyst. The gap analysis will also inform workforce needs as well as systems or tools the Agency will require to ensure data sharing and integration.

Skills Gap Analysis—Key Steps

The following steps are used to conduct a gap analysis of the skills necessary for data sharing and integration:

1. Create a list skills provided by Data Stewards and data analysts.
2. Determine the gap between skills necessary to achieve Agency goals and objectives and existing skills within the Agency.

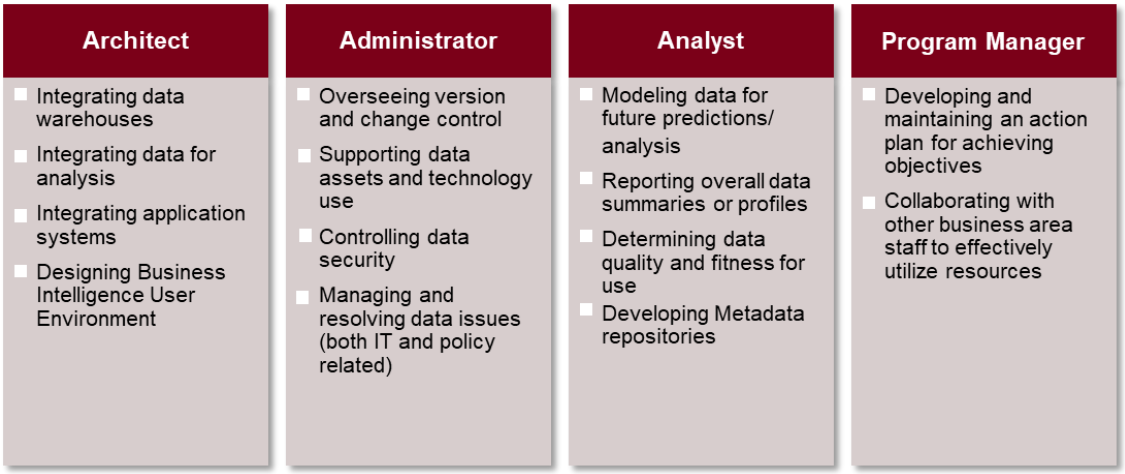


***Figure 9**. Example data skill areas*

7.4 Objectives Identification

While the overall goal of data sharing and integration is to improve enterprise-wide collaboration, it is important to further define the objectives of data sharing and data integration within the context of the business area the Agency is trying to support. Because data sharing and integration improvements can encompass a multitude of activities, the Data Domain Trustees and Data Management Committee will leverage the identified needs of the Agency to determine which data sharing and integration improvement areas the Agency should focus on. Figure 10 summarizes potential areas of improvement the Agency will consider when defining the objectives of data sharing and integration for a specific business area. Although the Agency will address each of the areas described, the purpose of this task is to emphasize which data sharing and integration activities are most important and to further inform the prioritization of action items described in the Data Management Action Plan (**DMAP**).

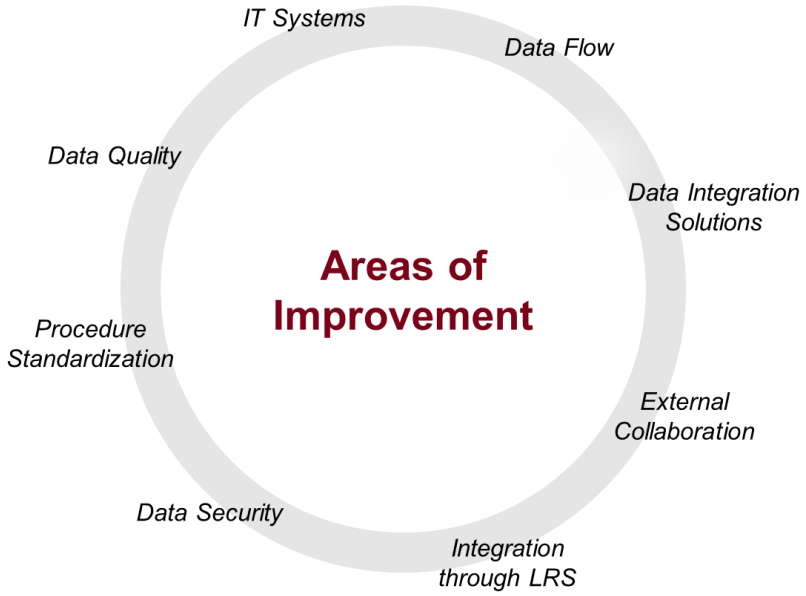


Figure 10. *Data sharing and integration improvement areas*

Each of the defined improvement areas can be used as a foundation for establishing specific objectives of the data sharing and integration process. By developing a more specific vision for data sharing and integration within the business area being considered, the Agency will better define the next steps for improving existing practices and processes.

Data Quality Assessment

Data quality refers to the ability of data to meet the needs of the data user or consumer. Within the DOT, data quality is used to describe to what extent a data source can effectively support key data activities, such as forecasting or modelling, and is a key metric for understanding how easily a data source is shared and integrated across the Agency. Therefore, to strengthen the Agency's data sharing and integration practices, the Agency will conduct a data quality assessment on all data sources identified within a business area. While data quality assessments are recommended throughout the life of a data source, the Agency will begin to assess the value of identified data sources through data profiling and the assessment of data business rules. **Figure 11** provides a summary of each step necessary for the Agency to assess data quality for data sharing and integration.

Data Quality

The ability of data to meet the needs of the data user or consumer.

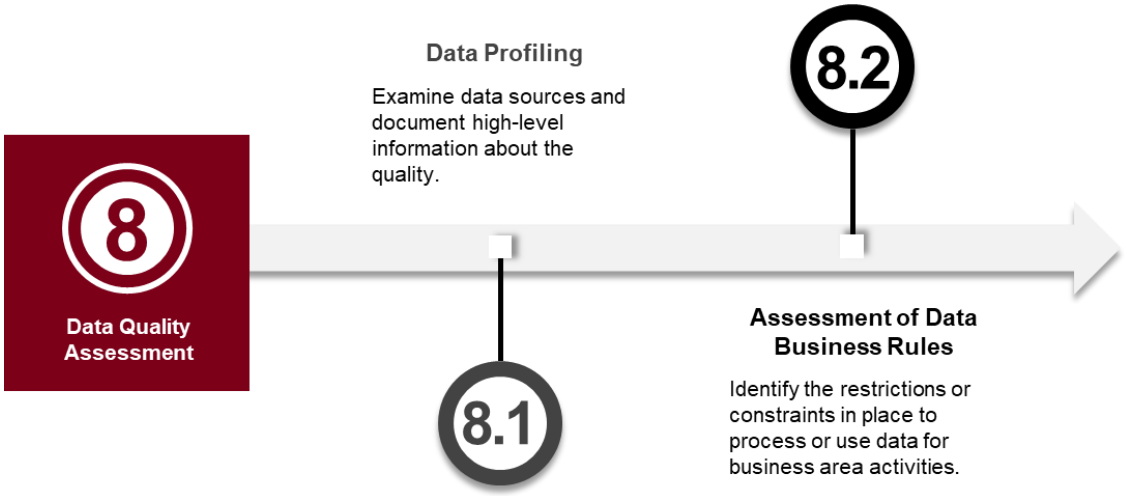


Figure 11. *Data quality assessment process*

8.1 Data Profiling

Data profiling involves the examination of data systems to better understand high-level information about a dataset. Common details exposed during profiling include column summaries (i.e., number of valid entries within a column), identification of column dependencies (i.e., the Peak Hour Factor (PHF) column depends on the average traffic volume and the maximum flow rate columns), and the identification of similarities and differences of syntax in related data sources. Most of the information identified through data profiling also exists within structural metadata.

Data Profiling

The examination of data systems to better understand high-level information about a dataset.

To properly assess the quality of the data sources identified withing a business area, the Agency will summarize key details of the data sources. The details selected will focus on information that is relevant to the data sharing and integration practices of the Agency and will be documented as supplemental information for each data source or as a part of the existing Metadata. Data profiling will be conducted by Data Stewards with input from the Data Domain Trustees.

Example—Data Profiling

Assume the Agency wants to profile a data source containing the average costs for minor bridge deck repairs throughout the State; the data source contains three columns—Project ID, Year, and Cost of Repair (per mile)—with a total of 1,500 entries. The following table provides an example of what data profiling may look like for this data source.

Element Profiled	Overview
Column Summary	• ***Project ID:*** 1,500 valid entries • ***Year:*** 1,500 valid entries, most frequently reported year is 2012 • ***Cost of Repair:*** 1,300 valid entries, average reported cost per mile is $2,500
Column Dependencies	There are no dependencies between columns within the data source.

8.2 Assessment of Data Business Rules

In addition to examining high-level information about individual datasets, the Agency will also assess the relationships or rules dictating the use of data throughout the Agency. Data business rules describe restrictions or constraints on how data is processed or used for business functions. In this task, the Agency will identify business rules and determine whether data is being properly merged, matched, created, updated, or deleted. Business rules can be both formal and informal, and therefore, the identification of the existing rules will require cooperation from both IT specialists and Data Stewards within the business area being assessed. To properly identify and assess the effectiveness of existing business rules, the Agency will utilize existing documentation on policy and procedures regarding how data is utilized and transformed throughout its life cycle. Identified business rules will be documented and assessed using a SWOT analysis.

Data Business Rules

Restrictions or constraints on how data is processed or used for business functions.

9 *IT Capabilities Assessment*

In addition to assessing the quality of the Agency's data, the Agency will also assess the capabilities of the existing Information Technology (IT) systems that support data sharing and integration throughout the Agency. IT Systems refer to data processing, storage, communications, inventory, and management utilized by the Agency. However, for the purposes of the capability assessment, IT systems are limited to the infrastructure, software, tools, and processes used to manage data. Within the Agency, the *Data Quality Action Plan* (DQAP) and the *Iowa DOT Strategic Enterprise Architecture Final Report*, provide insight on the state of IT systems throughout the entire Agency. However, while each help summarize the existing IT Systems being used, the Agency will assess the existing IT systems and capabilities to better understand the IT systems a specific business area depends on. The following tasks, described in **Figure 12,** assist in the assessment of IT system capabilities.

IT Systems

Data processing, storage, communications, inventory, and management utilized by the Agency.

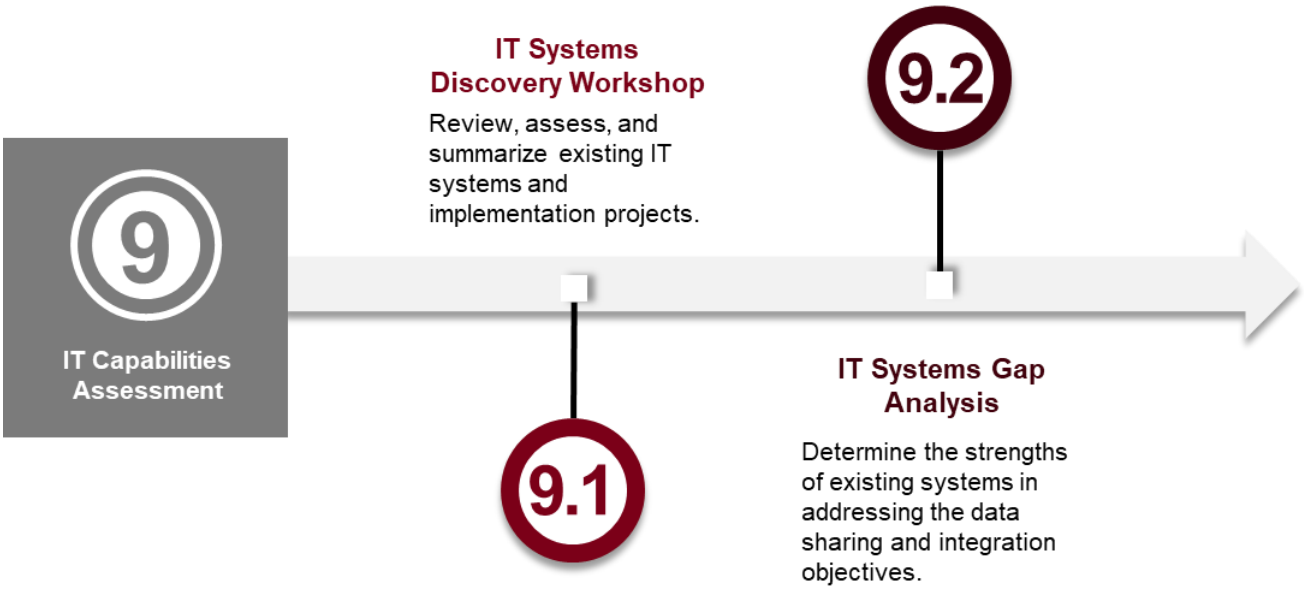


***Figure 12**. IT capabilities assessment process*

9.1 IT Systems Discovery Workshop

At the start of the assessment, the Agency will determine what currently exists in terms of IT infrastructure, policies, and procedures. The Agency will identify key stakeholders to conduct the discovery activities, including members of the IT Governance Board (when commissioned), the Data Management Committee, Data Domain Trustees for the business area being explored, and the appropriate Data Stewards. The final team of individuals will be responsible for conducting an in-depth systems discovery workshop for the business area being assessed.

The discovery workshop will be two-fold: the team will review, assess, and summarize existing documentation on IT systems and implementation projects conducted by IT and determine the strengths of these existing systems in addressing the data sharing and integration objectives defined in Task 7.4.

Summary of Existing Documentation

The first part of the workshop will focus on reviewing any documentation the Agency has on existing IT infrastructure, policies, and procedures. Relevant documents include the *Iowa DOT Enterprise Architecture Final Report* and any existing documentation on the flow of data within the organization. Additional information on the IT systems, including informal knowledge provided by data users and IT staff, should also be discussed at this time. Finally, the discovery team will discuss any current IT efforts or implementation projects that are ongoing; the DQAP is one such example.

Discussion of Strengths of Existing IT systems

Based on the process described, the Agency will identify how the existing IT systems for a specific business area work. Specifically, members of the discovery team will be able to describe how data is transformed throughout its useful life and the strengths of the existing IT systems.

9.2 IT Systems Gap Analysis

Based on the review of existing documentation and information provided by area experts, the discovery team will identify the key areas of improvement for existing IT systems. Areas of improvement will focus on aligning the strategic objectives of the Agency for data sharing and integration with proposed changes to existing data storage, integration, and access. A formalized maturity assessment using the guidance and tools being developed under the National Cooperative Highway Research Project (NCHRP) 08-115 is recommended during this stage.

10 *Data Flow Assessment*

Another necessary step in contextualizing the current state of data sharing and integration practices within the Agency is reviewing, refining, and implementing a data flow between the defined data sources and IT systems and tools. For the Agency to identify areas of redundancy or inefficiency, a better understanding of the data flow, specifically data consumption and data transfer, is necessary. The following activities, summarized in **Figure 13**, provide a path for identifying, validating, and improving upon the Agency's existing data flow.

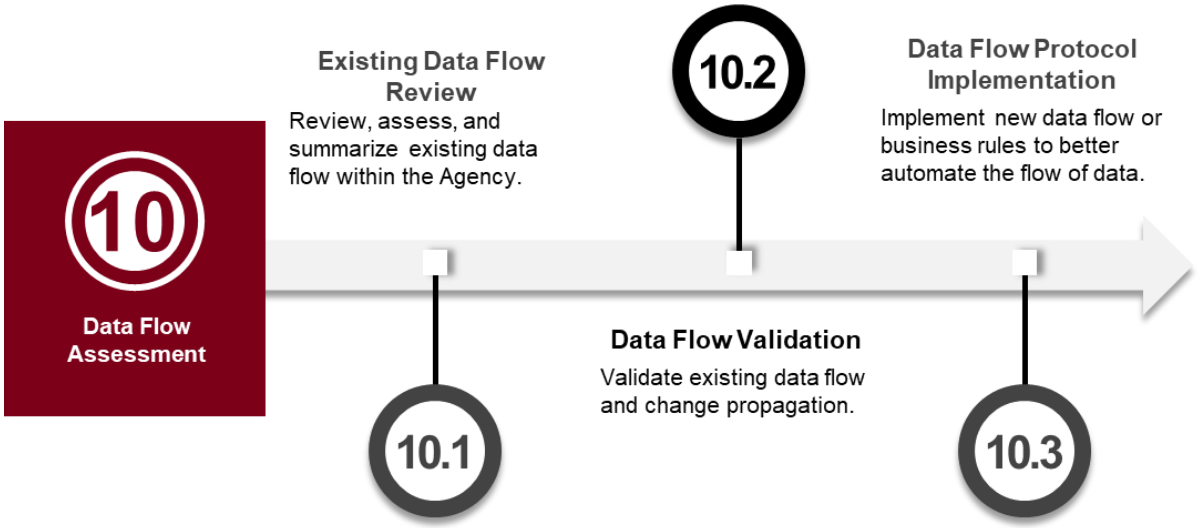


Figure 13. *Data flow assessment process*

10.1 Existing Data Flow Review

A data flow diagram or data lineage describes where data comes from or how systems and data are related. Data flow diagrams can be used to identify how data moves from collection to storage to application as well as the databases, tools, or rules utilized in the movement of the data. For the purposes of creating a clear representation of how data within a business area moves, the data flow diagram should identify the detailed movement of data such as the name of database where data is stored, the name of applications critical for transforming data, and the types of transformations that occur. The desired data flow diagram will focus on eliminating duplicated data processes and the replication of data sources.

Data Flow Diagram

A visual representation of where data comes from or how systems and data are related.

The current flow of information within the Agency has been documented and discussed through multiple Agency efforts. As previously mentioned, the Agency is currently developing a DQAP, which will identify the Agency's data management infrastructure and how data will be stored and transferred internally and externally. **Figure 14** shows the conceptual flow of data as documented by the Agency. The data systems are divided into three distinctive domains—data systems available to the wider Agency and public, data systems that are maintained and operated by IT, and data stores where full datasets can be extracted and transformed using analysis tools. Central to the data flow diagram is the Master Data Management system (MDM) maintained by IT. The MDM is a central location for data to be extracted, loaded, or transformed. Therefore, the MDM extracts and loads data from both the data stores and the public or internal data systems.

While the efforts of the DQAP have helped identify the flow of data between data systems, **Figure 14** provides minimal information on how specific data sources interact with each other. Detailed information on the interaction of specific datasets or applications is important for understanding the effects of naming conventions, data transformations, and other changes made to the data sources when new standards or procedures are implemented. The Agency, therefore, developed an additional diagram to describe the flow of data at the activity level. As depicted in **Figure 15**, data software or systems are colored based on the origin of the tool or system. The Agency reports five primary software types used to conduct most daily activities—commercial off-the-shelf, AASHTOWare, custom-built, externally provided, and provided through a 3rd party. While the figure highlights relationships between data within the pavement and bridge practice, similar data flow diagrams can be established for other business areas and data sources.

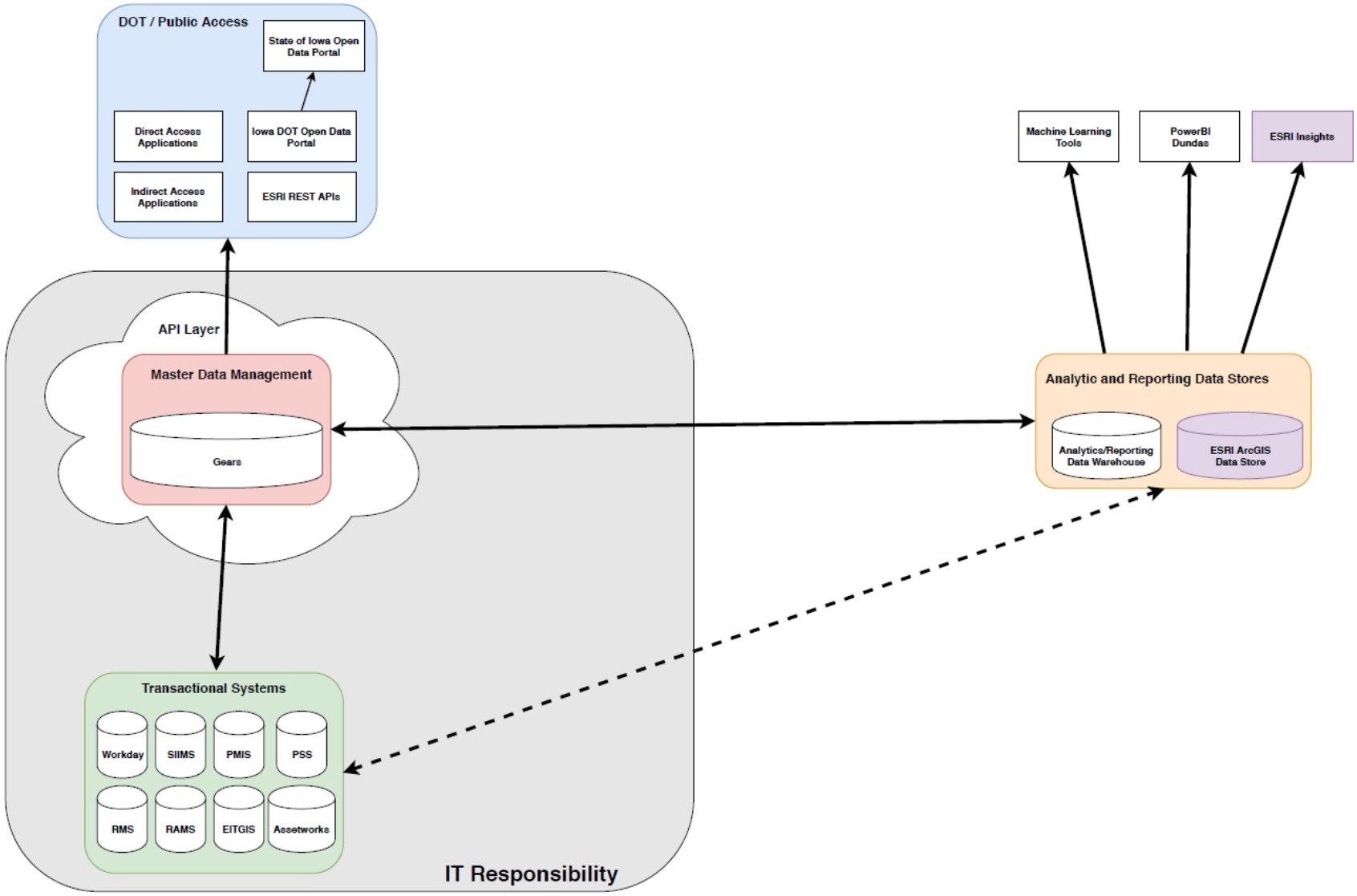


Figure 14. *Data management systems flow*

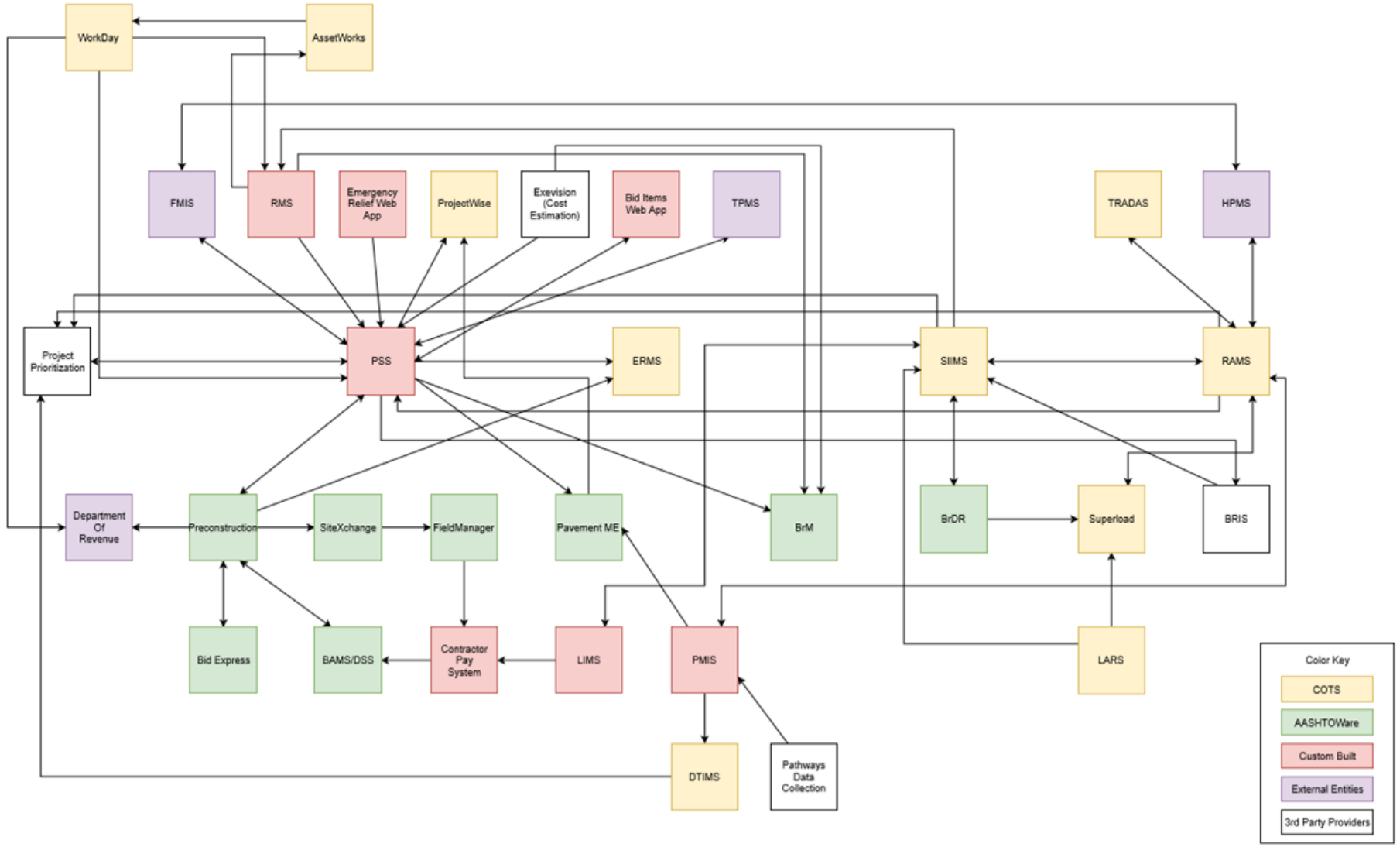


Figure 15. *Granular-level data activities flow*

An oversight team will further review the data flow diagrams to ensure each will be effective in achieving the objectives of the data sharing and integration process. The group will identify whether additional information on how data is shared and transformed between data tools and systems is necessary and whether the origin of specific data sources identified for a business area need to be captured in the existing data flow.

10.2 Data Flow Validation

Based on the review of the existing data flow diagrams, the Agency will refine and update the flow diagrams and document the protocols needed to address the identified areas of improvement. Changes will be made by IT personnel, but refinement will remain a collaborative effort among stakeholders. In addition to updating the data flow diagram, the Agency will also validate that changes made to the data flow align with the goals and needs of the business area.

Data Validation

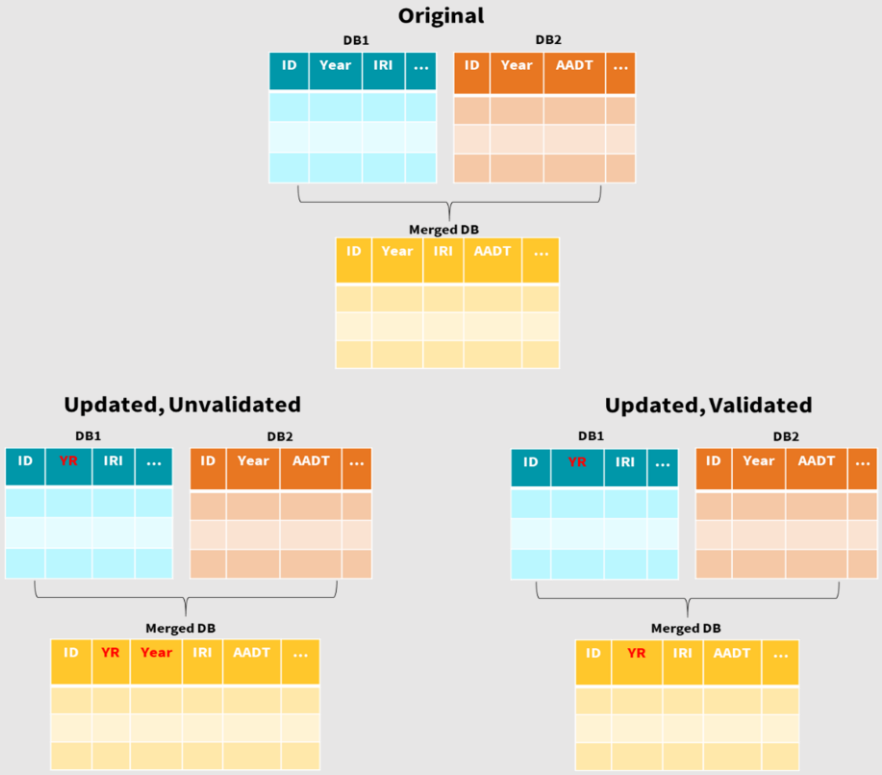
A way of checking the accuracy and validity of data used to better understand potential conflicts in the data system.

Data validation is a way of checking the accuracy and validity of data used to better understand potential conflicts in the data system. One example of validation is to intentionally change the name of one of the attributes in one data source, but not in a related database. If the proper business rules were implemented into the system, the name change in one data source would either propagate to data sources that depend on the data source through a series of transformations or the change in name would have no effect on data sharing or linkages between the data source that was changed and all related databases. If the expected response does not occur, the data flow and business rules will be reassessed and refined until the expected outcomes occur. A real-world example of the validation process is presented on the next page.

Example—Validation

Assume the Agency is conducting a study on the relationship between traffic and bridge condition within the State. The analysis requires both traffic data and pavement condition data. Database 1 (DB1) contains attributes related to pavement condition, and Database 2 (DB2) contains attributes related to traffic. The tables have two attribute fields which are used to merge the data—ID and Year. A Data Steward of DB1 decides to update DB1 with abbreviated headings. Therefore, the Year attribute in DB1 is changed to YR; DB2 remains the same.

If the data flow and data business rules for the data sources are properly established, then the changes made to the Year attribute in DB1 will not affect how DB1 and DB2 are merged. However, if the relationship between the two databases is not properly reported, the two databases will not be properly merged. To validate the change made to the Year attribute, the two databases will be merged. In an unvalidated, merged database, both the YR and Year attributes are kept because the relationship between two databases, regardless of the attribute names, was not established. In a validated, merged database, only one attribute for the year is kept (either YR or Year) and the merging process remains unchanged.



10.3 Data Flow Protocol Implementation

Once the data flow and business rules have been established and refined, the Agency will implement the updated data flow and business rules. The implementation process will be done iteratively by IT personnel and will be effectively monitored and communicated throughout the Agency. The Agency will aim to automate the monitoring of issues that may arise during data processing or extraction; however, human monitoring, conducted by IT, will also be necessary. Monitoring of issues related to data sharing and integration will be a top priority of IT, and appropriate responses to identified issues will be both prompt and well communicated internally.

Data Integration Solutions

Database integration focuses on storing, organizing, and sharing multiple datasets to best support the Agency's business needs. As part of the DQAP, the Agency is currently developing and implementing an MDM, which will provide a structure for data integration and related activities. However, the Agency will require multiple integration solutions to support all the key data-related activities. Therefore, the Agency will need to evaluate and implement the most effective data integration techniques for long-term benefits. In this section, the process of reviewing and recommending data integration solutions is explored. **Figure 16** provides an overview of the process.

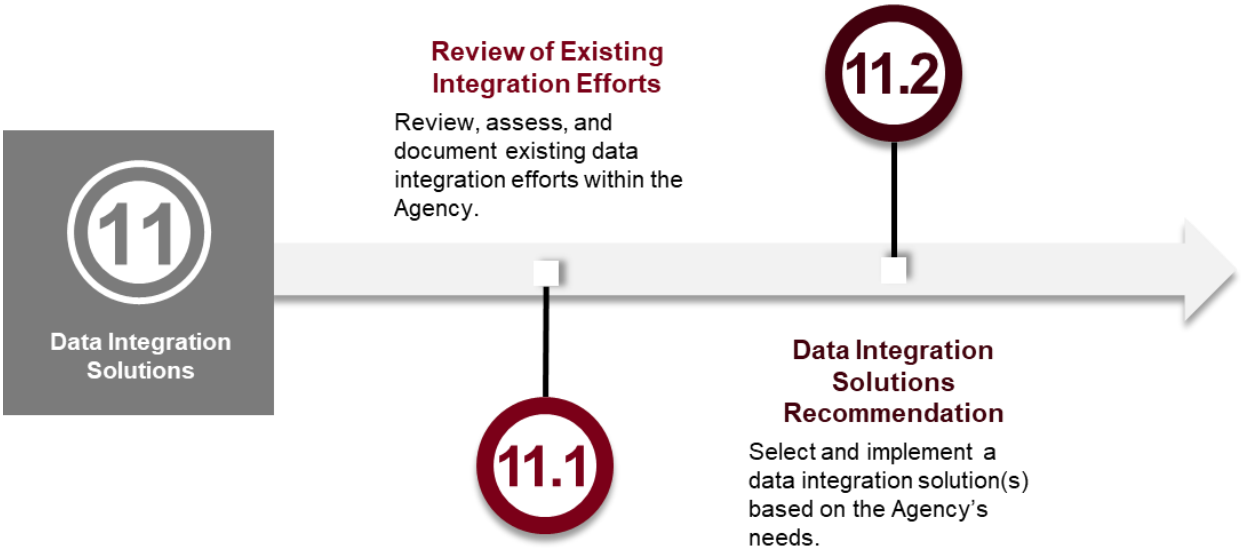


Figure 16. *Data integration solutions process*

Review of Existing Integration Efforts

Prior to selecting data integration techniques for adoption, the Agency will review existing data integration solutions that are feasible for implementation. IT and the Data Management Committee will conduct a comprehensive review of potential integration solutions based on industry practices and internal knowledge. In this section, three integration solutions common throughout the industry—data consolidation, data propagation, and data virtualization—are described and will be used as a starting point for the Agency's review. **Table 3** provides a summary of the pros and cons of each technique.

Table 3. *Summary of integration techniques pros and cons*

Technique	Description	Pros	Cons
Consolidation	Aggregation of multiple data sources to a single location	• Less complex data system/easier to update and maintain. • Eliminates issues of data redundancy through data centralization.	• Highly sensitive to network connectivity and the number of system users accessing data at one time. • Increased risk or loss if security breach occurs

Technique	Description	Pros	Cons
Propagation	Derivation of data from one or multiple data warehouses to local versions or instances	• Data maintained and updated locally. • More resilient to database "outages" because data is stored in multiple locations.	• Complex system requiring additional upkeep and experience to manage.
Virtualization	Utilization of a single, physical system with virtual instances provided directly to the user	• Reduced system costs because of virtualization. • Less complex system/easier to update and maintain.	• Reliant on internet connection. • Increased risk or loss if security breach occurs.

Data Consolidation

Data consolidation involves the aggregation of multiple data sources to a single location; data users refer to that location to access data as depicted in **Figure 17**. The technique is advantageous for updating, analyzing, or reorganizing data the users receive, and it also eliminates issues of data redundancy through location singularity. However, the data consolidation technique does incur risks. A centralized data location, which is created through consolidation, is highly sensitive to network connectivity and the number of users trying to access the database. Slow connections or limited copies of data sources will lead to restricted access for data users. Additionally, as all instances of data rely on a single database, the risks associated with a security threat or security breach are determinantal to the Agency using data consolidation as the main integration strategy.

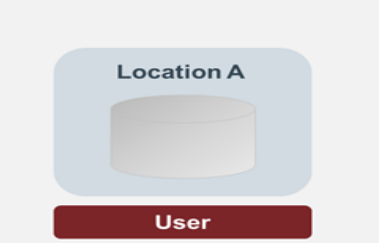


Figure 17. *Data consolidation*

Data Propagation

Data propagation refers to an integration strategy where data is derived from one or multiple data warehouses and propagated to local versions or instances as depicted in **Figure 18**. One of the most common types of data propagation is the implementation of a federated database system. The federated database maps multiple data sources of interest into a single federated database. For example, suppose a data user required pavement condition information and environmental hotspot data to perform an analysis. In a federated system, a new database would be temporarily created that maps both datasets into a single location. The datasets and locations of the datasets remain separated, but the mapped instance provides a shared location for users to manipulate the data.

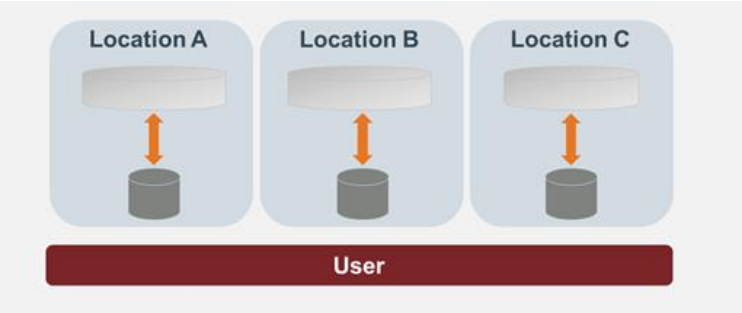


Figure 18. *Data propagation*

The benefits of data propagation are related to the decentralization of data. With separate databases for different datasets, data can be maintained in the Bureau or Division most familiar with the data. Additionally, because data is distributed throughout multiple locations, if one data location is down or

undergoing maintenance, the user can still access all other datasets not within that location. However, the data propagation requires additional upkeep and experience to manage as the multiple locations create a more complex data system. Security and management will also require additional resources.

Data Virtualization

Data virtualization is another commonly implemented data integration technique. Through cloud platforms, software, data, and storage can be hosted through a single, physical system with virtual instances provided directly to the user. Data virtualization provides economic benefit as software and data systems can be pooled from a single source and maintained or modified more easily. The limitations of virtualization are similar to data consolidation. As instances of the physical system are typically hosted online, interruptions to internet access will limit access to all users in the network. Additionally, as all instances of data rely on physical databases, the risks associated with a security threat or security breach are more determinantal to the Agency. **Figure 19** depicts what data virtualization may look like in the Agency,

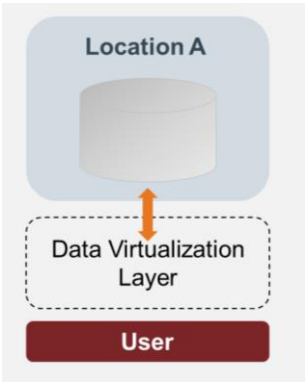


Figure 19. *Data virtualization*

11.2 Data Integration Solution Recommendation

After evaluating the potential integration solutions and assessing the benefits and limitations of each, the Agency will consider which technique best serves the objectives of the Agency given the limited resources available. Through a workshop, the IT Governance Board will facilitate a conversation with the DMC about the best integration solutions for the Agency moving forward. Due to the costs associated with adopting new integration solutions, the Agency will first discuss whether the existing management systems will be adequate in the long run. The group will make recommendations and communicate them to the executive leadership to ensure appropriate buy-in.

12 *Linear Referencing System Assessment*

Within the Agency, the Linear Referencing System (LRS) plays a crucial role in integrating location-based datasets for analysis and visualization; LRS is an essential tool for sharing and integrating disparate data sources. Since 1998, Agency has recognized the value of a Linear Datum and has since used a LRS to integrate data utilizing Geographic Information Systems (GIS). While the LRS has matured since its initial implementation, the Agency will periodically assess the effectiveness of the LRS used and identify areas of improvement to further the Agency's goals of stronger data architecture, integration, and collaboration. Through the tasks identified in Figure 20, the Agency will periodically assess the existing LRS and develop strategies for improvement.

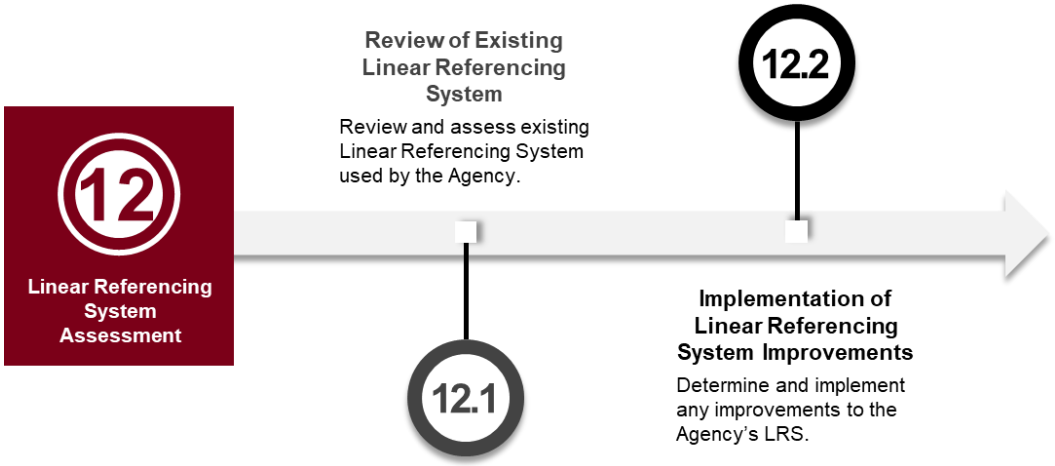


Figure 20*. Linear Referencing System assessment process*

12.1 Review of Existing LRS

While the Agency's existing LRS is both robust and widely accepted throughout the Agency, it is important that the Agency periodically reviews the existing system and identifies areas of strengths and opportunities for improvement. In doing so, the Agency will continue to mature the existing data architecture and integration practices. During this task, the Agency will focus on reviewing the existing LRS and systems that utilize the LRS through a Strength, Weaknesses, Opportunities, and Threats (SWOT) analysis. The SWOT analysis, conducted by the Data Management Committee, IT Governance Board, and Analytics staff, will occur in a workshop format and consider how the existing LRS compares to the best practices of industry. As a starting point, the Agency will compare the existing LRS to the best practices identified in NCHRP 814.

Best Practice—LRS[1]

A mature LRS is focused on:

- Use of a common LRS for data within the Agency,
- Establishment of an LRS quality protocol,
- Creation of infrastructure to integrate non-spatial data with spatial data,
- Automation of change propagation from the LRS to all related databases, and
- Collaboration of data managers and GIS staff to improve the consistency of the LRS.

[1]Harrison, F. D. (2015). *NCHRP Report 814: Data to Support Transportation Agency Business Needs: A Self-Assessment Guide.* Washington, D.C.: Transportation Research Board of the National Academies.

12.2 Implementation of LRS Improvements

Based on the assessment of the LRS, the Agency will determine key action items to further improve the Agency's LRS. Key action items may include establishing a standard procedure for monitoring and assessing the quality of the LRS, the identification of key strategies to better integrate new non-spatial data using the LRS, or creating an annual meeting for the Data Management Committee to discuss potential improvements or updates to the LRS with the Analytics staff. Action items that result from the SWOT analysis will be documented and tracked continuously as a part of the Agency's action plan. The Agency will prioritize action items based on the level of effort and importance of each action items as discussed in the (**DMAP**).

13 *Data Security Assessment*

Data security focuses on reducing risks, such as a data ransom attack or improper access to secured data, by enforcing established privacy, legal, and business agreements. The tasks associated with ensuring data security include identifying governing requirements for individual data sources, conducting a security assessment of existing procedures, creating or updating security policies and standards, and communicating and implementing the established security protocols throughout the Agency as depicted in Figure 23.

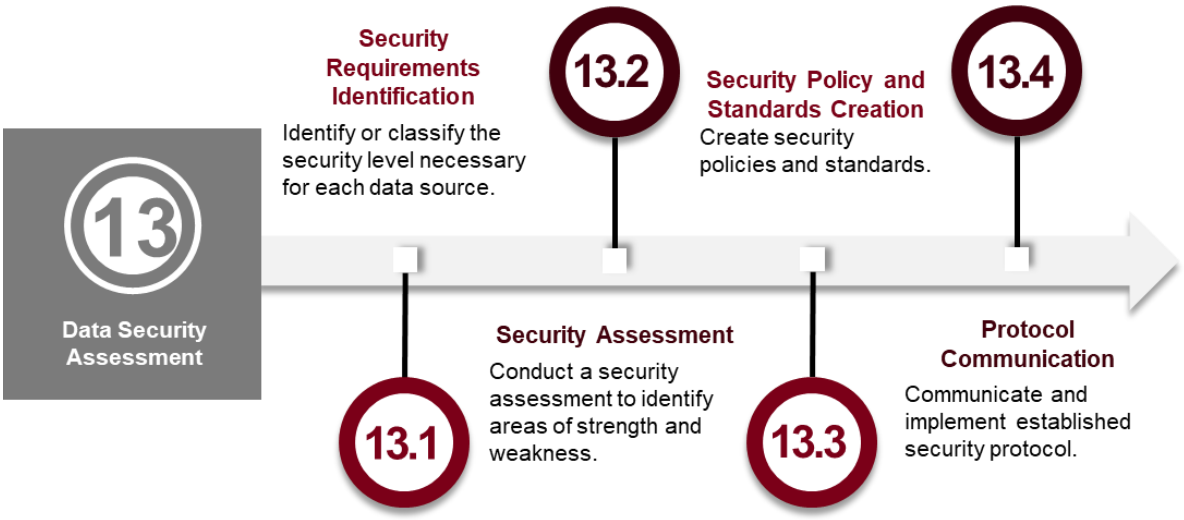


***Figure 21**. Data security assessment process*

13.1 Security Requirements Identification

At the core of data security is the identification or classification of data sources based on the sensitivity or confidentiality of the information provided. The confidentiality of a data source is driven by both internal and external regulation on the accessibility of a dataset. The confidentiality of a data item or data source can be classified as one of three levels—low, moderate, or high. **Figure 22** provides an overview of what each level means with respect to data owned or used by the Agency.

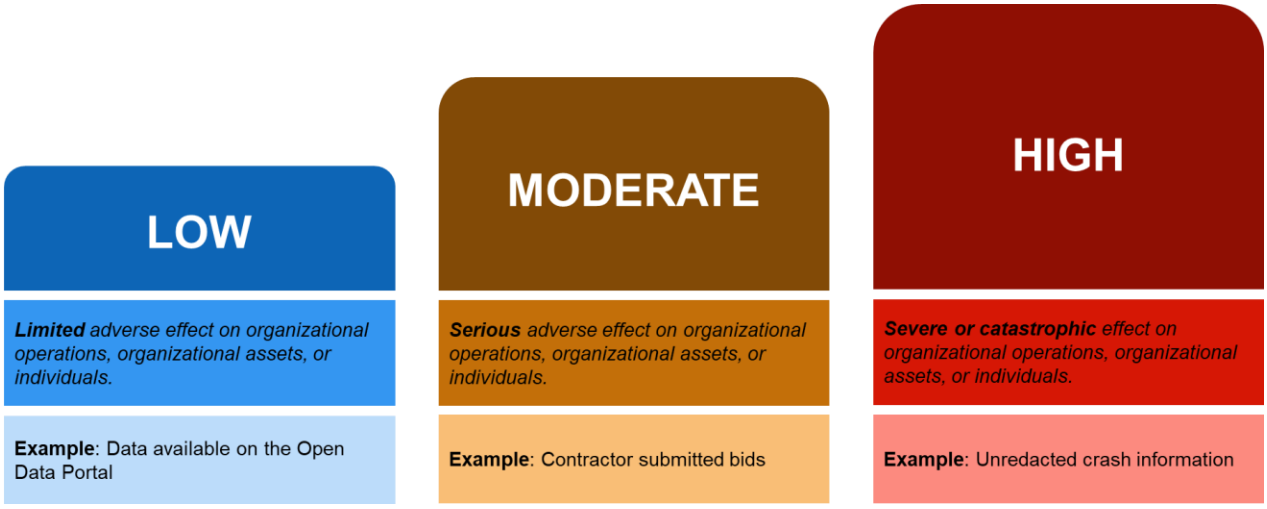


***Figure 22.** Levels of data confidentiality*[6]

Prior to assessing the existing security policies and standards, the IT Governance Board and Data Domain Trustees will assess (or reassess if the process has been previously completed) the data confidentiality of each of the data sources identified within the business area. The level of confidentiality assigned to each data source is the *most restrictive* level of confidentiality of the corresponding data attributes that make up the data source. The confidentiality of each individual attribute is based on the type of information the attribute provides, the subsequent risk the data poses to the Agency if released to the public, and the existence of policies (international, federal, or local) that restrict the accessibility of the information. During the creation of rights metadata, the IT Governance Board and Data Domain Trustees

[6] U.S Department of Commerce. (2004). Standards for Security Categorization of Federal Information and Information Systems. *Federal Information Processing Standards Publication.*

will recommend a level of confidentiality. By identifying the level of confidentiality or security necessary for a data source within the metadata, the Agency will more easily regulate and monitor data security compliance.

Example—Assessing Data Source Confidentiality

Assume the Agency is assessing the confidentiality of a data source containing six attributes. The risk levels of the six attributes are as follows:

- *Attribute 1*: Low
- *Attribute 2*: Moderate
- *Attribute 3*: Moderate
- *Attribute 4*: Low
- *Attribute 5*: High
- *Attribute 6*: Low

Based on the level of risk associated with each attribute, the data source will be assigned a High level of confidentiality which corresponds to the most restrictive attribute of the data source (Attribute 5).

13.2 Security Assessment

Security or risk assessments require IT staff to identify the existing state of data security within the Agency. While a security assessment can be conducted at an enterprise-level if resources are available, IT personnel should conduct a security and risk assessment for the IT infrastructure of the business area being evaluated. The procedure for conducting the security assessment will depend on the scope, resources, and goals of IT, as the process is typically conducted with limited input from other members of the Agency (although the Data Management Committee will receive updates on the process). The security assessment process will draw upon findings from the *Iowa DOT Strategic Enterprise Architecture Final Report* completed by the Agency in 2016. Recommendations from the assessment will inform decisions made about security polices and standards.

13.3 Security Policy and Standards Creation

As the Agency develops a better understanding of the existing security protocols implemented within a business area, it will also determine the policies, standards, and tools necessary to improve these existing data security protocols. The creation or updating of security polices and standards will involve the IT Governance Board, the Data Management Committee, and Data Domain Trustees, as the policies and standards will serve both IT systems and specific data sources within the specified business area. Collaboration and communication between IT and business area experts will help create robust security protocols.

IT-specific policies will provide technical rules on how data should be organized for security reasons and where password protection is required. During the creation of IT data security policies, the Agency will also identify Security Officers for the enforcement of data security policies. The creation of the position was a critical need identified in the *Iowa DOT Enterprise Architecture Final Report* conducted by the Agency in 2016. In addition to IT security policies, the Agency will also identify data-specific security policies. Data-specific security policies apply to specific software or databases for regulatory or business reasons (i.e., data must be converted into a .csv file to be used in an analysis software). Data specific polices will identified by IT staff and Data Stewards.

Data Security Policy Types

IT-Specific Polices: Technical rules on how data should be organized for security reasons and where password protection is required.

Data-Specific Policies: Policies that apply to specific software or databases for regulatory or business reasons.

13.4 Protocol Communication

Once the Agency has created or updated the security policies, the implementation phase will begin. While the enforcement of these policies and standards will

largely be the role of the identified Security Officers, the Agency will need to communicate any new or updated policies to data users throughout the Agency; communication via email campaigns and training videos is recommended. Given the importance of data security to the Agency's overall resiliency and efficiency, communication and training will be a priority of the Agency.

External Collaboration

In addition to developing a strong internal culture of data sharing and integration, the Agency also benefits from establishing relationships with external entities and stakeholders that may provide data or leverage Agency-produced data for analysis or communication. External stakeholders, which can include both public and private agencies, strengthen the Agency's existing data practice through the data each produces or leverages. However, for these partnerships to be effective, a procedure for sharing data with external stakeholders needs to be developed. **Figure 23** details key actions for identifying stakeholders and developing an engagement plan to enable and strengthen external collaboration.

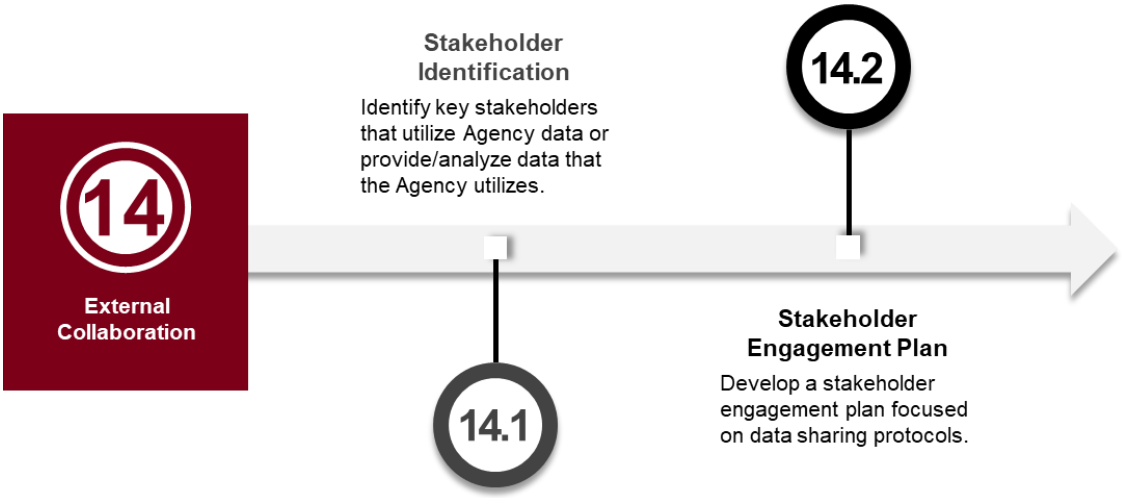


***Figure 23**. External collaboration process*

Stakeholder Identification

Collaboration with external agencies is centered on eliminating duplicated data processes and leveraging existing data and analysis by sharing resources valued by key stakeholders. Therefore, identifying key external stakeholders is an important first step in standardizing the data sharing process. External stakeholders will vary by business area, but may include institutes of higher education, other government entities (such as federal and local agencies and legislators), or private industry (including the media). The Agency will develop a list of external stakeholders based on the types of data and types of analysis the Agency needs or that external stakeholders typically leverage. Candidates for external partnerships include entities currently collecting data the Agency uses or would like to use, entities with the capability of conducting analyses the Agency currently conducts but in a more efficient, cost-effective manner, or entities the Agency has previously provided data or analysis to in the past. The Data Management Committee and Data Domain Trustees will identify and document the potential external stakeholders for further review by the Executive Board and IT oversight.

Example—Identifying External Stakeholders

A university within the state of Iowa is creating a small-scale inventory on signage retroreflectivity to support research. The data is collected on a State-owned route and therefore, includes data on signage that is owned and maintained by the Agency. Because the data collected by the university would also be useful for maintenance decision-making within the Agency, the university research team is identified as an external stakeholder.

14.2 Stakeholder Engagement Plan

With potential external stakeholders identified by the Agency, the next step in developing external collaboration is to create a stakeholder engagement plan. The Agency will identify one individual, knowledgeable in data or services that an external organization provides or leverages, to act as a liaison between the stakeholder and the Agency. The selected individual will be the point of contact for the external organization and will help facilitate conversations with the interests of the Agency in mind. Potential candidate s to serve in this role include Data Domain Trustees or Data Stewards.

Following the identification of Agency liaisons, the Agency will develop the stakeholder engagement plan. While stakeholder engagement plans vary in content, the Agency will develop a strategy for creating a Memorandum of Understanding (MOU), determining meeting type and frequency needed for each stakeholder, and developing data sharing protocols. Issues related to security, system compatibility, and existing governing policies developed by the Agency need to be assessed for external collaboration to be effective.

15 *Procedure Standardization*

In addition to understanding and documenting data sharing and integration processes for the existing data sources, the Agency will also develop procedures for new data sources and data software. One of the areas of opportunities identified during the SWOT analysis conducted in 2019 was the oversaturation of data-related products within the Agency; these products are not always vetted for system compatibility or functional redundancy. While new data and data software provide the Agency with the opportunity to expand or improve business practices, the new data and software need to communicate and improve what the Agency already supports. This section focuses on standardizing the evaluation and integration of new data, applications, and system acquisitions to avoid redundancy and incompatibility that will limit the effectiveness of adopting new data or technologies. **Figure 24** describes the key steps in developing a standardized procedure for the addition of new data or applications.

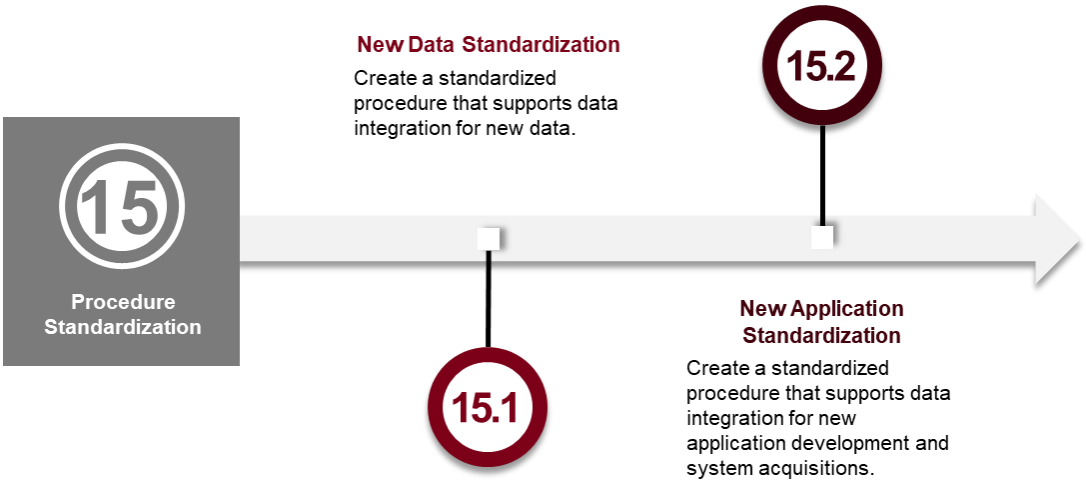


Figure 24. *Procedure standardization process*

15.1 New Data Standardization

The addition of new data into the existing system is a process that builds on the procedures described throughout this plan. A summary of the process is provided in the subsections that follow.

Identify what data is being collected and what purpose it serves

The first step towards integration of new data is the identification of what data is being collected and why the data needs to be collected. Because data collection or acquisition and integration are costly data management activities, the Agency needs to be certain that the data proposed to be collected is not redundant and that it provides value to the Agency. Developing a business case describing what is being collected (and in what format) and why the data is important for business functions of the Agency is

critical to the process. If the Data Management Committee and executive leadership find the business case for the new data compelling, the Data Management Committee and Data Domain Trustees will determine where the data should be mapped within the existing IT system.

Create appropriate data standards

Data standards enable data to be robust and resilient to changes made within the data management system. Before data is added to the Agency's IT systems, the Agency will follow the procedure for the creation of data standards creation discussed in Tasks 1-6. While the identification of existing standards and policies for the new dataset is necessary, the creation of metadata is crucial to integrating new data into the existing data system. Through the creation of metadata, the format of the new data and the relationship of that data to other datasets will be further identified.

Update the existing data sharing and integration procedures

The final step in integrating new data into the existing system involves the execution of the recommended integration strategies developed throughout this section. With the newly created or acquired data, the Agency will establish the flow of the data from collection or acquisition to its corresponding business activity, integrate the data into the Agency's LRS and data integration system, and update any data business rules or policies based on the addition of the data into the system. These tasks, which are detailed throughout the plan, enable the data to be smoothly shared and integrated using the plan's standardized method. Once the data is fully embedded into the Agency's data architecture, the updated data flow and integration will be communicated to relevant data users.

15.2 New Application Standardization

The procedure for integrating a new application or software into the existing data system follows a similar procedure to the one identified for new data. The first step of integrating a new application is to identify the purpose and requirements of the application or software. The Data Management Committee and Data Domain Trustees will investigate whether the application is necessary and whether it provides a unique perspective or analysis of data that is not provided by other applications or software. Additionally, the requirements of the application or software will also need to be thoroughly reviewed. The Agency will identify whether the format of the application is compatible with the existing system or whether the data it utilizes needs to be transformed. Once the purpose and requirements have been identified and the application has been approved, integration can occur using the methods described throughout this plan.

16 *Data Sharing and Integration Plan*

The final task in establishing a rigorous data sharing and integration procedure is to document the processes and actions in a formal data sharing and integration plan. The plan will detail existing data, IT systems, and the data flow within the Agency, recommend sharing and integration solutions, LRS improvements, and data security protocols, and define the procedures for supporting external collaboration and new data collection and application acquisitions. Because data sharing and integration is an agency-wide initiative, the plan will be developed by the Data Management Committee and IT with input from the Executive Board and Data Domain Trustees. The scope of the plan ultimately will be agency-wide, but its development will likely begin by considering data sharing and integration through the lens of one data business area. As additional business areas are considered, the data sharing and integration plan will be modified and refined to meet the changing needs of the entire Agency.