DATA GOVERNANCE

THE FOUNDATION FOR AWESOME ANALYTICS

April Blackburn
Chief of Transportation Technology
Florida Department of Transportation

TOPICS

1. How Big Data is Different
2. Varieties of Data
3. What is Data Governance
4. ROADS
5. Data Governance Considerations
6. People Structure is Essential
7. Other Considerations
8. Conclusion
9. Questions

BIG DATA ANALYTICS

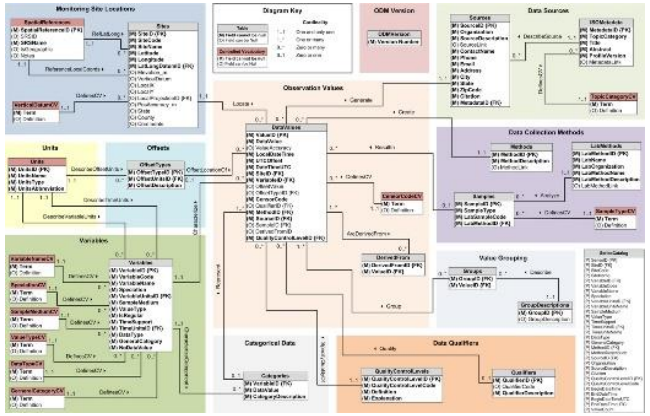
Big Data is Different

- Big Data technology allows us to capture and process **massive amounts of transportation data** from numerous sources.
- New ways to **manage unstructured and semi-structured** information is now available that was previously unexplored such as Data Lakes, In-Memory Processing, etc.
- The highest value comes from **making connections across data sources**, types, functions, etc. versus standalone environments / applications.
- Traditional BI/Analytics informs of **past activities and performance**.
- Big Data and Advanced Analytics is typically a **future-oriented view** and leverages some type of statistics and algorithms to present a Simulation, Descriptive Analytics or Predictive Analysis.

BIG DATA ANALYTICS

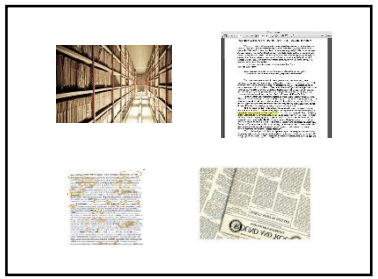
Varieties of Data

Big Data requires **governance** of large volumes of structured, unstructured and semi-structured data.



Structured Data

+



Legacy Documents

Web and Social Media

Images & Video

Highway Traffic Flow

Traditional BI/Analytics

IT typically provides information via extracting data from a data warehouse and presents to the business in pre-defined reports.

Big Data Analytics

Analytical platform is provided that lends itself to be more variable/flexible to analyze data from multiple sources.

DATA GOVERNANCE OVERVIEW

What is Data Governance?

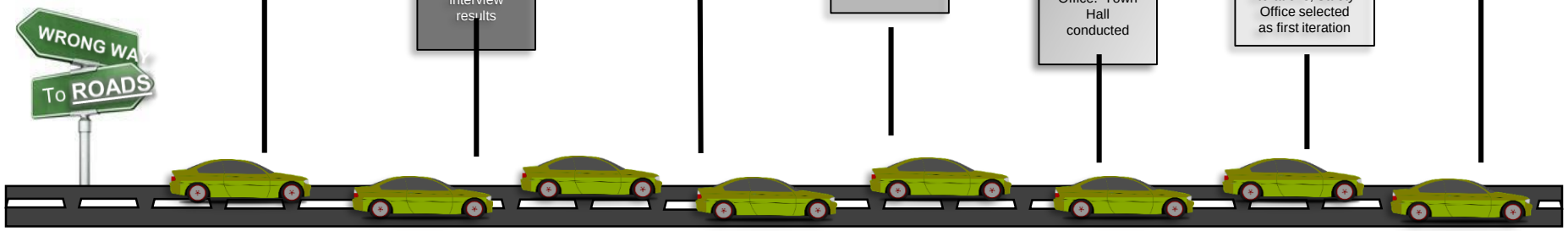
Data Governance (DG) refers to the overall management of the availability, usability, integrity, and security of the data employed in an enterprise. A sound data governance program includes a governing body or council, a defined set of procedures, and a plan to execute those procedures.

FDOT's vision for data governance is based on a foundation of reliable and consistent data that helps drive business, is well understood by internal and external stakeholders, and is collected, managed, and clearly / openly communicated.

ROADS MISSION

The goal of the ROADS initiative is to improve data reliability and simplify data sharing across FDOT to have readily available and accurate data to make informed decisions.

THE ROADS JOURNEY



MARCH 2015

Start of the ROADS Journey – moving FDOT to Reliable, Organized and Accurate Data Sharing

SPRING 2015

270 interviews and 237 surveys of FDOT identifying 63 gaps in data and information

SUMMER/FALL 2015

Establishment of data governance structure

FALL/WINTER 2015

Determination of Tool Requirements completed after reviewing survey and interview results

SPRING/SUMMER 2016

"Invitation to Negotiate" completed with interest from 45+ vendors. Review process and "short list" completed

SUMMER 2016

Applications and Reports Inventory assessments

SUMMER/FALL 2016

Charters created to define responsibilities and the path forward for the EDS and ROADS Executive Team

FALL 2016

ROADS Town Hall and Knowledge Sharing Sessions occurred to educate stakeholders on Metadata concepts

FALL 2016

Three short listed vendors performed test cases and oral presentations for ITN

WINTER 2017

ROADS is aligned under the Civil Integrated Management Office. Town Hall conducted

SPRING 2017

Public meeting announced SAS as intended award vendor for BI/DW ITN

SUMMER 2017

SAS contract executed. Pilot of select data governance processes. BI/DW Project – Initiation Phase and project kick-off including scoping of future iterations, Safety Office selected as first iteration

FALL 2017

Begin Knowledge Sharing Sessions on select data governance Best Practices

WINTER 2017-SUMMER 2021

BI/DW Project – Iterative Implementation Phases and continued Knowledge Sharing

Updated as of July 2017

BIG DATA ANALYTICS

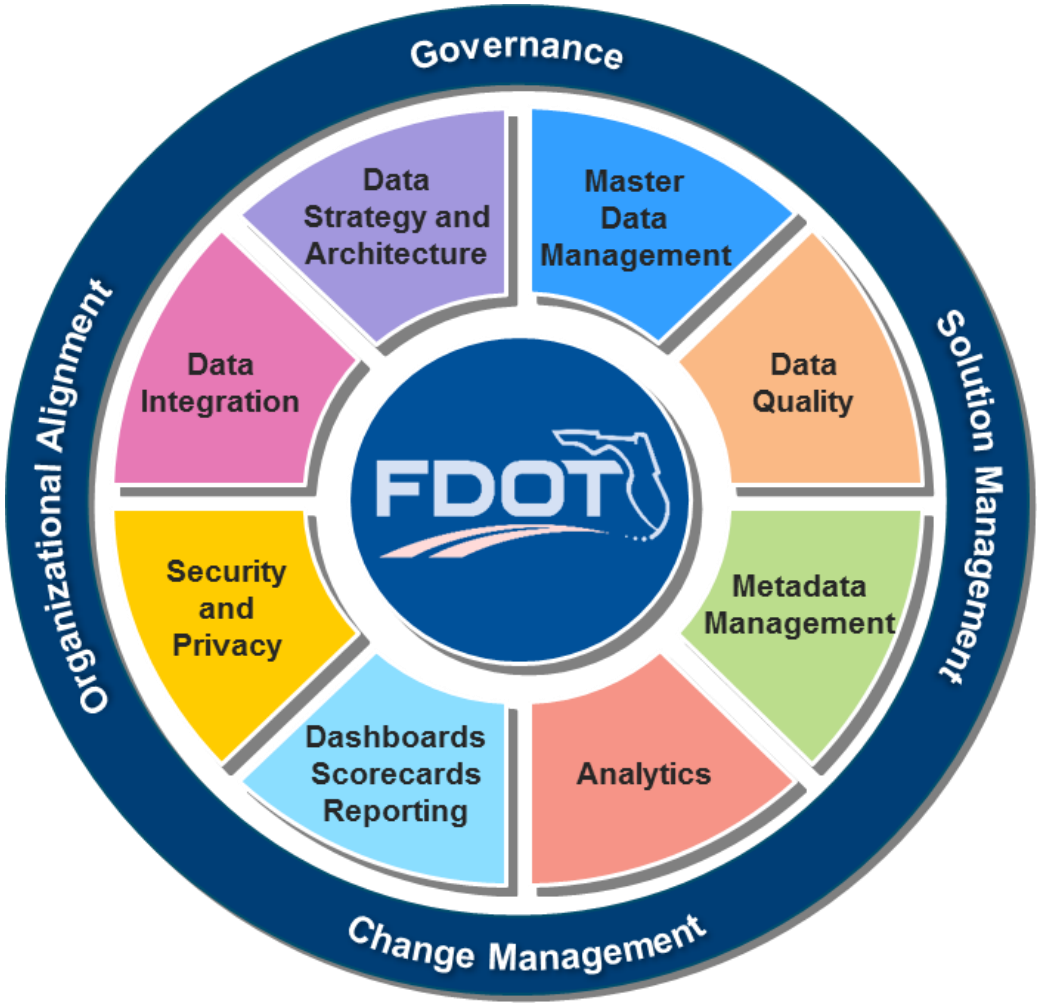
Data Governance Considerations

- A solid **Information Management Framework** is critical to help establish confidence in the “Big Data” that will be used in Advanced Analytics models that are put into production.

- To get a full 360-degree view of a citizen, development project, or highway network, consider consolidating disparate data into a data lake where raw data remains available for analysis.

- Be intentional about the amount of governance / cleansing that is done on the data. Be careful not to overdo it and remove some of the analytic value that would otherwise be found in the data.

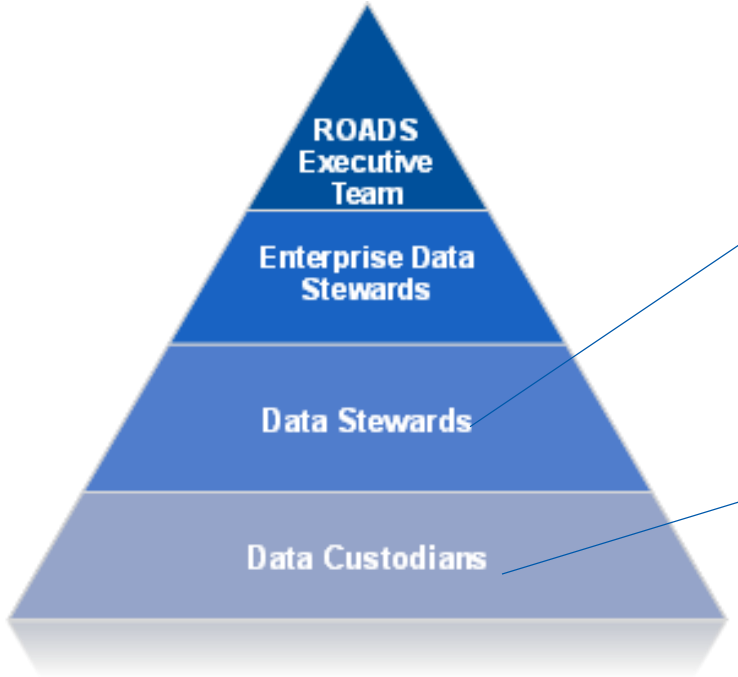
- Consider staging data in 3 categories of governance / cleansing
  - Raw – Data straight from source, immediately available
  - Basic – Some initial governance / cleansing for select non-financial related decision making
  - Gold Standard – Fully governed / cleansed data.



BIG DATA ANALYTICS

People Structure is Essential

Managing a formal data governance structure to make key decisions related to Big Data is critical.



Data Stewards are business-focused experts in their functional area. They understand their business, their data, and what data they create, store and use.

Data Custodians are cross-functional, technical experts responsible for the day-to-day management of data. They are familiar with the technical underpinnings of FDOT data.

BIG DATA ANALYTICS

Other Considerations

- Identify the analyst segments (audiences) you need to serve. Research analysts need raw data, financial analysts probably need more governed data.
- Define the level of governance needed for each audience and data source (Ex - Twitter: meta-data around hashtags so you can mine). Publish a 3-5 step standard based on Raw to Gold (gold being the top standard everyone can trust).
- Some data sources should not be managed (free, open) BUT you should not certify reports used with those data sets.
- PII tokenization schemes are important to allow exposure of data to analysts while still securing data.
- Look out for new trends in automatic data discovery and tagging to allow automated governance.

CONCLUSION – WRAP UP!

Utilize Data Governance to Implement Big Data Solutions

- Know the meaning of the data regardless of the source, make it accessible to both business and IT, and track lineage from the source to analytics.
- Install common definitions in order to help establish consistent conclusions.
- Understand what portions of the Big Data being collected should be maintained as private or sensitive to meet company compliance.
- Conduct sufficient training throughout the organization to help executives, managers, etc. grasp the fundamentals of Big Data and “Advanced Analytics” so they can ask the right questions of their data analysis teams.

THANK YOU

Questions?

April C. Blackburn
Chief of Transportation Technology
april.blackburn@dot.state.fl.us