LET'S STOP TALKING ABOUT TRANSPORTATION DATA LIKE IT'S NEW

November 15, 2019
Agenda

Reflecting on five years as Chief Data Officer

Why is this moment different?

Tools for the way ahead
Reflecting…

...ON FIVE YEARS AS CHIEF DATA OFFICER
Imagine if I walked in and said...
My first three months…

August

Assess
• Survey OA CIOs on “state of the practice” for data management, open data, etc.
• Survey key stakeholders on role of the CDO for external engagement, standards development, etc.

September

Review
• Compile survey results
• Identify gaps and desired end state
• Compare against CDO duties
• Recommend role of the CDO in achieving end state

October

Document & Implement
• Develop program management plan with key milestones

Invest in a baseline to ensure results are measurable
What I found...

Multiple data & information management efforts underway across the Department

- FAA Enterprise Information Management
- FHWA Data Governance and Business Planning
- FRA IT & Data Management Modernization
- FMCSA Data Management Program & Service Oriented Architecture
- MARAD Information Management Authority & Service Oriented Architecture
- NHTSA Enterprise Data Management & Corporate Information Factory
- PHMSA Data Governance
- Geospatial Coordination Council
- DATA Act Data Standards Working Groups

Common challenges, varied approaches
Why...

...IS THIS MOMENT DIFFERENT?
The Future of Data Analysis (1962)

Four major influences act on data analysis today

- The formal theories of statistics
- Accelerating developments in computers and display devices
- The challenge of more and ever larger bodies of data
- The emphasis on quantification in an ever wider variety of disciplines

- John Tukey
The present of data analysis

Four corollaries

The formal theories of statistics are freely available

Accelerating cost decreases in and ubiquity of computers and display devices

The challenge of more diverse and ever larger bodies of data

The emphasis on quantification and accountability for results in an ever wider variety of disciplines
Three Truths

The data is as small as it will ever be right now

The data has never been better than it is right now

The data may still not be useful for your purposes
Tools

FOR THE WAY AHEAD
One Axiom

“Using data effectively is not just about which database you use or how many data scientists you have on staff, but rather it’s a complex interplay between the data you have, where it is stored and how people work with it, and what problems are considered worth solving.”

DJ Patil and Hilary Mason

Data Driven: Creating a Data Culture
The problems worth solving

- Improve the collection, management, and integration of **data**
- **Identify risks** that contribute to fatalities and serious injuries
- **Collaborate with stakeholders** to foster changes to the transportation ecosystem
Safety Data Initiative Focus Areas

**Integrate** existing DOT data and new “big data” sources

Use advanced data analytics to provide **predictive insights** into safety risks

Create **data visualizations** to help policymakers arrive at solutions
Lessons Learned

• There is a **wealth of data outside of the federal government** that are not fully leveraged

• Private sector data could **help the Department understand** what is happening on the nation’s roadways

• Persistent safety issues can be further illuminated through **new data to contextualize** safety risk
The data you have

How Standards Proliferate:
(See: A/C chargers, character encodings, instant messaging, etc)

Situation: There are 14 competing standards.

14?! Ridiculous! We need to develop one universal standard that covers everyone's use cases. Yeah!

Soon:

Situation: There are 15 competing standards.
A simple specification

• Developed in partnership
  – Portland TriMet
  – Google

• Loosely specified
  – To ease adoption
  – To facilitate a minimum viable use case
Automated Vehicles and Data

- Provides new multimodal safety guidance, clarifies policy and roles, and outlines how to work with U.S. DOT as automation technology evolves

- Calls on stakeholders to identify opportunities for voluntary data exchanges

- Features efforts aimed at enabling voluntary data exchanges
The Work Zone Data eXchange (WZDX)

1. Simple, Open Specification
2. Broadly Adopted
3. Saves Lives
Where you keep your data

Secure Data Commons

The Secure Data Commons (SDC) is an online data warehousing and analysis platform for transportation researchers. On this portal, researchers can take advantage of pre-established programming environments to access and analyze a growing set of transportation-related data sets.

- Provides secure access to data and enables the ability to conduct research and analysis on these data sets
- Security of Data - Moderate level
- Designed for analysis using programming and statistical tool packages
- Analysis is performed within the SDC platform through cloud-based resources

The SDC platform is being developed as a collaborative environment for traffic engineers, researchers, data scientists, and anyone who is interested in carrying out research and analysis on different datasets related to traffic, weather, crashes, and others.
How you work with data

https://whatsthebigdata.files.wordpress.com/2015/10/datascience_unicorn.png
The business of data

- Portal
- Business Intelligence & Analytics
- Toolkits
- Master & Reference Data Management
- Privacy, Security, & Quality Assurance
- Governance & Standards
- Metadata (dictionaries, catalogs)
- Platforms (servers, warehouses, lakes)

Esteem and self-actualization

Love and belonging

Safety

Food and water

Where most of your data users want to start

U.S. Department of Transportation
Office of the Chief Information Officer
Questions