# Freight Data Dictionary

## Linking freight data sources across transportation modes, subjects, and geography

Mary Moulton
Leighton Christiansen
Xin Wang

The National Transportation Library
Bureau of Transportation Statistics
US Department of Transportation

FCSM Metatada Workshop
September 14, 2018


PORT OF NEW ORLEANS
M S C

# Overview

- About the National Transportation Library
- Project background
- Functions and features
- Information architecture
- Implementation
- Future development

# About NTL

Bureau of Transportation Statistics, Office of Information and Library Sciences (OILS)

Established in 1998, we provide to:

- Digital collections
- Data services
- Reference and research services
- Networking

We are an **open access** digital repository.
All items are in the **public domain** and available for reuse **without restriction**.

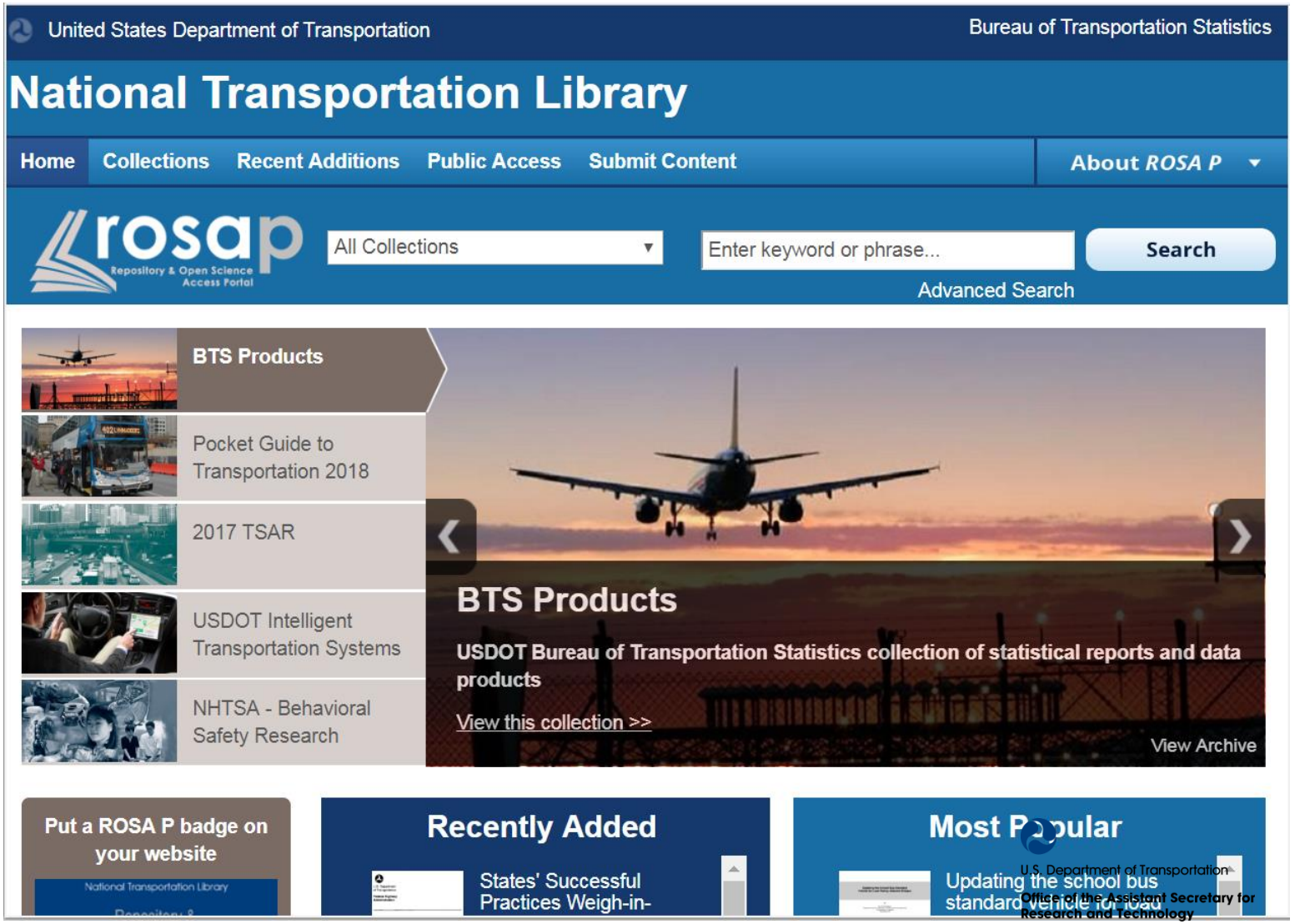
# NTL Mandates

- **Transportation Equity Act for the 21st Century (1998)**
  - "establish and maintain a National Transportation Library, which shall contain a collection of statistical and other information needed for transportation decision making at the Federal, State, and local levels."

- **MAP-21 (2012)**
  - Acquire, preserve and manage transportation information and information products and services for use by DOT, other Federal agencies, and the public
  - Central repository for DOT research results and technical publications
  - Central clearinghouse for transportation data and information of the Federal Government
  - Coordinate among and cooperate with multiple external parties to develop a "comprehensive transportation information and knowledge network"

- **White House Office of Science and Technology Policy memo (2013)** requiring all Executive Departments and Agencies spending more than $100 million/year on R&D to ensure public access to peer-reviewed publications and digital datasets arising from federally-funded scientific research

# Repository and Open Science Access Portal (ROSA P)



# Project Background

# Project Background

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

# National Cooperative Freight Research Program (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.

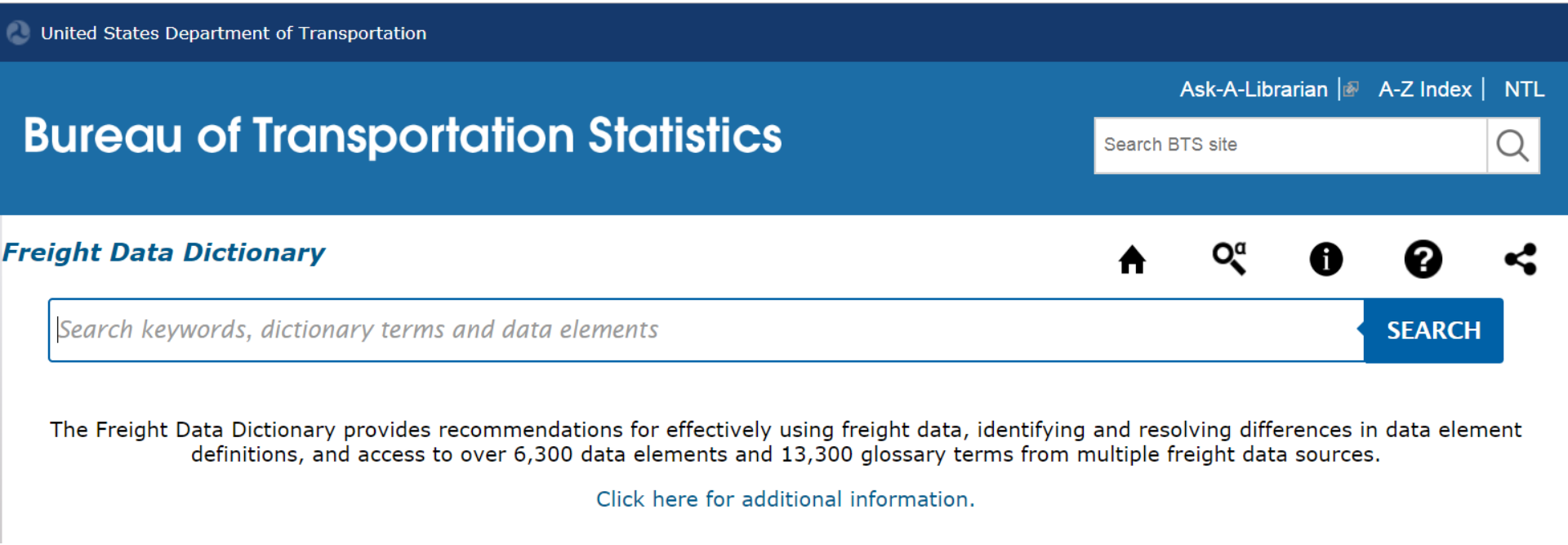
# Why is BTS Interested in the FDD?

- BTS is a freight and transportation statistics aggregator and publisher.
  - Several BTS products are represented in the FDD.
- BTS identified as logical host for FDD.
- FDD can provide a model architecture and platform for BTS metadata harmonization efforts.
  - In 2016 BTS launched a Data Management and Data Curation project.
- FDD provides potential model for other transportation modal data dictionaries.

# Functions & Features

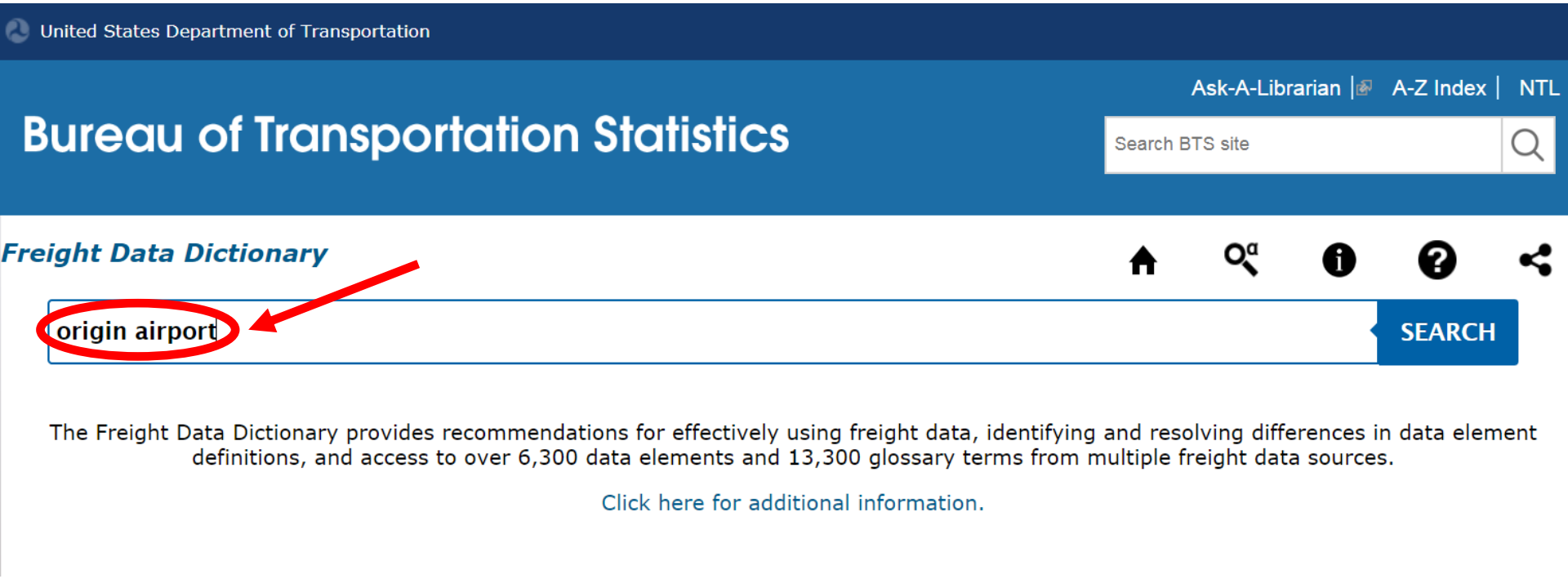
# FDD Home Page

https://fdd.bts.gov/freight-data-dictionary/



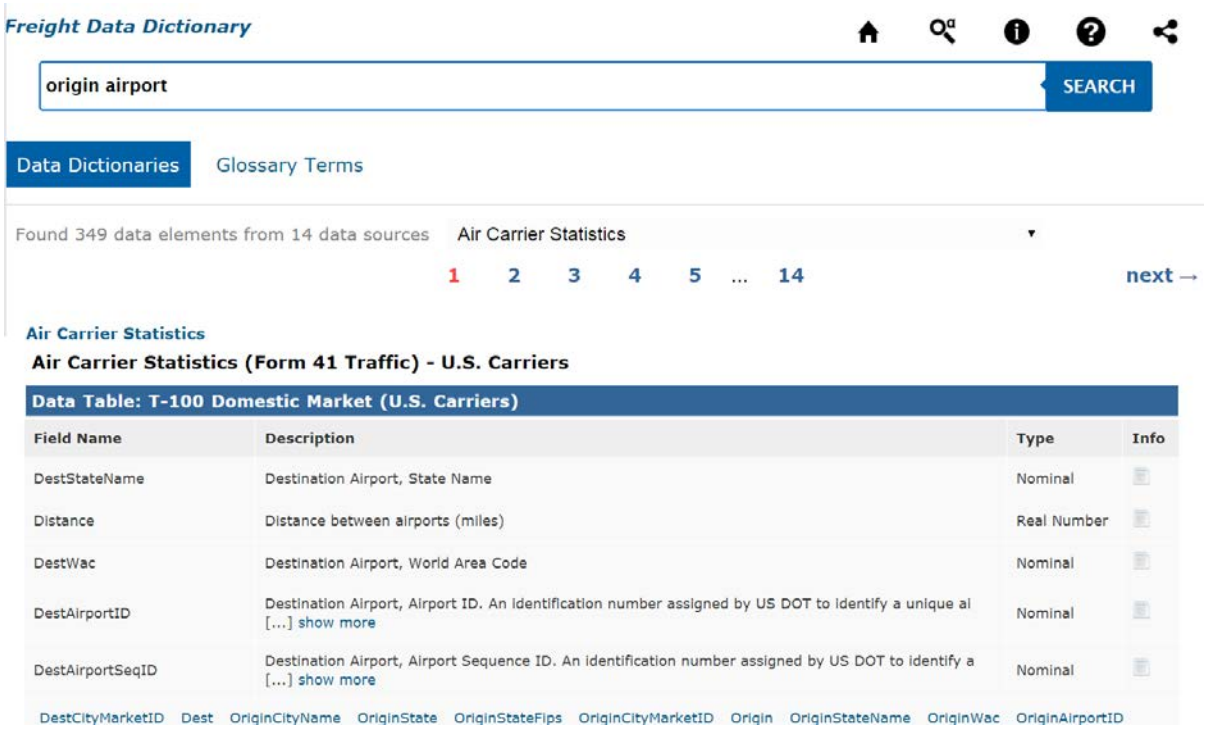
# FDD Simple Search:

## *Origin Airport*



# FDD Simple Search:

## *Origin Airport*

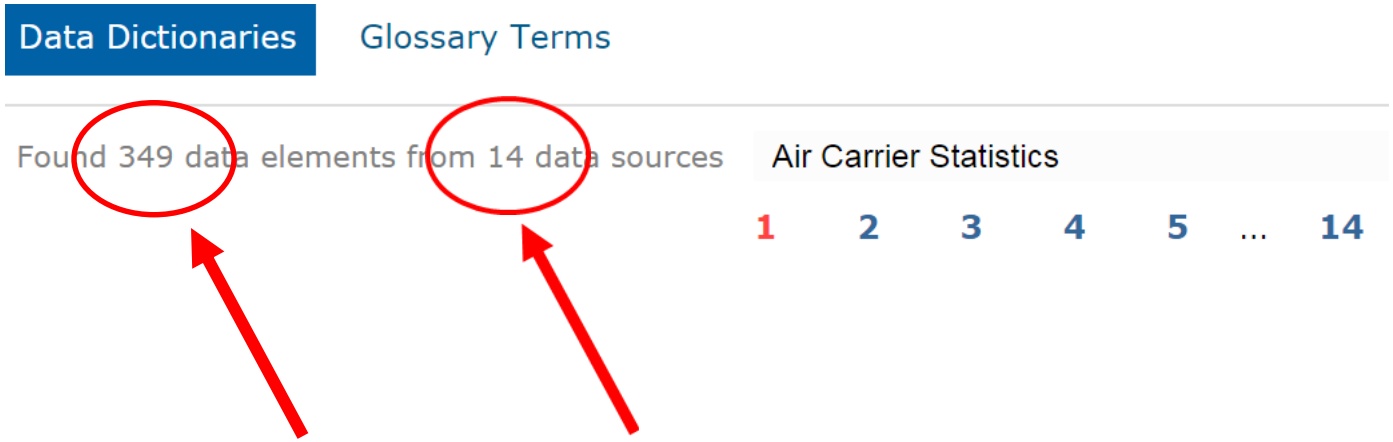



There are two distinct main tables in the system:
***Data Dictionaries*** and ***Glossary Terms***

(Don’t squint: we will zoom in on each of these.)

# FDD Simple Search:

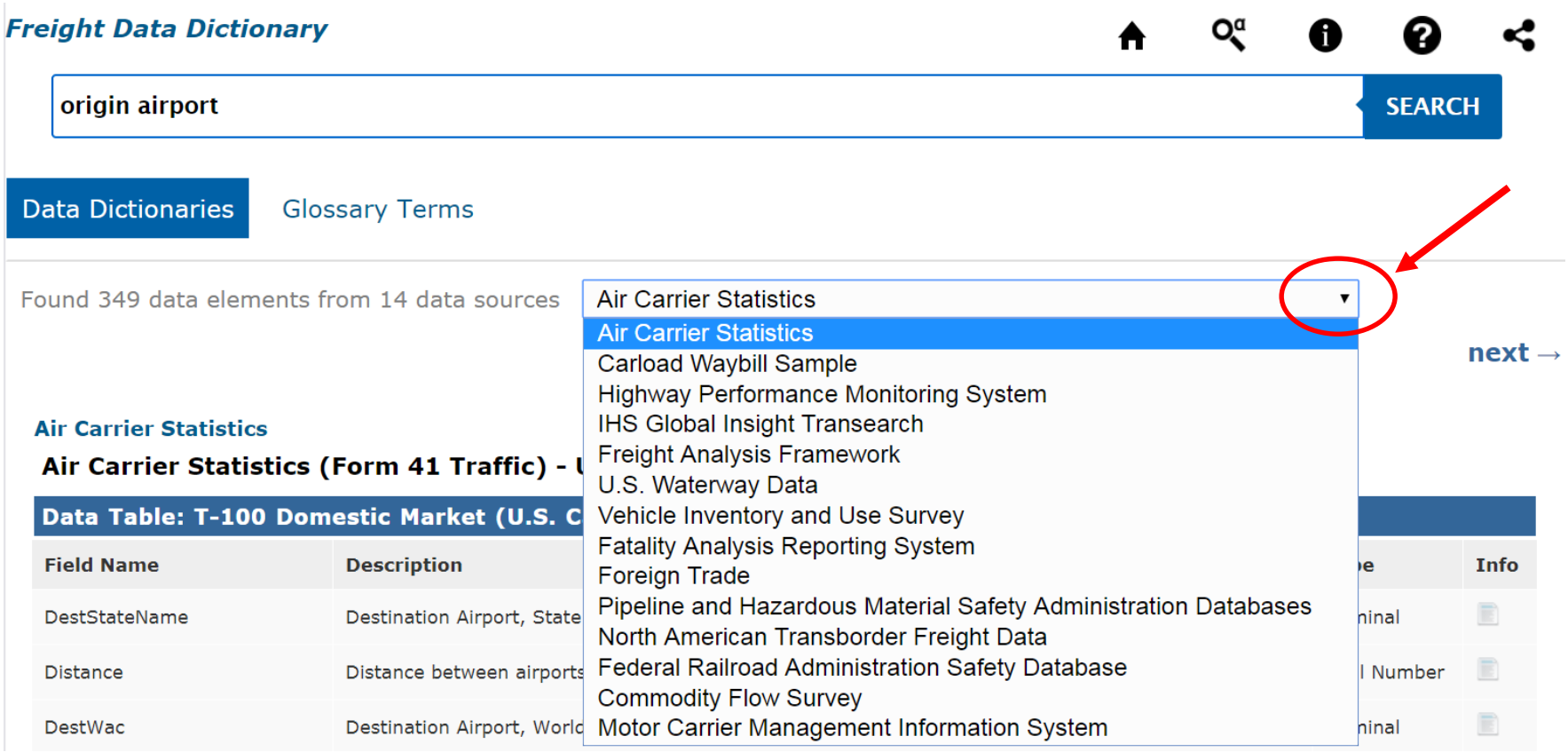
## *Origin Airport 1*




Data Dictionary Number and Sources

# FDD Simple Search:

## *Origin Airport 2*




Data Dictionary Sources

# FDD Simple Search:

## *Origin Airport 3*

Found 349 data elements from 14 data sources  Air Carrier Statistics

1 2 3 4 5 ... 14 next →

**Air Carrier Statistics**

**Air Carrier Statistics (Form 41 Traffic) - U.S. Carriers**

**Data Table: T-100 Domestic Market (U.S. Carriers)**

| Field Name | Description | Type | Info |
|---|---|---|---|
| DestStateName | Destination Airport, State Name | Nominal | |
| Distance | Distance between airports (miles) | Real Number | |
| DestWac | Destination Airport, World Area Code | Nominal | |
| DestAirportID | Destination Airport, Airport ID. An identification number assigned by US DOT to identify a unique ai [less] | Nominal | |
| DestAirportSeqID | Destination Airport, Airport Sequence ID. An identification number assigned by US DOT to identify a [less] | Nominal | |
| Origin | Origin Airport | Nominal | |

DestCityMarketID Dest OriginCityName OriginState OriginStateFips OriginCityMarketID OriginStateName OriginWac OriginAirportID OriginAirportSeqID

Data Dictionary Source Table

# FDD Simple Search:

## *Origin Airport 4*

**Data Table: T-100 Domestic Market (U.S. Carriers)**

| Field Name | Description | Type | Info |
|---|---|---|---|
| DestStateName | Destination Airport, State Name | Nominal | |
| Distance | Distance between airports (miles) | Real Number | |
| DestWac | Destination Airport, World Area Code | Nominal | |
| DestAirportID | Destination Airport, Airport ID. An identification number assigned by US DOT to identify a unique ai [less] | Nominal | |
| DestAirportSeqID | Destination Airport, Airport Sequence ID. An identification number assigned by US DOT to identify a [less] | Nominal | |
| Origin | Origin Airport | Nominal | |

DestCityMarketID Dest OriginCityName OriginState OriginStateFips OriginCityMarketID OriginStateName OriginWac OriginAirportID OriginAirportSeqID

Similar Elements | Complete Table Profile

Added Elements

# FDD Simple Search:

## *Origin Airport 5*

### Similar Elements

**Similar "Place Identifier" Data Elements** Close (X)

| Field Name | Description | Type | Info |
|---|---|---|---|
| Dest | Destination Airport | Nominal | |
| Origin | Origin Airport | Nominal | |
| DestWac | Destination Airport, World Area Code | Nominal | |
| OriginWac | Origin Airport, World Area Code | Nominal | |
| DestState | Destination State Code | Nominal | |
| OriginState | Origin State Code | Nominal | |
| DestCityName | Destination City | Nominal | |
| DestStateFips | Destination State FIPS (U.S. Federal Information Processing Standard Codes) | Nominal | |
| DestAirportID | Destination Airport, Airport ID. An identification number assigned by US DOT to identify a unique ai [...] show more | Nominal | |
| DestStateName | Destination Airport, State Name | Nominal | |
| OriginCityName | Origin City | Nominal | |
| OriginStateFips | Origin State FIPS (U.S. Federal Information Processing Standard Codes) | Nominal | |

# FDD Simple Search:

## *Origin Airport 6*

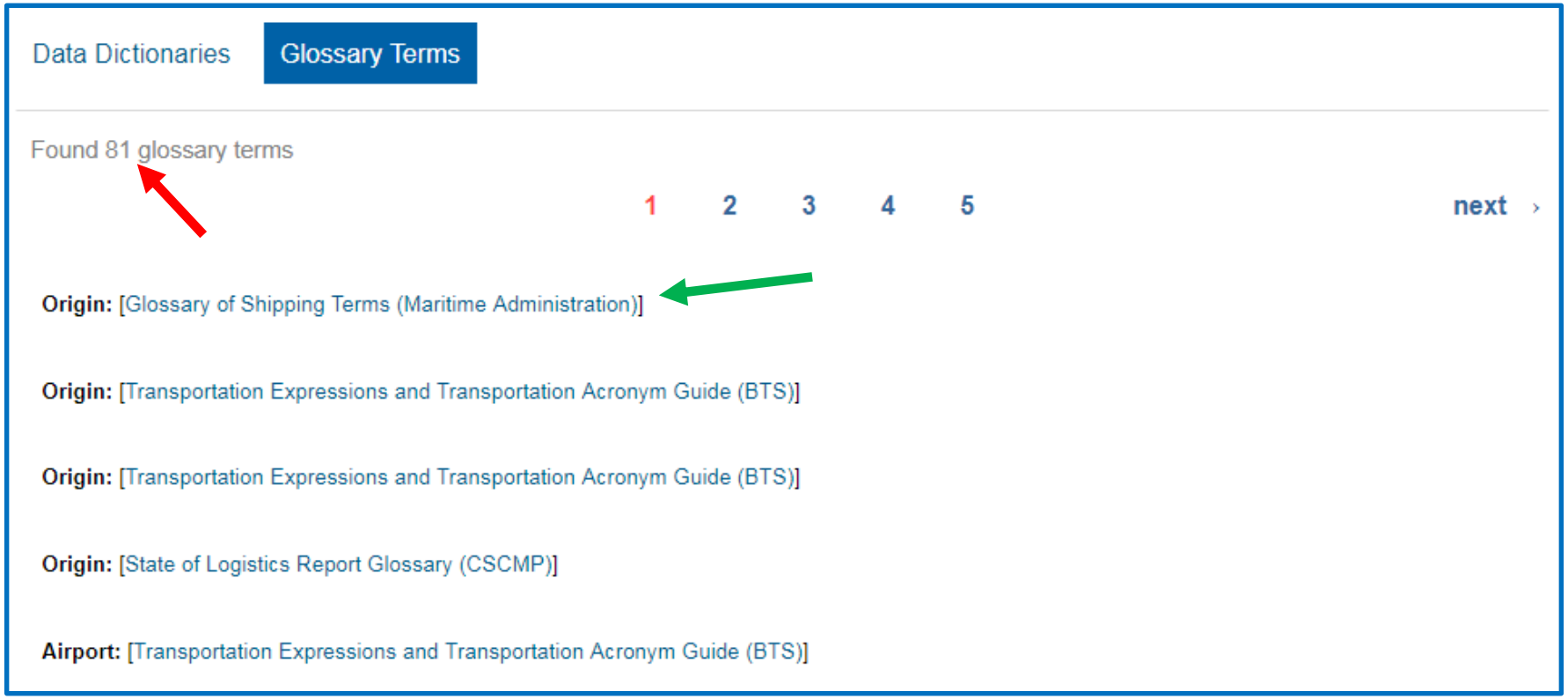
### Complete Table Profile

**Complete Data Table Profile** Close (X)

| Field Name | Description | Type | Info |
|---|---|---|---|
| DepScheduled | Departures Scheduled | Integer | |
| DepPerformed | Departures Performed | Integer | |
| Payload | Payload, in Pounds | Real Number | |
| Seats | Available Seats | Integer | |
| Passengers | Non-Stop Segment Passengers Transported | Integer | |
| Freight | Non-Stop Segment Freight Transported (pounds) | Real Number | |
| Mail | Non-Stop Segment Mail Transported (pounds) | Real Number | |
| Distance | Distance between airports (miles) | Real Number | |
| LoadFactor | Load Factor: Ratio of Passenger Miles to Available Seat Miles | Ratio | |
| RampToRamp | Ramp-to-Ramp Time, in Minutes | Integer | |
| AirTime | Air Time, in Minutes | Integer | |
| UniqueCarrier | Unique Carrier Code. When the same code has been used by multiple carriers, a numeric suffix is used [...] show more | Nominal | |

# FDD Simple Search Example

## Glossary Terms



# Information Architecture

# Sources

- 28 sources compiled (2 commercial, the rest public)
- 6,322 total number of data elements selected
- For each data source, the minimum required entities are the data source name, a table containing the elements, and the elements themselves.

# Sources

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.

# Classification schema

- Role-Based Classification Schema (RBCS) organizes and categorizes data elements across multiple data sources
- Top level groups derived from analyzing freight data classification schema
- Secondary level groups differentiates data elements that *identify* objects from data elements that *describe the features* of an object

# Classification schema

Primary, top level groups

- Commodity
- Event
- Humans
- Industry
- Link
- Mode
- Place
- Time
- Unclassified (elements that do not fit in other roles)

Commodities (C) generated by the industry (I) are moved by various transport modes (M) from one place (P) to another (P) along the transportation network (L) within a time period (T). During the transport process, a chain of possible events (E) may occur that involve various stakeholders or individuals (H).

# Classification schema

Secondary classification groups

- Time elements: time period for reporting or freight movement
- Place elements: O-D freight movement or event location
  - Place identifier (e.g. city name, county, state, country ... or geo point)
  - Place feature (e.g. population, area)
- Commodity elements
  - Commodity identifier (standard commodity codes)
  - Commodity feature (e.g. liquid, bulk, value)
- Link elements
  - Link identifier (e.g. roadway name, waterway name)
  - Link feature (e.g. width, length)

# Classification schema

Secondary classification groups, cont'd.

- Mode elements
  - Mode identifier (e.g. truck, rail, air, pipeline)
  - Mode feature (e.g. unit train, vehicle class)
- Industry elements
  - Industry identifier (NAICS, SIC)
  - Industry feature (e.g. number of employees, sales)
- Event elements
  - Event identifier (e.g., an accident report number, a dredging operation)
  - Event feature (e.g., number of fatalities, number of port calls)
- Human elements
  - Human identifier (e.g., investigating officer, reporting agent)
  - Human feature (e.g., drunk driver)

# Glossaries

- 13,554 terms from 13 glossaries compiled into a single glossary
- Entries include glossary terms and their definitions, link to glossary term source

# Recommended Data Types

| Data Type | Description |
|---|---|
| Nominal | Values exist in name only, can be counted not measured |
| Binary | Values involve 2 things (e.g. yes or no, true or false) |
| Date/Time | Time of day, day of week or month, year, time period |
| Real Number | Values can be measured, can be expressed in non-whole numbers (miles, tonnage, …) |
| Integer | Values expressed only in whole numbers (number of trucks) |
| Currency | Monetary values |
| Ratio | Relation between 2 numbers (e.g. passenger miles per available seat miles) |
| Percentage | Values expressed as fraction of 100 (e.g. percentage truck traffic) |
| Geometry | Representation of GID data (e.g. point, line, polygon) |

# Implementation

# Objective

The primary goal of this acquisition was to provide the solution and services necessary for BTS to offer freight vocabulary control for transportation industry and community in the manner of efficiency, agility, innovation, and potential cost savings.

# Timeline

- BTS received the finished project from Texas University at Austin in 2015.
- BTS allocated the funding for the Migration of the System to Microsoft Azure Cloud in 2016.
- The National Transportation Library migrated the system to the DOT internal testing environment in mid-2017.
- NTL created the Project Charter, Performance Based Statement of Work (PBSW), and other related documents in late summer 2017.
- The Project was approved and funded in October 2017 for FY 2018.

# Milestones

- Setup the Freight Data Dictionary in DOT network (early 2017)
- Convert Unix based oracle to Windows based Oracle (early 2017)
- Convert Oracle to  Microsoft SQL Server in Azure Cloud (mid 2017)
- Implement Azure Index to the application (mid 2017)
- Recode the application in PaaS in Azure (early 2018)
- System testing (April 2018)
- Develop Web based Public Access API (May 2018)
- Move to staging and Production environment (May 2018)

# Future

# Challenges

- FDD lacks export feature
- Units of measurement are US
- Data elements cannot be displayed individually
- Source code not available
- Search is very simple – no Boolean operators
- No user documentation
- Point of contact for each data source not identified
- Planned as a collaborative platform

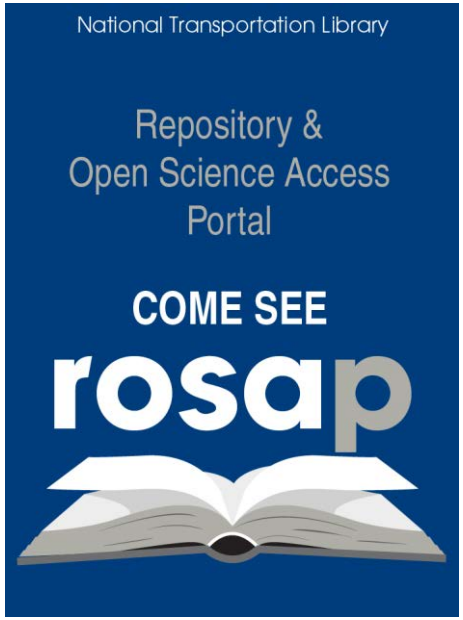
# Future opportunities

- Collaborate with freight data community on governance and submission of new terms
- Forthcoming project for Federal Aviation Administration data dictionary using the same technology
- Planned enhancements:  Boolean search, individual source search, download function, term suggestion form
- Outreach

# Questions or comments?

**ROSA P**

https://rosap.ntl.bts.gov/




**Freight Data Dictionary**

https://fdd.bts.gov/freight-data-dictionary/

# Contact us

**Mary Moulton**
Digital librarian
https://orcid.org/0000-0002-1791-068X
mary.moulton@dot.gov
202-366-0303

**Leighton Christiansen**
Data curator
https://orcid.org/0000-0002-0543-4268
leighton.christiansen@dot.gov
202-366-2759

**Xin Wang**
Systems librarian
xin.wang@dot.gov
202-366-9014


https://transportation.libanswers.com/

# References


- Walton, C Michael; Seedah, Dan P K; Choubassi, Carine; Wu, Hui; Ehlert, Andy; Harrison, Robert; Loftus-Otway, Lisa; Harvey, Jim; Meyer, Joel; Calhoun, Jacob; Maloney, Lucia; Cropley, Stephen; Annett, Ford. **Implementing the Freight Transportation Data Architecture: Data Element Dictionary.** NCFRP Report, Issue 35, 2015, 161p. http://trid.trb.org/view/1367451
- **Freight Data Sharing Guidebook**. NCFRP Report 25, Cambridge Systematics; North River Consulting Group; University of Washington, Seattle, Issue 25, 2013, 68p. http://trid.trb.org/view/1251804
- Quiroga, Cesar; Koncz, Nicholas; Kraus, Edgar; Villa, Juan; Warner, Jeffery; Li, Yingfeng; Winterich, David; Trego, Todd; Short, Jeffrey; Ogard, Elizabeth. **Guidance for Developing a Freight Transportation Data Architecture.** NCFRP Report, Issue 9, 2011, 105p. http://trid.trb.org/view/1085296