REVIEW OF FREIGHT DATA CLEARINGHOUSES

JUNE 2016

FDOT

SECTION 4

REVIEW OF FREIGHT DATA CLEARINGHOUSES

ATION MATRIX OF DATA CLEARINGHOUSES

	State	State	Multiple years of data	Exportable files/formats	Query options personalize	Roadway Data (AADT and # of lanes minimum)	Basemap of roads	Georeference information	Tonnage, Value, TEUs	Intuitive Interface	Interactive	Navigation
	Flagship GIS - cel Data	AL	1	1	1	0	1	1	0	1	2	2
	aView - Remote and GIS Shapefiles	AL	1	1	0	0	0	1	0	1	2	2
	State Geo-Spatial ta Clearinghouse	AK	0	1	0	0	1	1	0	1	2	2
	rizona State Land Department	AZ	2	1	1	1	1	1	0	1	1	1
	University of Arizona braries GIS & Geospatial Data	AZ	0	1	1	0	0	1	0	1	2	2
	Arkansas GIS Office	AR	0	1	1	1	1	1	0	1	2	2
	Cal-FRED: The Online California Freight Data Repository	CA	2	1	0	1	1	1	1	1	3	3
	State of California GeoPortal	CA	2	1	0	1	1	1	0	1	2	2
	University of Connecticut GIS Data	CT	2	1	0	0	1	1	0	1	2	2
10	State of Delaware	DE	2	1	0	0	1	1	0	1	2	2
11	Florida Housing Data Clearinghouse	FL	2	1	1	0	0	1	0	1	3	2
12	Florida Geographic Data Library	FL	2	1		1	1	1	0	1	2	2
13	Florida Department of Transportation Statistics Office	FL	0	1	0	1	1	1	0	0	2	2

4.0 REVIEW OF FREIGHT DATA CLEARINGHOUSES

The Review of Freight Data Clearinghouses provides analytical and technical recommendations for freight data clearinghouses, data fusion analyses, and data visualization to increase the accessibility and marketability of multimodal freight data. This section contains the following four subsections:

- Background of Data Clearinghouses;
- Guidelines for Evaluation of Data Clearinghouses (further detailed in Appendix E);
- Data Clearinghouse Evaluation (further detailed in Appendix F); and,
- Data Fusion Analysis.

The first subsection (Section 4.1) provides a background on data clearinghouses including major components of a data clearinghouse, which include: servers, datasets, metadata, search engines, collective metadata, advanced query functions, and data architecture. According to the Merriam Webster dictionary, a data clearinghouse is defined as: *"A central agency for the collection, classification, and distribution especially of information; broadly: an information channel for distributing information or assistance".* The first subsection also provides information regarding the inventoried datasets and data architecture.

Inventoried Datasets: The Review of Freight Data Clearinghouses explored 89 major freight datasets. These datasets were heterogeneous in nature and can be differentiated in three forms: structure, syntax, and semantics. The inventoried data clearinghouses were categorized using a suitable classification method to determine their strengths and weakness. The utilization of a suitable classification method facilitated structured review of clearinghouses and ensured data architecture was consistent. *Data Architecture:* Based on literature review of national sources, Section 4.1 provides a comprehensive analysis of the different components of an ideal data architecture framework.

The second subsection (Section 4.2) focuses on selected guidelines for evaluation of data clearinghouses. The first set of guidelines are adopted from Data Archiving and Networked Services' (DANS) methodology for "Data Seal of Approval". DANS is an institute of the Royal Netherlands Academy of Arts and Sciences (KNAW). The evaluation of data clearinghouses also includes an intuitive methodology developed as part of this task. The guidelines utilized in this analysis as well as other important guidelines for evaluating data clearinghouses are provided in Appendix E.

The third subsection (Section 4.3) applies the guidelines described within Section 4.2 to evaluate 52 different data clearinghouses. A brief background of the process followed for the evaluation and major findings are summarized within this subsection. The complete rating evaluation matrix is provided in Appendix F.

Section 4.4 provides a discussion regarding the importance of data fusion within freight data analysis. Data fusion is an intuitive approach but utilization of a structure approach is imperative in assisting freight data analysts in choosing the most appropriate method to conduct data fusion. Several examples are highlighted within this subsection to provide better understanding of data fusion methods.

4.1 BACKGROUND OF DATA CLEARINGHOUSE

Freight data clearinghouses provide significant benefits for users who need to access data sources and related online services for freight analysis and visualizations. With majority of data sources available at one centralized location (clearinghouse), users can reduce their time and efforts in acquiring data, analyzing and assessing data and then changing the data format to suit their needs. It represents a potential solution to alleviate the data assessment, collection, and visualization burdens on analysts and modelers which will enable them to spend their time to provide cost-effective solutions for policy-makers at regional and local levels. Data clearinghouses also provide a benefit for private sector industry data analysts who use similar information for economic and financial optimization of their services. Examples include: Fleet routing, trip planning, obtaining permits, and performance measures for the transportation system.

4.1.1: What is a Data Clearinghouse?

According to the Merriam Webster dictionary, a data clearinghouse is defined as: *"A central agency for the collection, classification, and distribution especially of information; broadly: an information channel for distributing information or assistance".*

The major components of data clearinghouse are explained below:

- *Servers:* A data clearinghouse comprises of a distributed system of agency servers which are located on the Internet and they contain field-level descriptions of available and planned datasets, applications, and services.
- *Datasets:* The datasets included in the clearinghouse form the main component of the data clearinghouse. They are inventoried and then may or may not be checked for their quality and size/space requirements.
- *Metadata:* Metadata for the different datasets is collected in a standard format to facilitate query and consistent presentation across multiple participating sites.
- *Search engine:* Clearinghouse search engine technology generally includes spatial query, text query and search of Metadata, field level search based on topics, geography, time and other important fields.
- *Collective metadata:* The definition of a dataset generally corresponds to individual identifiable data products (e.g. file, layer, service) for which metadata are customarily collected. Collections of datasets may also have generalized metadata that could be inherited by individual datasets.
- *Advanced query functions:* To provide search interoperability among different servers of geospatial metadata, a standard search and retrieve protocol specification has to be selected.
- *Data architecture:* This component is not a quantifiable component of the clearinghouse but provides the framework for the manner in which datasets and other components of data clearinghouses should be organized and integrated for freight transportation-related applications or business processes.

Two important components, the inventoried datasets and data architecture are considered for further study in the sections below.

4.1.2: Inventoried Datasets

Section 1 and 2 of this report included a comprehensive assessment of different public and private datasets. This assessment aimed to assist local, regional, and state freight coordinators in freight transportation-related applications from a holistic perspective which included multiple modes, different levels of spatial and temporal details, all commodities and supply chain nature of freight data. The Data Inventory Matrix provided as an attachment in Section 1 illustrated the complete list of the inventoried datasets.

Heterogeneous Nature of Datasets

One of the challenges encountered while adding different datasets to a data clearinghouse is the heterogeneous nature of datasets. These heterogeneities can be classified in three forms - structure, syntax, and semantics.

- *Structural heterogeneity* indicates differences in how the data is stored in the various databases.
- Syntactic *heterogeneity* indicates differences in the representation of the data; in other words, data types and formats.
- *Semantic heterogeneity* indicates differences in interpretation of the meaning of the data.

Resolving freight data heterogeneity across multiple data sources is necessary to facilitate the integration of data elements, enable interoperability between multiple systems, and smooth the exchange of data and information between clearinghouses. Many methodologies have been explored providing robust classification schemes for dataset. These classification schemes and their responsible developer are illustrated in Table 4.1 below.

Table 4.1: Different Types of Freight Data Classification Schemes

Title	Description	Developer
CODMRT	» **C**ommodity, which describes the type of freight being moved and contains information such as value, weight, and handling characteristics. » **O**rigin, which describes the geographic starting point of a freight trip. » **D**estination, which describes the geographic ending point of a freight trip. » **M**ode, which describes the vehicles and infrastructure used to transport goods. » **R**oute, which describes the sequence of specific individual facilities (e.g., sections of roads, railroad tracks, etc.) that are used to transport freight between the origin and destination on a specific mode. » **T**ime, which is defined as the time period for which the freight data was collected (i.e., the freight forecast time period).	TRB Committee on Freight Transportation Data (2003)

Title	Description	Developer
FFFN	» **F**reight node data, which represents consolidated or individual endpoints that generate or receive freight flows and are the key points of production, consumption, or intermediate handling for goods. » **F**reight network data, which defines major route patterns and critical infrastructure being used to convey freight shipments through the various modal systems. » **F**reight flow data, which provides information on commodity flows and provides insight on the economic and trade environment of regions. Typical commodity flow records will contain information on the O-D of shipments, type of commodity, weight, and/or value of the commodity shipment, and mode of shipment. » **N**eighborhood freight data, which provides information on safety, congestion, land use, and emissions.	NCFRP Report 14 [1]
Role-Based Classification Schema (RBCS)	» Time » Place, which can be an identifier or a feature » Link » Mode: » Commodity: » Industry: » Event: » Human: » Unclassified	NCFRP Report 35 [2]
Data Quality Assessment Framework (DQAF)	» Prerequisites of quality » Integrity » Methodological soundness » Accuracy and reliability » Serviceability » Accessibility	INTERNATIONAL MONETARY FUND
Complexity Measure	» Spatial Coverage » Commodity Coverage » Industry Coverage » Modal Coverage » Temporal Coverage » Collection Frequency/Update » Accuracy » Access » Usability	NCFRP 12 and RS&H

Utilization of these classification methods will provide the following benefits:

- Designing a structured nature of data clearinghouse.
- Refining components of data architecture (Refer Section 4.1.3 for more details).
- Determining appropriate data fusion methods which are necessary to merge two or more datasets together.

RBCS is expected to be the one of the most user friendly classification scheme for building a robust data clearinghouse. The classification schemes can also be used for developing textual search and for building bridges among different datasets.

4.1.3: Data Architecture

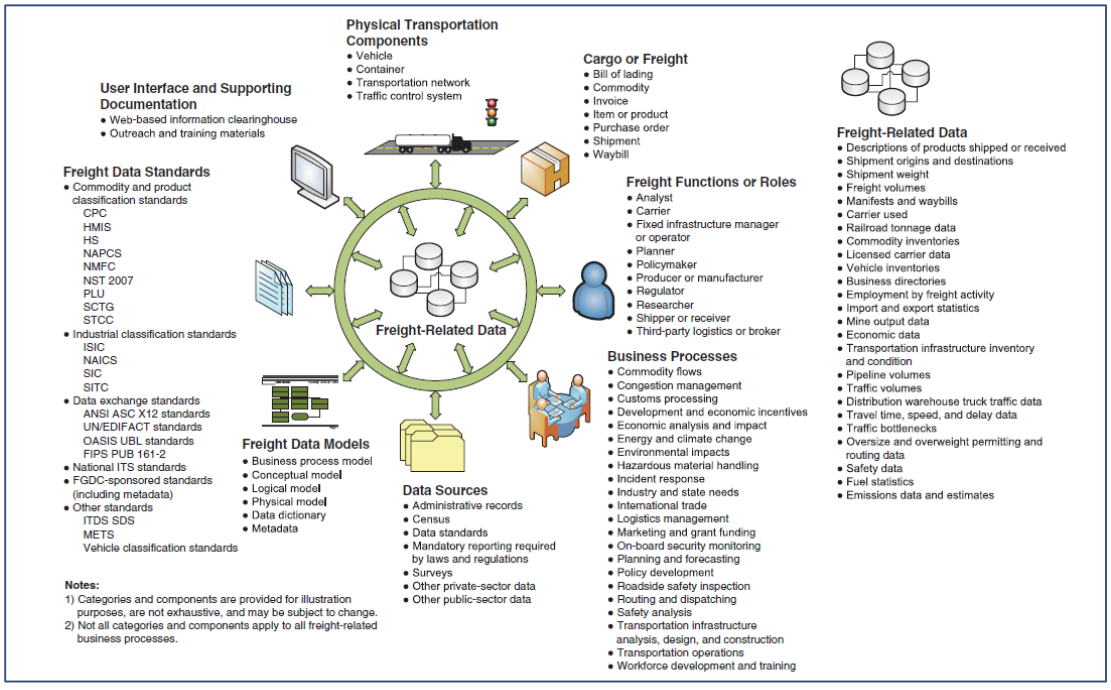
According to NCFRP 35 [2] data architecture is defined as:

> "*The national freight data architecture is the manner in which data elements are organized and integrated for freight transportation-related applications or business processes. The data architecture includes the necessary set of tools that describe related functions or roles, components where those roles reside or apply, and data flows that connect roles and components at different domain and aggregation levels.*"

Figure 4.1 illustrates the different components of the proposed National Freight Data Architecture Framework. As per NCFRP 9 [3], data architecture includes following components:

- Physical transportation components
- Cargo or freight
- Freight functions or roles
- Business processes
- Data sources
- Freight-related data
- Freight data models
- Freight data standards
- User interface and supporting documentation

Figure 4.1: National Freight Data Architecture Framework and Components



Standard frameworks of data architecture can be classified into centralized, distributed, decentralized and hierarchical. Data architecture for a state based data clearinghouse should include a holistic, all-encompassing approach with all data elements organized and integrated for multiple freight transportation-related applications or business processes. The FDOT handles a lot of public records requests and public information requests as part of day-to-day activities, a freight data clearinghouse can facilitate and enhance the coordination between FDOT, MPOs/TPOs, local governments, and freight industry. It can elevate how the FDOT responds to the freight industry's call for support. NCFRP 35 [2] identified freight data functions and applications. An overview of these freight related functions and applications is provided in Table 4.2. The freight functions listed in Table 4.2 were used to characterize current and potential applications of selected profile data sheets provided in Section 2 of this report.

Table 4.2: Freight Planning and Decision-Making Public Sector Functions

Function	Description
Congestion Management	Identify and monitor recurring and non-recurring congestion along road corridors and evaluate and recommend mitigation strategies
Operations/Services	Develop, operate, and maintain transportation modes; improve the movement of goods and people and increase the safety and efficiency of the transportation system through enhanced management and operations coordination
Safety Planning and Analysis	Implement and maintain integrated multimodal safety and transportation planning; the ultimate goal is to reduce crashes, injuries, and fatalities
Freight Mobility Planning	Incorporate goods movement into the regional transportation planning process
Emergency Preparedness and Security Planning	Increase the safety and security of the transportation system through enhanced coordination and communications among emergency responders
Economic Development Planning	Estimate the impacts of transportation planning on local population and employment
Freight Transportation and Land Use Planning	Coordinate regional freight transportation planning and land use development
Environmental Planning	Investigate activities involving mobile emissions planning, environmental protection, land use management, and air quality efforts
Regulation and Enforcement	Conduct activities such as licensing, inspection, size and load specifications, work hours regulation, and taxes/fares
Intermodal Trade Corridor Planning	Develop intermodal corridors to ensure efficient freight movement and reduce congestion
Terminal and Border Access Planning	Manage terminals and borders to ensure efficient movement of people and goods across modes

Function	Description
Hazardous Materials Planning	Improve safe movement and monitoring of hazardous materials transported using the freight system
Roadway Pavement and Bridge Maintenance Planning	Study the effects of fleet use on infrastructure, such as expected pavement deterioration
Modal Shift Analysis	Investigate policies and incentives that foster modal shift changes, including measuring the impact of shifting from one mode to another
Freight Performance Measurements	Develop measures to monitor the performance of the freight transportation system, including its subsystems and components
Sustainable Transportation Investment	Investigate ways to fund the existing transportation system and future projects
Financial Planning	Investigate grants, loans, and subsidies to support the transportation system; also involves tax policy, road user fee assessment, and other activities such as public-private partnerships
Interregional Connectivity	Develop intermodal corridors to ensure efficient freight movement and reduce congestion
Security Planning	Integrate emergency response and other calculations into transportation planning
Transportation Equity Planning	Incorporate transit equity principles and legislation such as SAFETEA-LU into regional transportation planning

4.2 GUIDELINES FOR EVALUATION OF DATA CLEARINGHOUSES

This subsection includes a discussion of the standard guidelines used for the evaluation of data clearinghouses, as well as, the methodology for the intuitive criteria developed specifically for the data clearinghouses analysis. The subsection also includes a discussion for literature review of other guidelines used and further detailed within Appendix E.

4.2.1: Standard Guidelines for Evaluation of Data Clearinghouses

Fundamental to the guidelines are six principles that together determine whether or not the data may be considered as sustainably archived:

- The data is available on the Internet;
- The data is accessible, while taking into account relevant legislation with regard to personal information and intellectual property of the data;
- The data is available in a usable format;
- The data is reliable;
- The data can be referred to a published document; and,
- The data is updated at a regular frequency.

These six principles are integral to the guidelines, which focus on three stakeholders: the data producer, the data clearinghouse and the data consumer.

- The data producer is responsible for the quality of the data.
  - Example: Florida Department of Transportation is a data producer for Roadway Characteristics Inventory (RCI) dataset.
- The data repository is responsible for the quality of storage and availability of the data.
  - Example: Transportation Statistics Office website is the data repository for road GIS data.
- The data consumer is responsible for the quality of use of the digital data.
  - Example: Florida Department of Transportation is a data consumer for Transearch dataset.

Tables 4.3 and Table 4.4 provide a list of guidelines and a list of rating metrics, respectively, which have been developed for self-assessment of data repositories to get a “Data Seal of Approval” [4]. A comprehensive review and analysis of these guidelines indicated that they are robust and attempt to eliminate any limitations within a data repository or a data clearinghouse.

A data repository is designated as Trusted Digital Repository (TDR) according to the requirements of the “Data Seal of Approval” if it meets guidelines 1-16. The guidelines can be categorized as follows:

- Guidelines 1-3 are related to data producers;
- Guidelines 4-13 are related to repositories; and,
- Guidelines 14-16 are related to data consumers.

These guidelines are listed in Table 4.3 and a detail description of the guidelines 1-16 is provided within Appendix E.

Table 4.3: List of Guidelines for Evaluation of Data Clearinghouse

Category	Guideline Number	Guideline Details
	0	Clearinghouse context
Related to Data Producers	1	The clearinghouse has an explicit mission to provide access to and preserve data in its domain.
	2	The clearinghouse maintains all applicable licenses covering data access and use and monitors compliance.
	3	The clearinghouse has a continuity plan to ensure ongoing access to and preservation of its holdings.
Related to Repositories	4	The clearinghouse ensures, to the extent possible, that data are created, curated, accessed, and used in compliance with disciplinary and ethical norms.
	5	The clearinghouse has adequate funding and sufficient numbers of qualified staff managed through a clear system of governance to effectively carry out the mission.
	6	The clearinghouse adopts mechanism(s) to secure ongoing expert guidance and feedback (either in-house, or external, including scientific guidance, if relevant).
	7	The clearinghouse guarantees the integrity and authenticity of the data
	8	The clearinghouse accepts data and metadata based on defined criteria to ensure relevance and understandability for data users.
	9	The clearinghouse applies documented processes and procedures in managing archival storage of the data.
	10	The clearinghouse assumes responsibility for long-term preservation and manages this function in a planned and documented way.
	11	The clearinghouse has appropriate expertise to address technical data and metadata quality and ensures that sufficient information is available for end users to make quality-related evaluations.
	12	Archiving takes place according to defined workflows from ingest to dissemination
	13	The clearinghouse enables users to discover the data and refer to them in a persistent way through proper citation.
Related to Data Consumers	14	The clearinghouse enables reuse of the data over time, ensuring that appropriate metadata are available to support the understanding and use of the data.
	15	The clearinghouse functions on well-supported operating systems and other core infrastructural software and is using hardware and software technologies appropriate to the services it provides to its Designated Community.
	16	The technical infrastructure of the repository provides for protection of the facility and its data, products, services, and users.

The clearinghouse should analyze potential threats, assess risks, and create a consistent security system. It should describe damage scenarios based on malicious actions, human error, or technical failure that pose a threat to the repository and its data, products, services, and users. It should measure the likelihood and impact of such scenarios, decide which risk levels are acceptable, and determine which measures should be taken to counter the threats to the clearinghouse and it's Designated Community. This should be an ongoing process.

The guidelines 3, 9 and 16 were not considered for the evaluation of example data clearinghouses within this report. These guidelines would need additional information requiring additional resources and time to conduct interviews or survey of the developers/owning agencies. But, all guidelines are recommended for a clearinghouse which would be developed in future through the Freight and Modal Program. Guideline 0 is a reference guideline and is not used in the evaluation.

Table 4.4 illustrates the rating system utilized, which is broken into four categories to develop a Strengths, Weaknesses, Opportunities, and Threats (SWOT) metrics as follows:

- 4-5: **S**trengths
- 2-3: **O**pportunities
- 0-1: **W**eaknesses
- -1: **T**hreats

Table 4.4: Ratings and SWOT Metrics

SWOT Metric	Rating	Definition
Threats	-1	Cannot be implemented
Weaknesses	0	N/A: Not Applicable
Weaknesses	1	No: Not considered yet
Opportunities	2	Theoretical: Have a theoretical concept – URL needed for the initiation document
Opportunities	3	In progress: In the implementation phase provide a URL for the supporting document
Strengths	4	Implemented: This guideline has been fully implemented for the needs of the repository
Strengths	5	A URL for the supporting document is provided

4.2.2: User Developed Guidelines

Specific guidelines were developed to evaluate various clearinghouses and are classified in the following categories:

- Data Coverage
- Technical Support
- User Interface
- Standardization
- Adaptability

Table 4.5 details the use of these guidelines and provides rating factors for analytical review of each clearinghouse. The rating system varies depending on the individual guideline but the system is consistent and homogenous throughout the list. Two overarching rating systems were employed:

- Simple Binary Code System 0-1, where "0" indicates No and "1" indicates Yes; and,
- Simple Likert Scale 0-4, where "0" is least/worst and "4" is most/best.

The guidelines which are marked in bold were not considered for the evaluation of example data clearinghouses in the analysis summarized in Section 4.3. Additional information is required which can only be gathered through interviews or survey of the developers/owning agencies.

Table 4.5: List of Guidelines for Evaluation of Data Clearinghouse based on User Support

Category	Guidelines	Ratings
Data Coverage	Multiple years of data	0 – No, 1 – 2-3 years, 2 – more than 3 years
	Roadway Data available (AADT and # of lanes minimum)	0-No, 1-Yes
	Base map of all public roads	0-No, 1-Yes
	Tonnage, Value, TEU Data Available	0-No, 1-Yes
	Relevance of Content	Scale 0-4 0 is worst and 4-best
Technical Support	Email Support / Telephone Support	0-No, 1-Yes
	Text Instructions	0-No, 1-Yes
	Online help	0-No, 1-Yes
	Frequently Asked Questions	0-No, 1-Yes
	Workshops	0-No, 1-Yes
	System improvements after feedback	0-No, 1-Yes
User Interface	Intuitive Interface	0-No, 1-Yes
	Interactive	Scale 0-4 0 is worst and 4-best
	Navigation	Scale 0-4 0 is worst and 4-best
	Visualization	Scale 0-4 0 is worst and 4-best
	Semantic search function	Scale 0-4 0 is worst and 4-best
	Query options	0-No, 1-Yes
Standardization	Data Analysis	0-No, 1-Yes
	Georeferenced information	0-No, 1-Yes
	Usage of Standards	0-No, 1-Yes
	Standards for data transfer	Scale 0-4 0 is worst and 4-best
	Literature database	0-No, 1-Yes

Category	Guidelines	Ratings
Adaptability	**Use of computing resources**	0-No, 1-Yes
	Flexible plug-ins	0-No, 1-Yes
	Upload routines	Scale 0-4, 0 – rare and 4-frequent
	Scalability of the system	Scale 0-4 0 is worst and 4-best
	Exportable files/formats	0-No, 1-Yes
	Mobile Device Functionality	0-No, 1-Yes

4.2.3: Other Evaluation Criteria

There are several other methodologies developed by Open Archival Information Systems (OAIS) organizations to evaluate the repositories and clearinghouses, including:

- NESTOR Seal for Trustworthy Digital Archives [5]
- ISO16363: Audit and Certification of Trustworthy Digital Repositories [6]
- Digital Repository Audit Method Based on Risk Assessment (DRAMBORA) [7]
- PREPARDE, Guidelines on Recommending Data Repositories as Partners in Publishing Research Data
- Interdisciplinary Body of the International Council for Science (ICSU) World Data System [8]
- DIN 31644: Criteria for trustworthy digital archives [9]

Appendix E provides list of guidelines for the first five methodologies above. The guidelines for the final methodology are not available publicly. For more details, cited references should be explored. It is important to note that this list is not an exhaustive list but it provides a starting point for reviewing data clearinghouses.

4.3 DATA CLEARINGHOUSE EVALUATION

A comprehensive search was completed to compile a list of 84 data clearinghouses with a freight component or freight related data. All these clearinghouses are owned by a state or federal agency. After a detailed review of these clearinghouses, it was determined that 52 have satisfactory amount of datasets and user friendliness to further analyze. These data clearinghouses were evaluated using the guidelines outlined in Tables 4.3, 4.4, and 4.5. The rating matrices are included in Appendix F.

Important highlights of the rating matrices are as follows:

- All data clearinghouses are evaluated for Guidelines based on user support. Some data clearinghouses were not evaluated for Standard Guidelines based on Data Seal of Approval as there was not enough information available.
- The evaluation of example data suggests that none of the clearinghouses have ideal ratings for every guideline.
- The following clearinghouses provided the highest overall evaluation ratings and are recommended for adoption/further exploration:
  - Massachusetts Department of Transportation GIS
  - Minnesota Geospatial Information
  - Washington Department of Transportation GeoPortal
  - Virginia Economic Development Partnership GIS
  - North Central Texas Council of Governments GIS
  - Rhode Island GIS Data Catalog
  - USDOT Bureau of Transportation Statistics TranStats
  - Freight Analysis Framework Version 4 (FAF4)
- It is recommended that these data clearinghouses are further evaluated with interviews and surveys of the developers and owning agencies.

4.4 DATA FUSION ANALYSIS

Data fusion combines multiple sources to obtain improved information, for example data that is cheaper, higher quality, and/or more relevant. As per the Joint Directors of Laboratories (JDL) workshop [10] data fusion can be defined:

> *"A multi-level process dealing with the association, correlation, combination of data and information from single and multiple sources to achieve refined position, identify estimates and complete and timely assessments of situations, threats and their significance".*

It is noteworthy to introduce three different terms:

- "Data Fusion"
- "Information Fusion"
- "Data Integration"

"Data Fusion" and "Information Fusion" are generally used as synonyms but the major difference between the two is that "Data Fusion" is synthesis of raw data and "Information Fusion" is combination of already processed data. For this analysis, both are considered synonyms as majority of the datasets considered in this report are processed data. The Data Integration primer [11] defines data integration as:

> *"The method by which multiple datasets from a variety of sources can be combined or linked to provide a more unified picture of what the data mean and how they can be applied to solve problems and make informed decisions that relate to the stewardship of transportation infrastructure assets."*

Data fusion and data integration are synonymous concepts. Some of the benefits of data integration are identified below:

- Integrated decision-making
- Safety analysis
- Consistency
- Clarity
- Completeness
- Lower data acquisition and storage costs
- Informed and defensible decisions
- Enhanced program development
- Greater accountability

4.4.1: Data Fusion Classification Schemas

Data fusion can be classified into five different schemas [12]. Figure 4.2 illustrates this framework. A detail description of each schema can be found in the Appendix E.

4.4.2: Factors Impacting Data Fusion Methodology

The selection of technique varies and depends on following factors:

- Cost which involves cost of data, data storage, data fusion software, operations and maintenance costs
- Time
- Accuracy of data fusion needed
- Characteristics of data like level of detail, spatial coverage, temporal coverage, etc.

- Availability of resources like skills, software, data platforms, etc.
- Purpose of data fusion
- Standards of data
- Limitations in access and usage of data

Figure 4.3 provides an outline of the key activities in the data integration process, along with the factors affecting each activity of the process. Analyzing requirements is the first step in the process. Once the requirements are known, a data and process flow modeling can be developed leading to the definition, evaluation and selection of alternatives. After an alternative is selected, the database design and specification can be pursued. Finally, the development, testing and implementation of the chosen database integration strategy can be implemented.

Figure 4.2: Data Fusion Classification Schemas

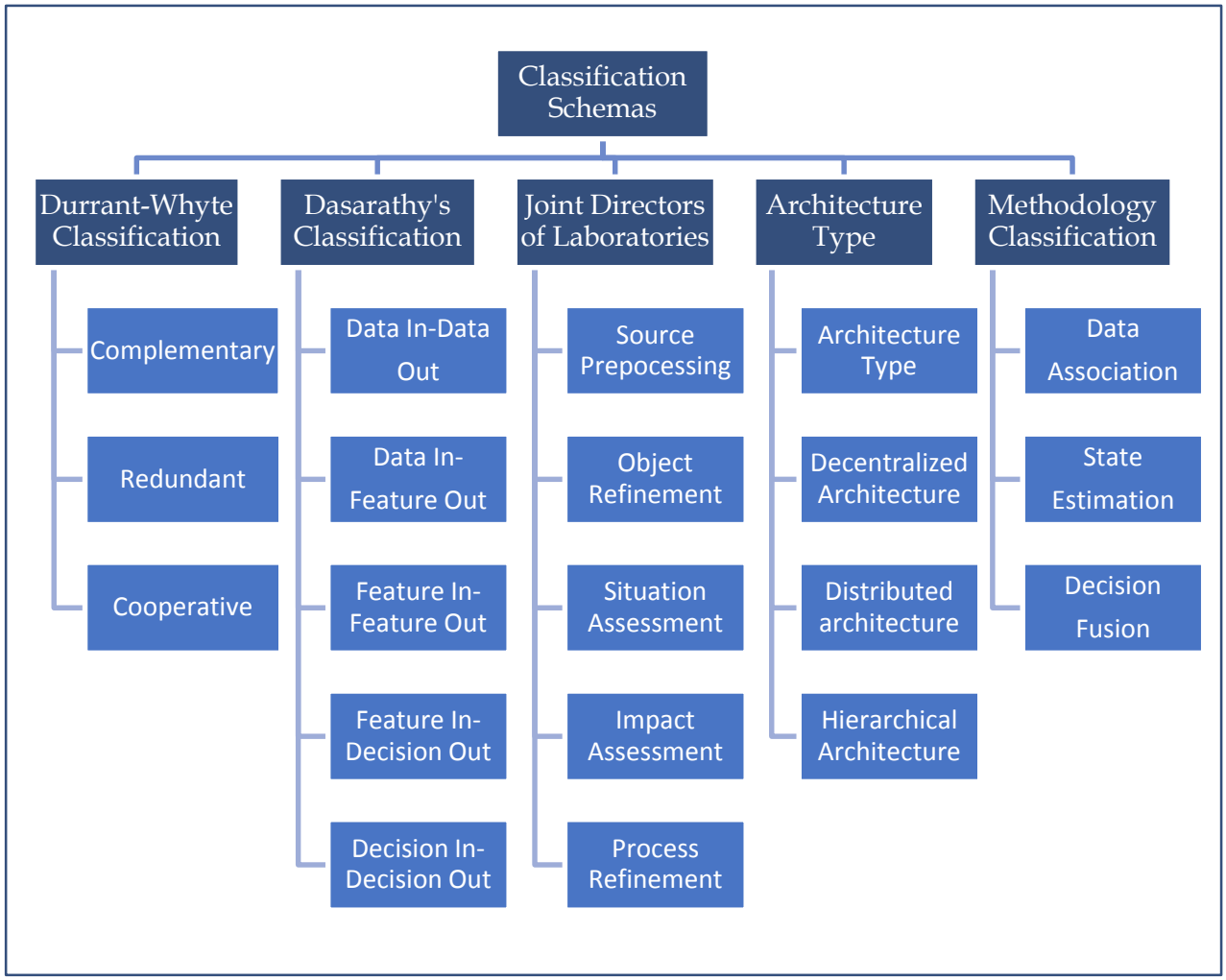


Figure 4.3: Activities of Data Integration/Fusion Process and Factors Impacting the Activities