# Implementing the U.S. Department of Transportation's Freight Data Dictionary

U.S. Department of Transportation
National Transportation Library

Research Data Alliance
11th Plenary
Berlin, Germany,
March 21-23, 2018

National Cooperative Freight Research Program (NCFRP) report 35, Implementing the Freight Transportation Data Architecture: Data Element Dictionary https://trid.trb.org/View/1367451, identified freight data sources, compiled and classified data elements, identified differences in data element definitions, and produced software for a searchable inventory of freight data dictionaries. The report indicates that the U.S. Department of Transportation's (USDOT) Bureau of Transportation Statistics (BTS) will permanently host the web-based interface, with staff from the National Transportation Library (NTL) providing updates and maintenance.

NTL is currently installing and testing the freight data dictionary (FDD) code, preparing for public launch in early 2018.

While technical launch of the FDD is an important step, in order to make the FDD a robust, sustainable and useful tool, the freight data community will need to be engaged in its maintenance. This includes active updating of data sources, dictionary elements, and software, as well as identifying new data sources from which the FDD draws.

This poster will give a progress report on current BTS activities to launch the web-based FDD. The poster will also explore future areas for freight community involvement in the long-term maintenance of the FDD.

## Project Background

National Cooperative Freight Research Program (NCFRP) studies on the necessity and development of a freight data dictionary (FDD) began in 2011.

A national Freight Data Dictionary is proposed that offers a centralized, controlled, authoritative vocabulary capable of supporting:

- Enhanced data inputs
- Improved accuracy, efficiency, and flexibility in freight data interchange
- Freight data analysis and interoperability across the transportation sector
- Improved analysis and decision making at all levels of government


For early research, see:

NCFRP Report 9: *Guidance for Developing a Freight Transportation Data Architecture* http://www.trb.org/Publications/Blurbs/164644.aspx &

NCFRP Report 25: *Freight Data Sharing Guidebook* http://www.trb.org/Publications/Blurbs/169010.aspx

## Metadata

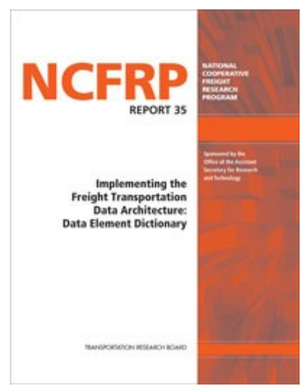
Structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage an information resource.

- Improves content discovery
- Classifies and categorizes
- Analyze term frequency
- Link multiple key words to a preferred term
- Provide multilingual support


## NCFRP Report 35

***Implementing the Freight Transportation Data Architecture: Data Element Dictionary***

http://www.trb.org/Main/Blurbs/173083.aspx




- Identifies "readily available" data sources associated with freight
- Provides examples of freight data uses and applications
- Presents an inventory of data elements and glossary terms found in the selected sources into a uniform typology
- Identifies differences in data element definitions
- Provides metadata tools and resources to guide data users on the appropriate steps and procedures for combining data from multiple freight data sources

Result: a searchable and sustainable web-based application containing the study findings, an inventory of freight data dictionaries, and a discussion feature to be used by practitioners to exchange ideas and information.

NCHRP Report 35 calls for FDD to be hosted by U.S. DOT's Bureau of Transportation Statistics (BTS). BTS's National Transportation Library (NTL) staff are tasked with building the FDD, given their expertise with controlled vocabularies.

**The Freight Data Dictionary will function as a dictionary, not as a controlled vocabulary, nor as a repository. The FDD will record how different freight entities define similar terms differently, or, record how similar concepts may be signified with different terms. This should allow users to understand how terms are used in their specific contexts.**

**The FDD will NOT attempt to enforce preferred definitions.**

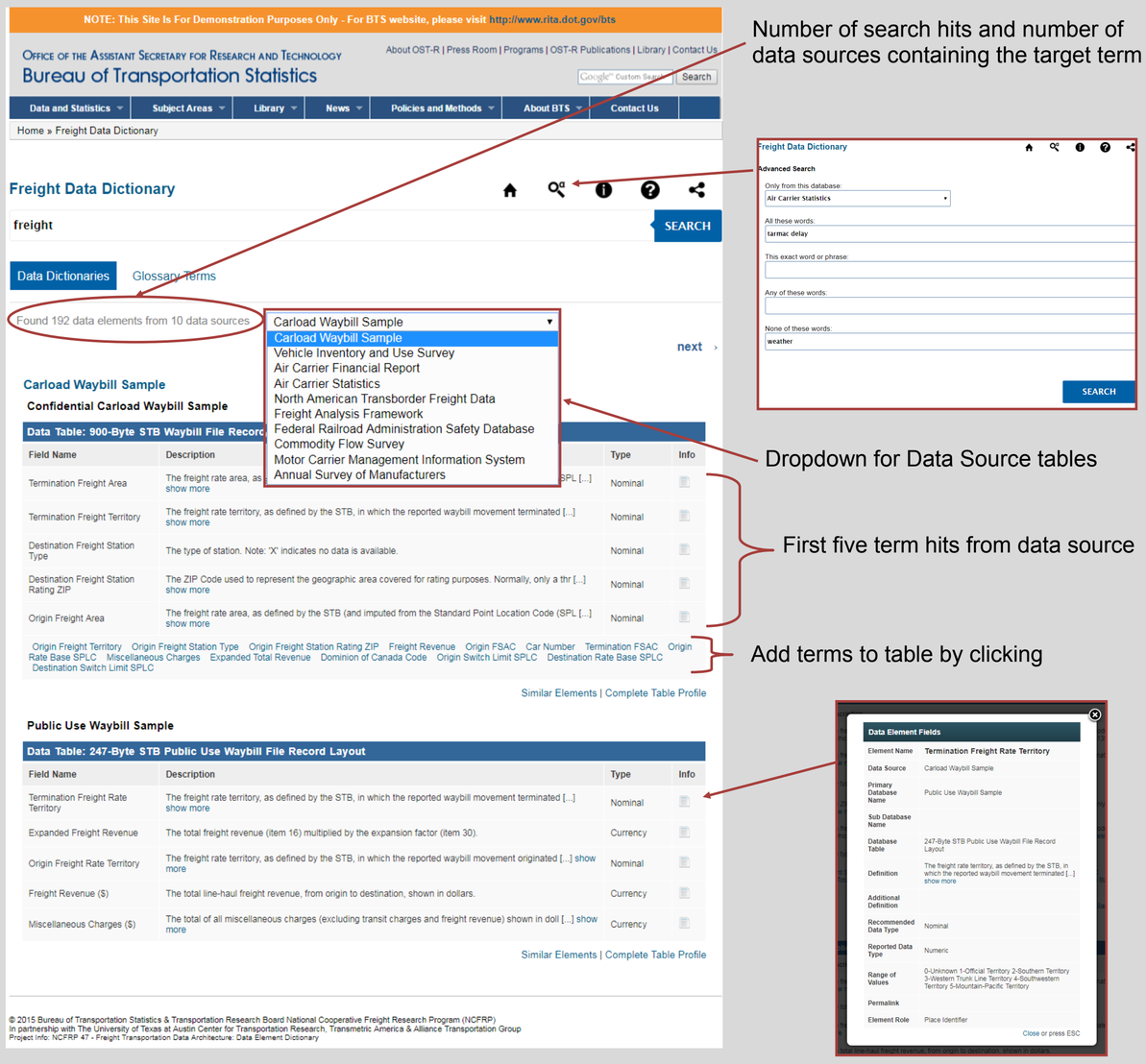
## Data Sources

The current FDD system is populated with freight data dictionary terms from 28 entities.

1. Air Carrier Statistics
2. Air Carrier Financial Reports
3. Annual Survey of Manufacturers
4. Border Crossing/Entry
5. CTA Intermodal Terminals Database
6. Carload Waybill Sample
7. Commodity Flow Survey
8. County Business Patterns
9. Fatality Analysis Reporting System
10. Federal Railroad Administration Safety Database
11. Foreign Trade
12. Freight Analysis Framework
13. Highway Performance Monitoring System
14. IHS Transearch
15. Motor Carrier Management Information System
16. Motor Carrier Safety Measurement System
17. National Agricultural Statistics Service
18. National Ballast Information Clearinghouse Database
19. National Corridors Analysis and Speed Tool Database
20. North American Transborder Freight Database
21. Pipeline and Hazardous Material Safety Administration
22. Service Annual Survey
23. Survey of Business Owners
24. Topologically Integrated Geographic Encoding and Referencing
25. U.S. Waterway Data
26. Vehicle Inventory and Use Survey
27. Vehicle Travel Information System
28. Woods and Poole Economics, Inc.

***While this list is a good start, there are likely dozens more data sources which should be included in the U.S. DOT FDD. If you have ideas, please email the authors.***

## Functions & Features




Number of search hits and number of data sources containing the target term




Dropdown for Data Source tables

First five term hits from data source

Add terms to table by clicking




## Community Engagement

The greatest challenge of the Freight Data Dictionary is *not* technical. Once the FDD is launched, the greatest demand will be engaging the freight data community in enhancing, maintaining, and improving the dictionary. To help build user interest, the authors gave presentations on the FDD at the Transportation Research Board 2018 Annual Meeting, including a lively conversation with the Freight Data Users' Forum. https://tinyurl.com/trb-fdd

**Key future tasks for the freight data community:**

- Verify data sources and fill in missing definitions through outreach to freight data source entities
- FDD user testing, beginning in Spring 2018
- Build FDD Governance Body
- Expand data sources through outreach
  - Expand FDD to include international freight data
- Set FDD versioning policy
- Recruitment to freight data users community and FDD community

**FDD Improvement Ideas from Freight Data Users' Forum:**

- Create Data Source catalog within FDD so that users can link directly to freight data sources
- Improve advanced search
- Build API for FDD harvesting
- Review metadata elements to include units of measure where appropriate
- Build function for data element (definition) comparison from different data sources
- Build a governance body
- Review NISO Standards on controlled vocabularies

***Join the Freight Data Dictionary Community.***

***Leave your card or email leighton.christiansen@dot.gov***

## Technology

- Initial Development: Java, JavaScript, jQuery, Ajax, PHP and Oracle Database
- Converted Oracle database to Microsoft SQL Server
- Created development, testing, staging, and production environments in Microsoft Azure Cloud
- Basic system security designed to meet DOT requirements
- Code conversion and debugging in Azure development environment is underway
- Azure indexing will be added which will decrease the load on database, decrease search time and increase the system efficiency
- Web API will be created for integration to other applications

## FDD Implementation Timeline

- Setup the Freight Data Dictionary in DOT network (accomplished)
- Convert Unix based Oracle to Windows based Oracle (accomplished)
- Convert Oracle to Microsoft SQL Server in Azure Cloud (accomplished)
- Implement Azure Index to the application (tested)
- Recode the application in PaaS in Azure (in progress)
- System testing (on Schedule – April 2018)
- Develop Web based Public Access API (on schedule – May 2018)
- Move to staging and Production environment (on schedule – May 2018

## Authors

Leighton L Christiansen* http://orcid.org/0000-0002-0543-4268,Data Curator

Vinod Koduri, Senior Web Engineer

Mary Moulton https://orcid.org/0000-0002-1791-068X, Digital Librarian

Pavan Tuniki, FDD Programmer

Xin Wang, Systems Librarian

Ted Westervelt, NTL Acting Director

* = *Corresponding author. Please email at* leighton.christiansen@dot.gov

**Recommended Citation:** Christiansen, Leighton L http://orcid.org/0000-0002-0543-4268; Vinod Koduri; Mary Moulton https://orcid.org/0000-0002-1791-068X; Pavan Tuniki; Xin Wang; and, Ted Westervelt. 2018. "Implementing the U.S. DOT Freight Data Dictionary." Research Data Alliance 11th Plenary. Berlin, Germany.

**Acknowledgements:** The authors thanks the National Transportation Library staff for their input, support, and teamwork.