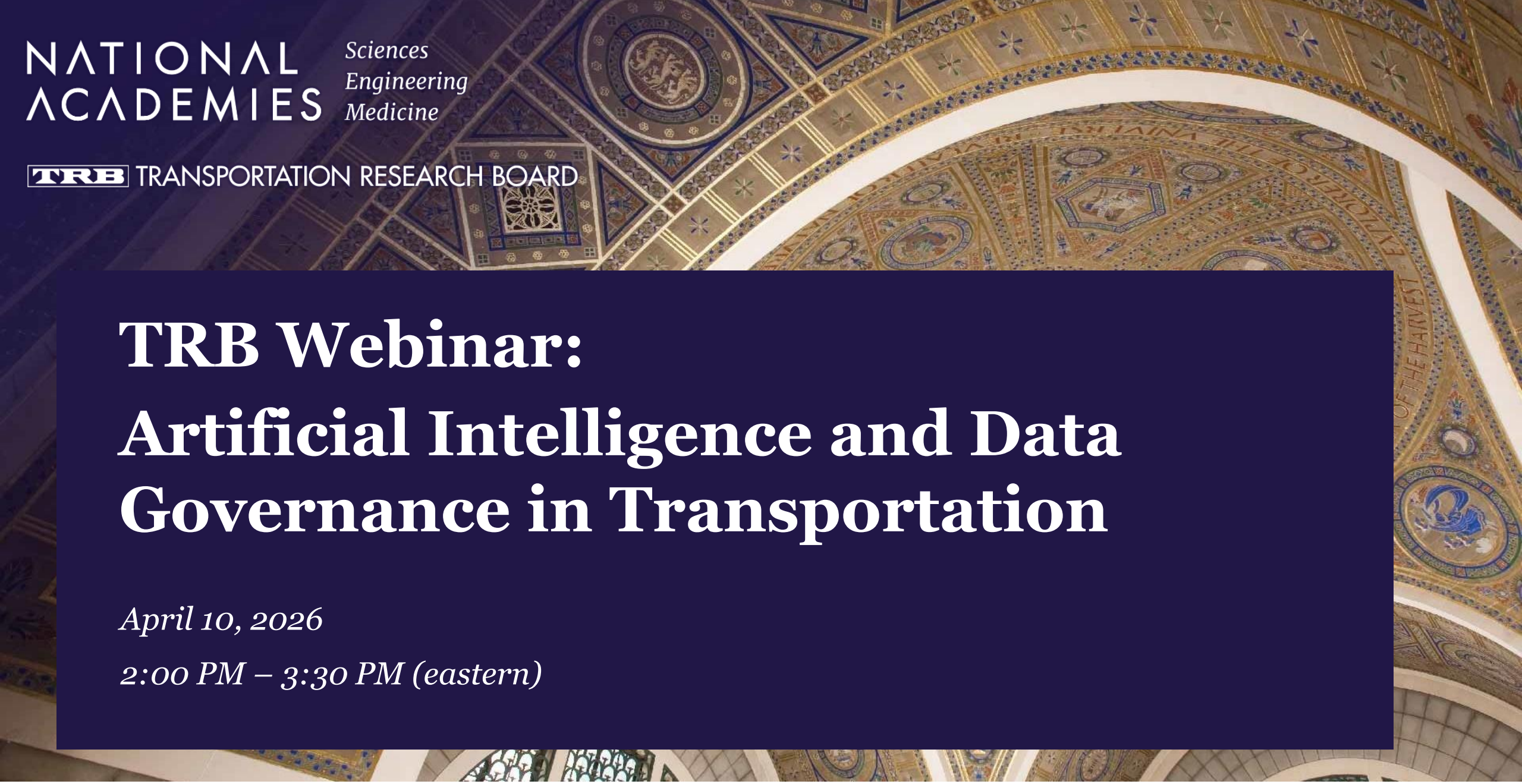


TRB Webinar:
Artificial Intelligence and Data Governance in Transportation

April 10, 2026

2:00 PM – 3:30 PM (eastern)

PDH Certification Information

1.5 Professional Development Hour (PDH) – see follow-up email

You must attend the entire webinar.

Questions? Contact Andie Pitchford at TRBwebinar@nas.edu

The Transportation Research Board has met the standards and requirements of the Registered Continuing Education Program. Credit earned on completion of this program will be reported to RCEP at RCEP.net. A certificate of completion will be issued to each participant. As such, it does not include content that may be deemed or construed to be an approval or endorsement by the RCEP.

ENGINEERING RCEP
REGISTERED CONTINUING EDUCATION PROGRAM

NATIONAL ACADEMIES Sciences Engineering Medicine
TRB TRANSPORTATION RESEARCH BOARD

AICP Credit Information

1.5 American Institute of Certified Planners Certification Maintenance Credits

You must attend the entire webinar

Log into the American Planning Association website to claim your credits

Contact AICP, not TRB, with questions

Purpose Statement

This webinar will examine how AI is being applied across transportation systems. From identifying potential security risks at airports to predicting high-risk roadway locations and detecting trespassers along rail corridors, strong data governance is foundational to these efforts. Presenters will share case studies highlighting successes, challenges, and lessons learned as they advance AI-readiness and strengthen data governance in a rapidly evolving environment.

Learning Objectives

At the end of this webinar, participants will be able to:

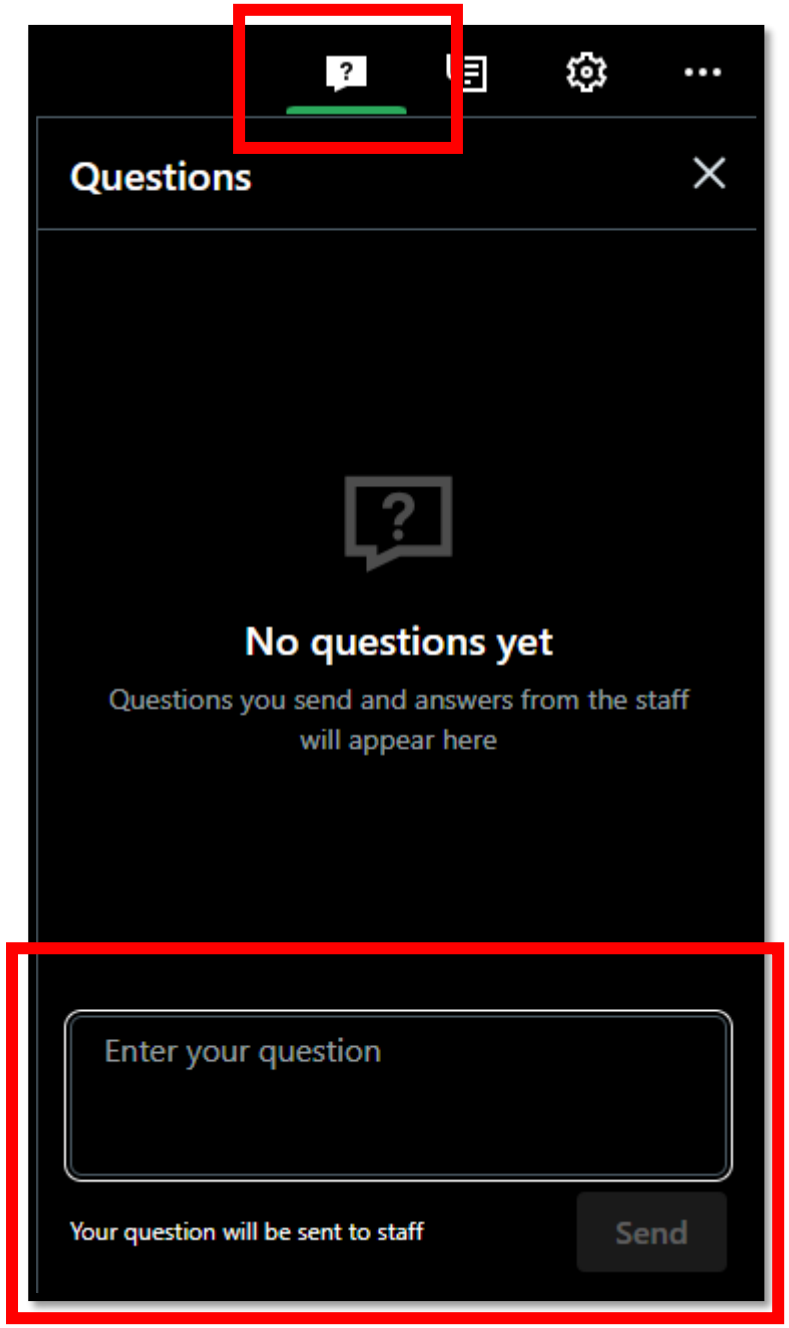
(1) Explain what a data literacy plan is and why it is important for organizations

(2) Describe the core principles that a data governance program can be governed by and what a comprehensive path to AI readiness looks like

(3) Describe the risks to organizations when they do not address data governance issues

Questions and Answers

- Please type your questions into your webinar control panel
- We will read your questions out loud, and answer as many as time allows



Today's Presenters

Jesse Newberry
HNTB
jnewberry@hntb.com

Joel Jundt
South Dakota Department of Transportation
joel.jundt@state.sd.us

Jennifer Volkening
Utah Department of Transportation
jvolkening@utah.gov

Gregory Ciparelli
Connecticut Department of Transportation
gregory.ciparelli@ct.gov

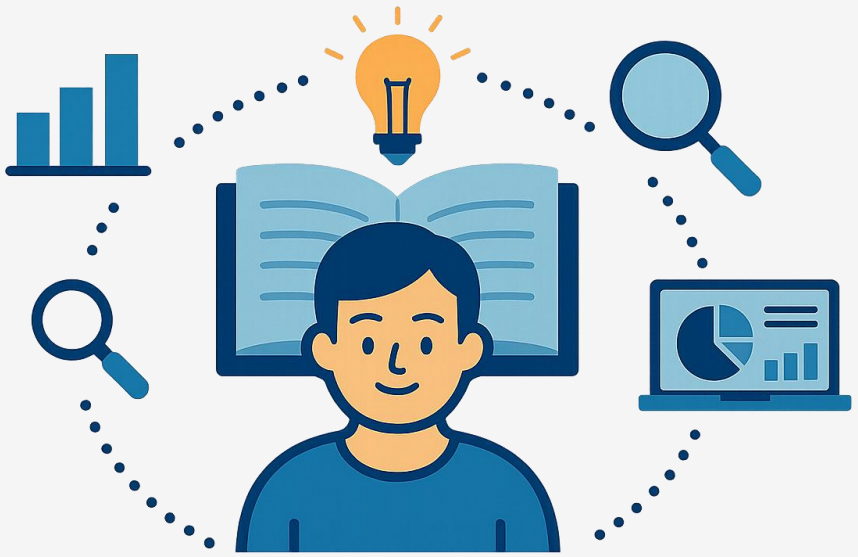
Benmjain McCulloch
Texas Department of Transportation
benjamin.mcculloch@txdot.gov

NATIONAL ACADEMIES Sciences Engineering Medicine
TRB TRANSPORTATION RESEARCH BOARD

Data Literacy is Everyone's Job

Gregory Ciparelli
Connecticut Department of Transportation
4/10/2026

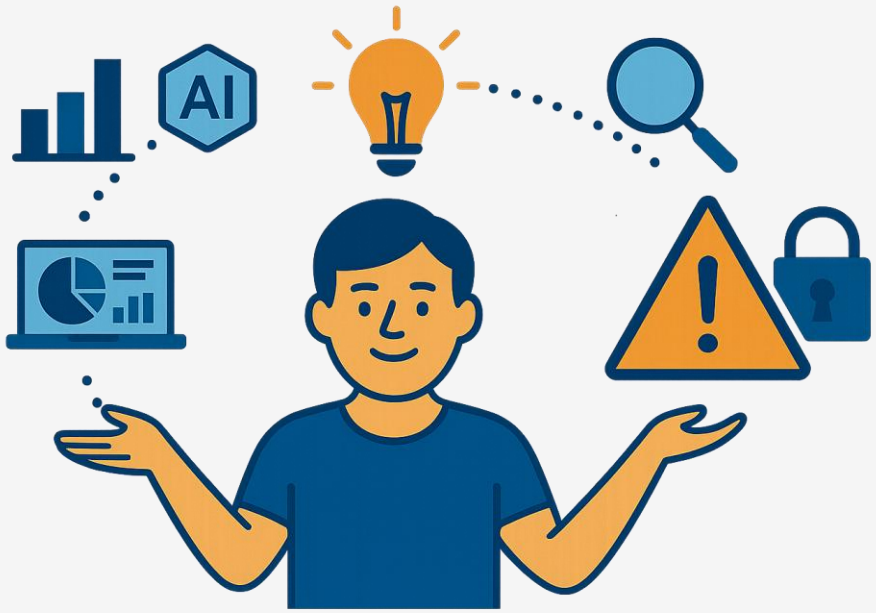
What is Data Literacy?



The ability of employees to understand and work with data and data analysis tools to the appropriate degree that it aids in **sound** analysis and data-driven decision making.

This graphic was generated or refined utilizing authorized AI tools and has been reviewed by CTDOT personnel for accuracy

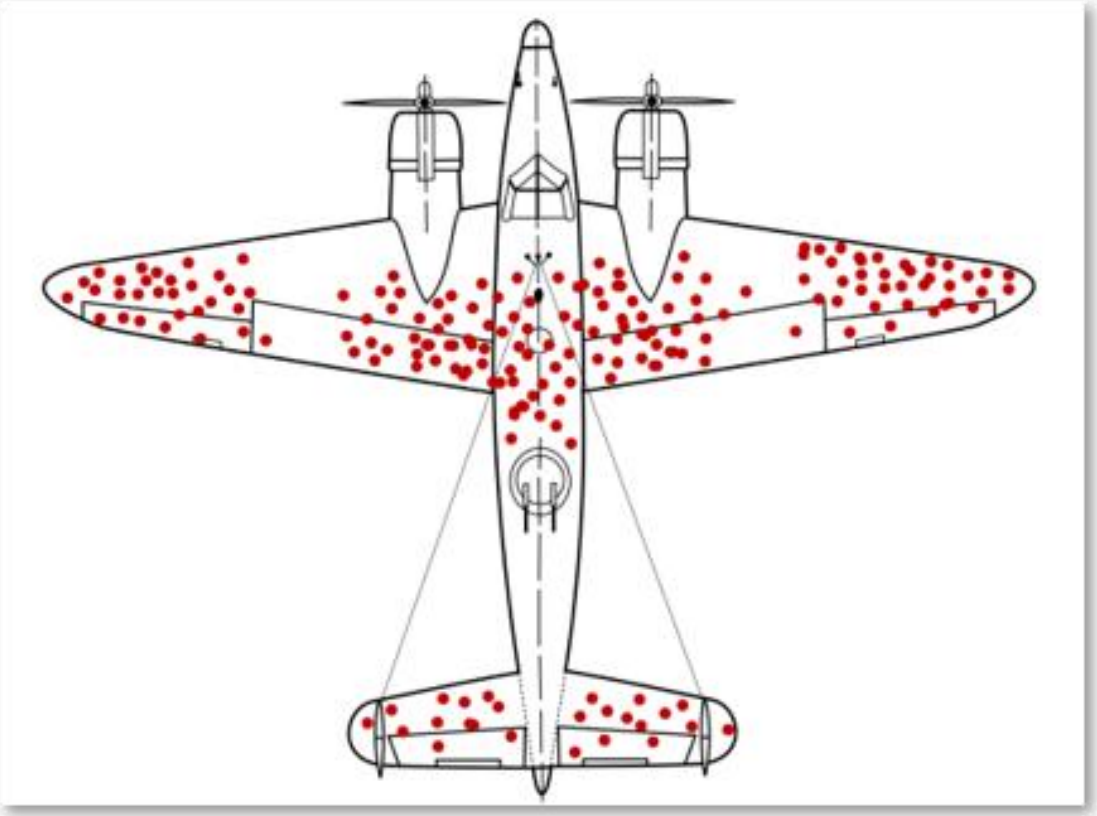
Why is Data Literacy Important?



Increased access to more data and robust analysis tools (including AI) has unleashed a great amount of power for data literate organizations, while simultaneously creating significant risk for those that cannot foster a culture of data and AI literacy.

This graphic was generated or refined utilizing authorized AI tools and has been reviewed by CTDOT personnel for accuracy

What types of risks?



False/pseudo Precision
Visual Framing Mistakes
Confirmation Bias
Lack of Authoritative Data
Data Inconsistency
Base Rate Neglect
Correlation ≠ Causation
Selection Bias
Survivorship Bias
Aggregation Paradox
Overfitting
Post Hoc Assumptions
Poorly Trained Models
PII Exposure

Also, This…

Image from Fox Sports Broadcast

Scale Distortion

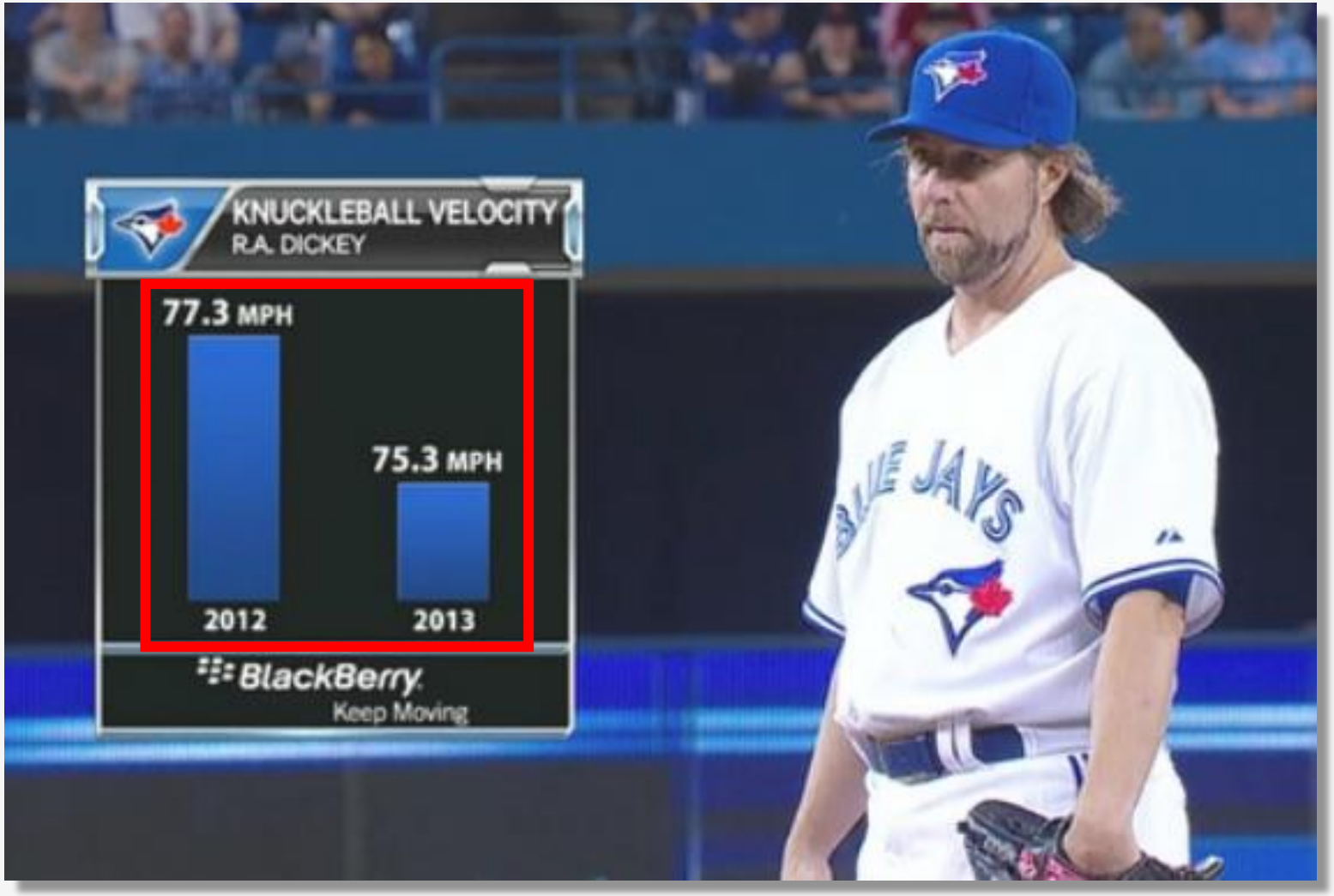
Truncated & Not Labeled Y axis

- Is the baseline 73.66mph?
- Did the speed drop by more than half?

Context of professional sports broadcast - Low Risk

For government entities & public policy outreach?

Can be perceived as purposefully misleading



Good news, AI Can Help

If you know to ask…

“Should I use this graphic in my presentation on data visualization and tell me why or why not.”

Identifies the core issue and makes recommendations – these are basic prompts with lots of context (the photo).

Should I use this graphic in my presentation on data visualization and tell me why or why not?

Copilot · Web search is OFF

Yes — you *can* use this graphic in your data-visualization presentation, and in fact it would make a **strong teaching example**, but only if you use it **as an illustration of what *not* to do**.

Here's why:

• The bar chart **dramatically exaggerates** a small numerical difference.
– The values differ by only 2 mph (77.3 vs 75.3), but
– The bar heights differ by far more than 2.6% in the graphic.
This happens because the y-axis is truncated or scaled improperly.

• The misleading visual scale makes a modest change look like a major drop.
This is one of the most common—and most important—data visualization pitfalls.

• Using a real-world broadcast graphic helps make the lesson relatable and memorable.
It shows that even professional graphics can mislead when scale is mishandled.

Recommendation:
Use the graphic specifically as an example of **poor visualization practice**. Explain the error, show how it misleads the viewer, and (optionally) present a corrected version with a proper y-axis for contrast.

“Can you explain this graphic to me?”

Open & Generic, but photo is well know/documented enough to get a valid response.

Can you explain this graphic to me?

Copilot · Web search is OFF

This graphic is a well-known illustration used to explain **survivorship bias**, a classic statistical reasoning error.

Here's what the image shows and what it means:

1. What the graphic depicts
• The image shows the outline of an aircraft viewed from above.
• Numerous red dots mark locations where returning aircraft were hit by enemy fire.
• These marks represent **damage observed on planes that made it back from missions**.

2. The key insight
The initial instinct—during the historical context in which this type of analysis became famous—was to reinforce the areas with the most bullet holes.
But that conclusion is **incorrect**.

Why?
Because **the only planes available to study were the ones that survived**.
Planes shot in different areas (such as the engine or cockpit region) **did not return**, so those areas appear undamaged in the data—but only because the aircraft that were hit there were lost.

3. The correct interpretation
Rather than reinforcing the heavily marked areas, analysts realized they needed to reinforce the **unmarked areas**, because:

• The red-dot regions show **damage aircraft could survive**
• The blank regions show **damage that caused aircraft to be lost**

Thus, the absence of marks is more meaningful than the presence of marks.

4. Why this graphic is important
It demonstrates how:
• Data can be biased by the way it is collected
• Missing information may be more important than observed information
• Incorrect conclusions arise when we ignore what we *can't* see

Why Is This A Challenge?

Data & Insight Isn't Inherently Easy to Understand

- Its Own Language
- Contains Nuance
- Requires Context
- Has Limitations
- AI Outputs can be black-box in nature

Data/Analysis Confidence is Critical To Sustained Success

- Integrate from Authoritative Systems
- Robust & Actively Managed Metadata
- Define Data Management Processes & Practice Data Governance
- Direct Lines of Communication

Not One Size Fits All

Reading Data/Using AI

Understanding what data is and what it represents

Understand how to generally use configured Artificial Intelligence Tools/Large Language Models

Requires ability to understand the context of the inputs and outputs

Working With Data/Supporting AI

Creating, Collecting, Acquiring, QA/QC

Understand how the data they provide impacts proper use by all supported tools, including AI

Tools can be built by technical staff to support

Organizational Data Literacy

Analyzing/Prompting Data/AI

Filtering, Querying, Aggregating, Instructing AI Tools

Can be handled through analysis specific tools

May require higher level technical understanding and prompt engineering knowledge

Narrating With Data/AI Outputs

Using data/AI to communicate analysis results

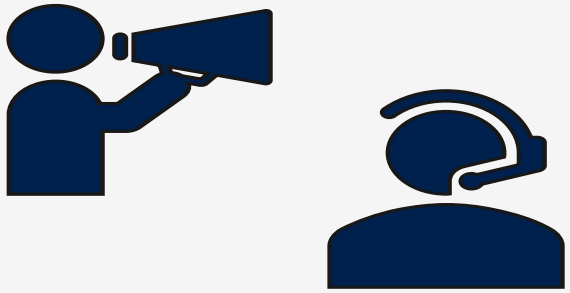
Requires mix of technical skill, effective graphic design, and compelling storytelling to succinctly convey results – check and refine, check and refine

The right level of literacy for the right role in the Organization

CTDOT Path to Literacy

Clear coordination, communication, outreach with Data Community

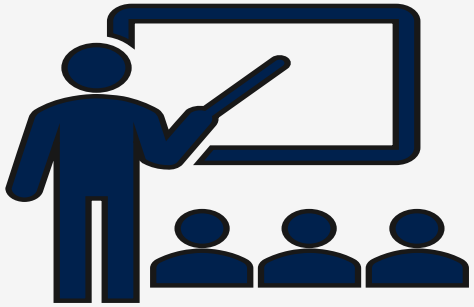
Communicate



Weekly Enterprise Coordination – IT, DM, Users, Stewards, Consultants

SharePoint Communication Page – News Posts, Lists, Documents

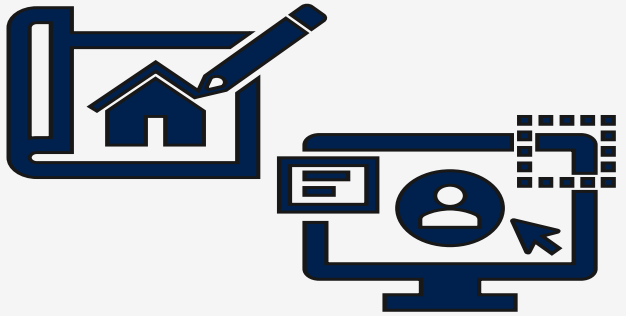
Educate



Regular User Groups – Platform Specific, but OPEN

Guidance/Standards Publication

Design & Build



Business Analysis & Requirements Gathering

Stakeholder Feedback – Agile approach

Data Warehousing – Medallion Architecture

- Data Maturity, Authority, & Access

Internal AI Solution Pilot with Citations

- DOTBOT

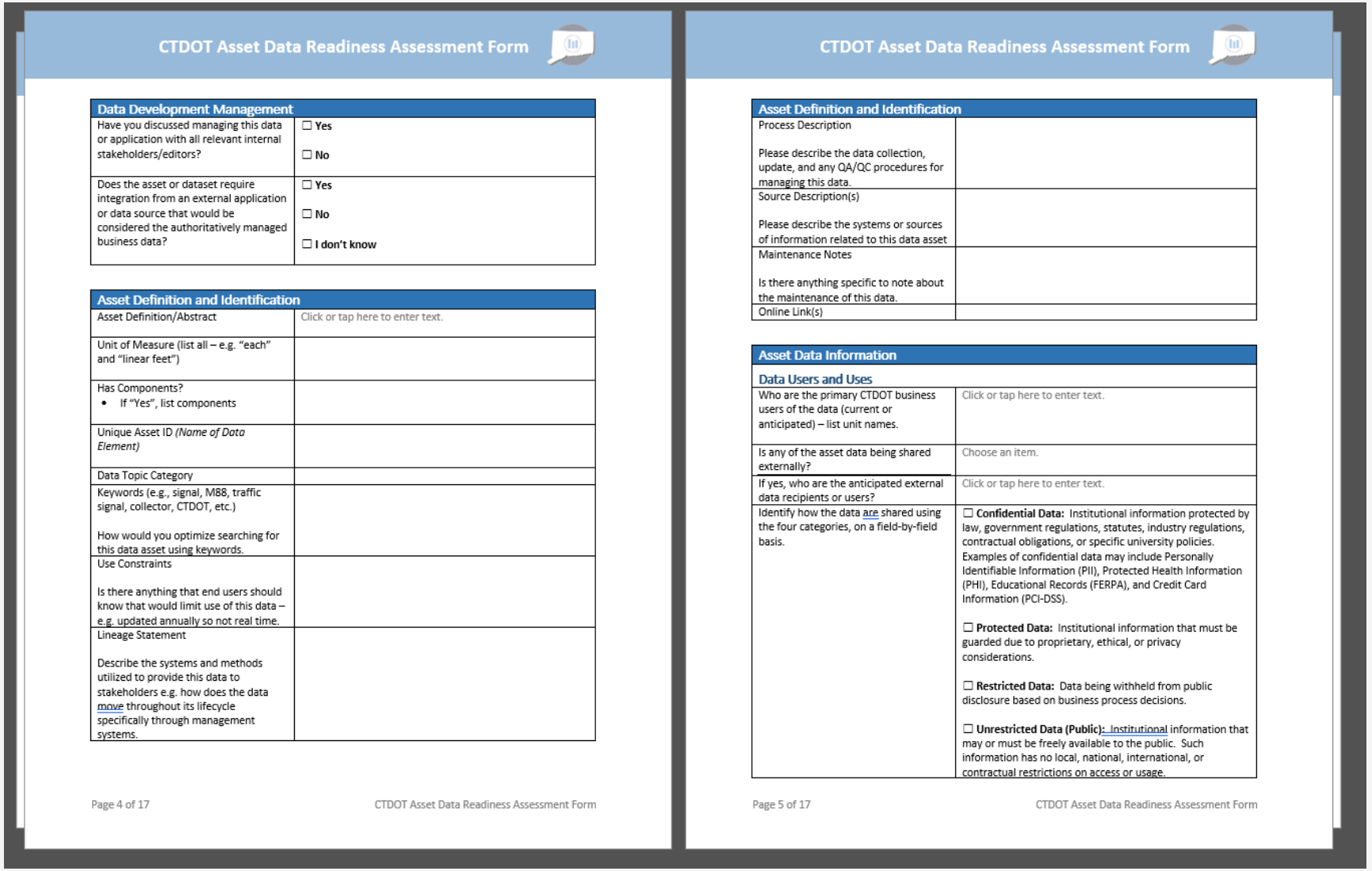
Internal Communication/Coordination/Education

Data Asset Documentation

17 Pages

Connects Data and Business

Documented context and nuance, critical information for users & AI tools



CTDOT Asset Data Readiness Assessment Form

Data Development Management	
Have you discussed managing this data or application with all relevant internal stakeholders/editors?	☐ **Yes** ☐ **No**
Does the asset or dataset require integration from an external application or data source that would be considered the authoritatively managed business data?	☐ **Yes** ☐ **No** ☐ **I don't know**

Asset Definition and Identification	
Asset Definition/Abstract	Click or tap here to enter text.
Unit of Measure (list all – e.g. "each" and "linear feet")	
Has Components? • If "Yes", list components	
Unique Asset ID *(Name of Data Element)*	
Data Topic Category	
Keywords (e.g., signal, M88, traffic signal, collector, CTDOT, etc.) How would you optimize searching for this data asset using keywords.	
Use Constraints Is there anything that end users should know that would limit use of this data – e.g. updated annually so not real time.	
Lineage Statement Describe the systems and methods utilized to provide this data to stakeholders e.g. how does the data move throughout its lifecycle specifically through management systems.	

Page 4 of 17 CTDOT Asset Data Readiness Assessment Form

CTDOT Asset Data Readiness Assessment Form

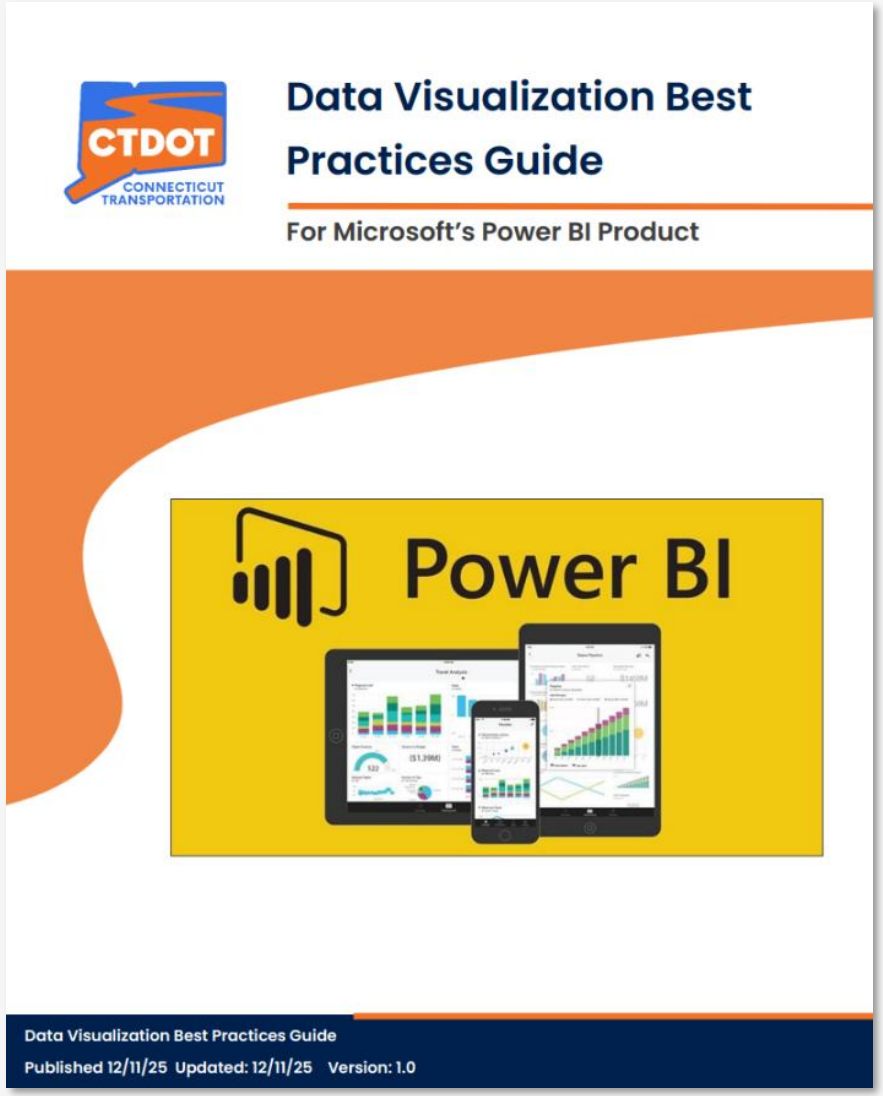
Asset Definition and Identification	
Process Description Please describe the data collection, update, and any QA/QC procedures for managing this data.	
Source Description(s) Please describe the systems or sources of information related to this data asset	
Maintenance Notes Is there anything specific to note about the maintenance of this data.	
Online Link(s)	

Asset Data Information	
Data Users and Uses	
Who are the primary CTDOT business users of the data (current or anticipated) – list unit names.	Click or tap here to enter text.
Is any of the asset data being shared externally?	Choose an item.
If yes, who are the anticipated external data recipients or users?	Click or tap here to enter text.
Identify how the data are shared using the four categories, on a field-by-field basis.	☐ **Confidential Data:** Institutional information protected by law, government regulations, statutes, industry regulations, contractual obligations, or specific university policies. Examples of confidential data may include Personally Identifiable Information (PII), Protected Health Information (PHI), Educational Records (FERPA), and Credit Card Information (PCI-DSS). ☐ **Protected Data:** Institutional information that must be guarded due to proprietary, ethical, or privacy considerations. ☐ **Restricted Data:** Data being withheld from public disclosure based on business process decisions. ☐ **Unrestricted Data (Public):** Institutional information that may or must be freely available to the public. Such information has no local, national, international, or contractual restrictions on access or usage.

Page 5 of 17 CTDOT Asset Data Readiness Assessment Form

Internal Communication/Coordination

User Guide Development



Published Guide to Creating Effective Data Visualization Products

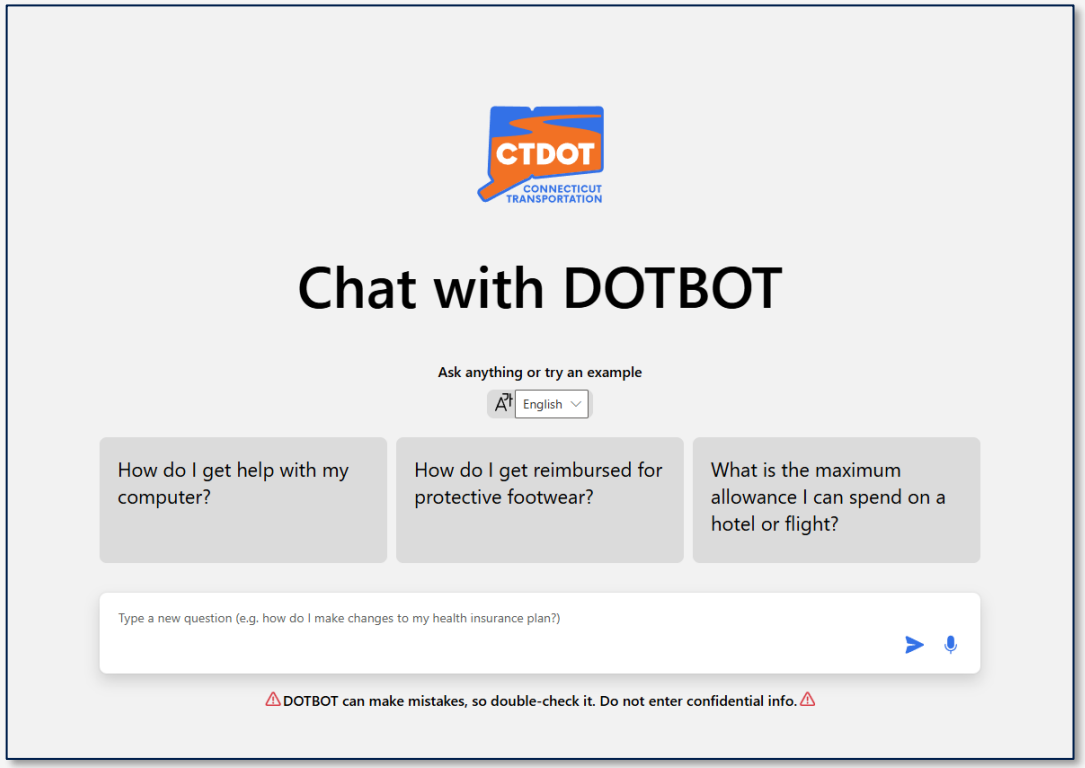
- **ADA Accessibility**
  - Tooltips, color contrast, alt text, etc.
- **Design Styles**
  - Flat, Skeuomorphic, Realistic/Literal
- **Data Storytelling**
  - Consider audience & insight
- **Chart Types**
  - What to use for the type of analysis
- **CTDOT Resources**
  - Branding, logos, colors, fonts
- **External Resources**
  - **Visual Vocabulary**
    - Interactive Chart Selection Guide
  - **Using Colors Effectively**
    - Schemes, Palettes, Rules
  - **General Concepts on Interactive Reports**
    - Reduce clutter, read left to right top to bottom, group elements, etc.

Data Warehouse

- Raw Data
- Clean/Validated
- Aggregated/Business Ready

Accessible to all staff as authoritative data

Internal AI Solution



DOTBOT
CTDOT's Internal RAG AI Chatbot

How RAG Chatbots Work

- **Indexing:** Documents (PDFs, web pages, etc.) are split into chunks, converted into numerical representations (embeddings), and stored in a vector database.
- **Retrieval:** When a user asks a question, the chatbot creates an embedding for the question and searches the database for the most similar document chunks.
- **Augmentation:** The retrieved relevant chunks, along with the original question, are given to the LLM as context.
- **Generation:** The LLM uses this combined information to generate a factual, context-aware answer, often with citations

Key Benefits of a RAG Chatbot

- **Accuracy & Trust**: Reduces hallucinations by grounding answers in specific data, allowing for source attribution.
- **Up-to-Date Info**: Can use fresh data from additional sources without retraining the entire LLM.
- **Domain-Specific Knowledge**: Accesses private or specialized data (e.g. company policies, technical manuals).
- **Control & Compliance**: Offers better control over information sources, crucial for sensitive industries/entities

An Agency Roadmap

Assess employee skills and create a baseline expectation for appropriate roles

- Consumers, Collectors, Stewards, Analysts, Developers, Engineers, Architects, Subject Matter Experts all have different needs

Define and use a common language

- Industry Terminology & Data Terminology
- Data Glossary

Everyone learns differently – consider method of delivery

- Visual Learners, Hands-On DIY, Guided Tutorials, Short & Long Videos
- Provide Applicable Testing Datasets

Support intellectual curiosity & build a culture of constant learning

- Have environments for people to explore ideas and where it's safe to be wrong
- Provide resources for evolving platforms and capabilities, especially as they gain traction

Data literacy & technical literacy are different

- Not everyone needs to have skills of AI developer or prompt engineer, Architect, DBA, GIS Developer, or even proficiency in COTS software configuration

Thank You

Gregory Ciparelli
Chief Data Officer
Connecticut Department of Transportation

Data Governance and AI: Key Takeaways from TxDOT

Texas Department of Transportation

4/10/2026

TxDOT Data Governance Goals

- Establish a strong data governance and procurement framework.
- Ensure high-quality, accurate, and trusted data.
- Standardize data and improve interoperability across systems.
- Modernize data tools, infrastructure, and practices.

Making Data Available at TxDOT

- TxDOT's enterprise data platform (EDP) built on Snowflake.
- Includes data from 30+ key systems: financial, project development, construction, pavement, consultant contracts, crash reports.
  - In progress: average annual daily traffic count, bridge, traffic management center, utility, and maintenance data.
- **Unified datasets help unlock AI capabilities.**

Agency Challenges

- 25 districts and 32 Divisions statewide using a variety of tools or purchasing data to meet their needs.
- As legacy systems are replaced there is inconsistent quality of data and metadata in the new systems.
- Business rules are not consistently applied during data entry.

Lack of Data Governance Leads to AI Risks

- Misinformed asset and safety prioritization.
- Incorrect predictions for costs and schedules.
- Dependence on vendors, no ability to audit AI models, lack of transparency.
- Higher risks due to unjustifiable or biased decisions.
- Pilot projects failing to scale, low adoption

Image by DALL-E 3

Key Takeaway

AI is not a software deployment,

it is a data maturity program.

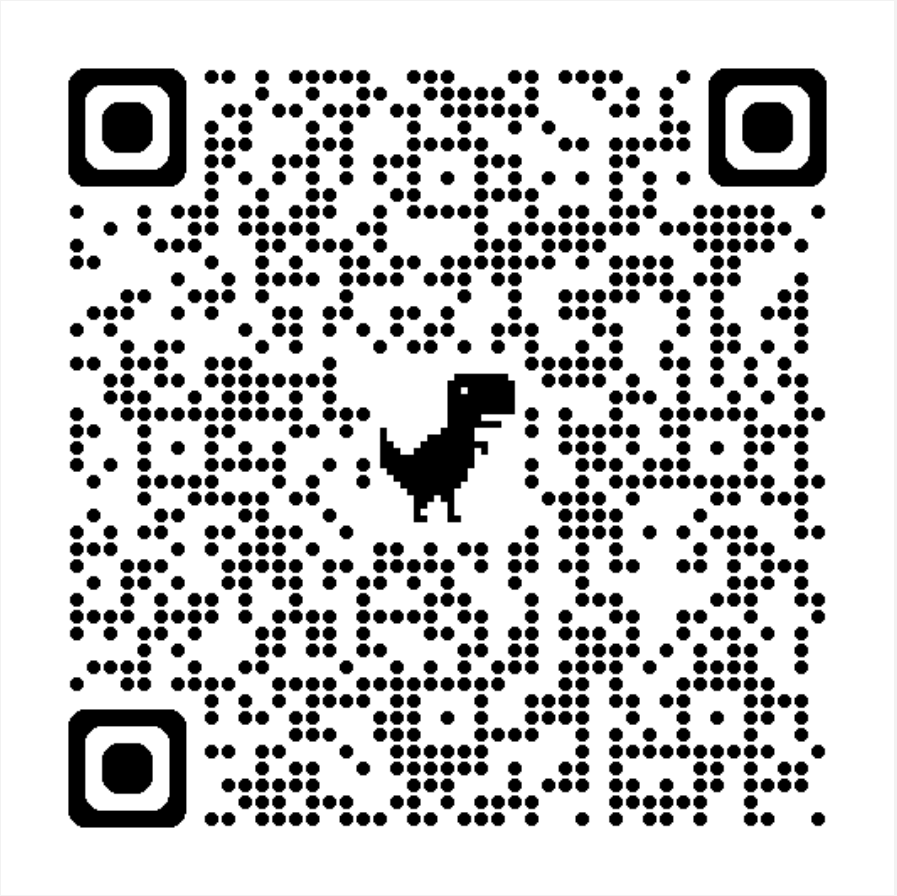
Thank you!

Ben McCulloch

Texas Department of Transportation

Strategic Data Scientist

benjamin.mcculloch@txdot.gov



UDOT

Keeping Utah Moving

From Governed Data to Responsible AI

Jennifer Volkening

Our Vision: A UDOT Where Reliable Data Empowers Every Decision

UDOT stakeholders will have access to reliable and accurate data to enable and support data-driven decision-making, performance management, legal and regulatory compliance requirements, while achieving transparency.

Core Principles

1. Treat Data as an Organizational Asset. Data will be formally defined, modeled, and governed to minimize redundancy and maximize its value as a strategic resource.

2. Define Authority and Accountability. We will establish clear authority for making data-related decisions through the Data Governance Steering Committee, ensuring accountability at all levels.

3. Proactively Manage Data Quality. Data quality will be defined, measured, and integrated into business processes to ensure it is fit for its intended purpose.

4. Integrate Governance into Operations and Culture. Governance will become part of our daily work, not a separate project, to ensure long-term sustainability.

5. Manage Based on Risk and Value. The extent of data management will be defined based on risk, value, and return on investment, focusing our efforts where they matter most.

6. Ensure Security and Privacy by Design. Safeguards and access controls will be embedded in our systems and processes to protect sensitive data and comply with all regulations.

GOALS

Protecting the integrity, privacy and security of the business's data

Implementing Data Management software that ensures clean, consistent data

Establishing standardized, repeatable processes for data entries and reporting

Developing a culture that relies on clean, consistent data when making decisions

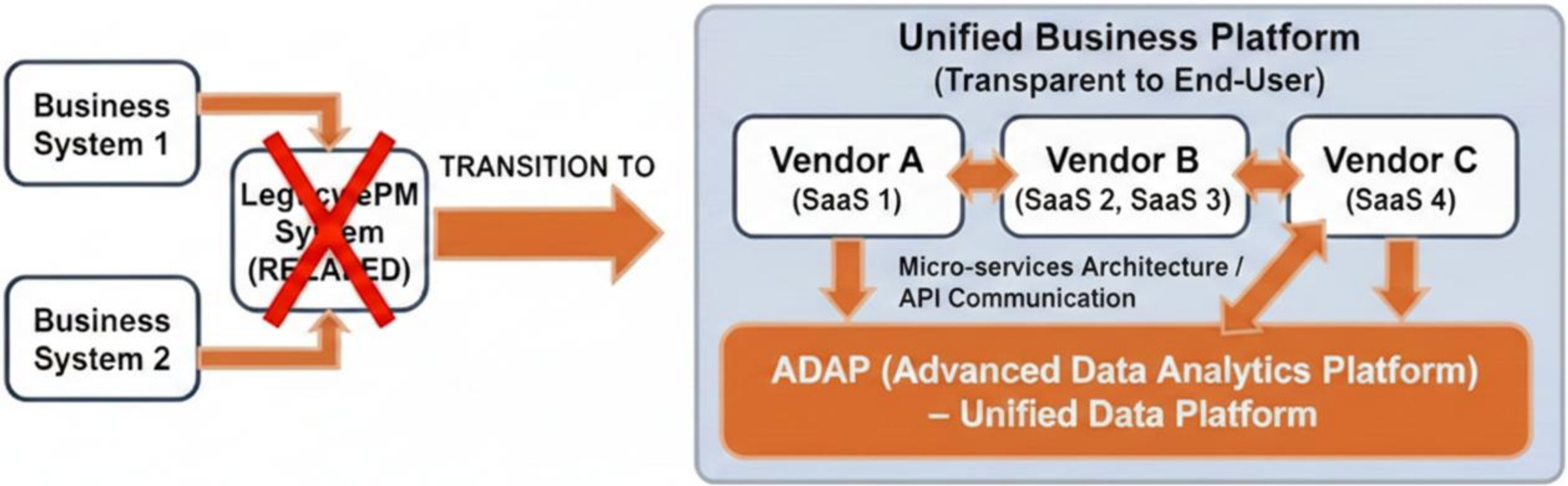
OBJECTIVES

Develop and maintain a data dictionary

Develop and maintain a business documentation process

Develop and maintain error/audit reports

Modernizing to a Unified Business and Data Platform



Data Governance Maturity Assessment

THE GOAL

To implement a comprehensive data governance framework that makes data management a core part of our identity.

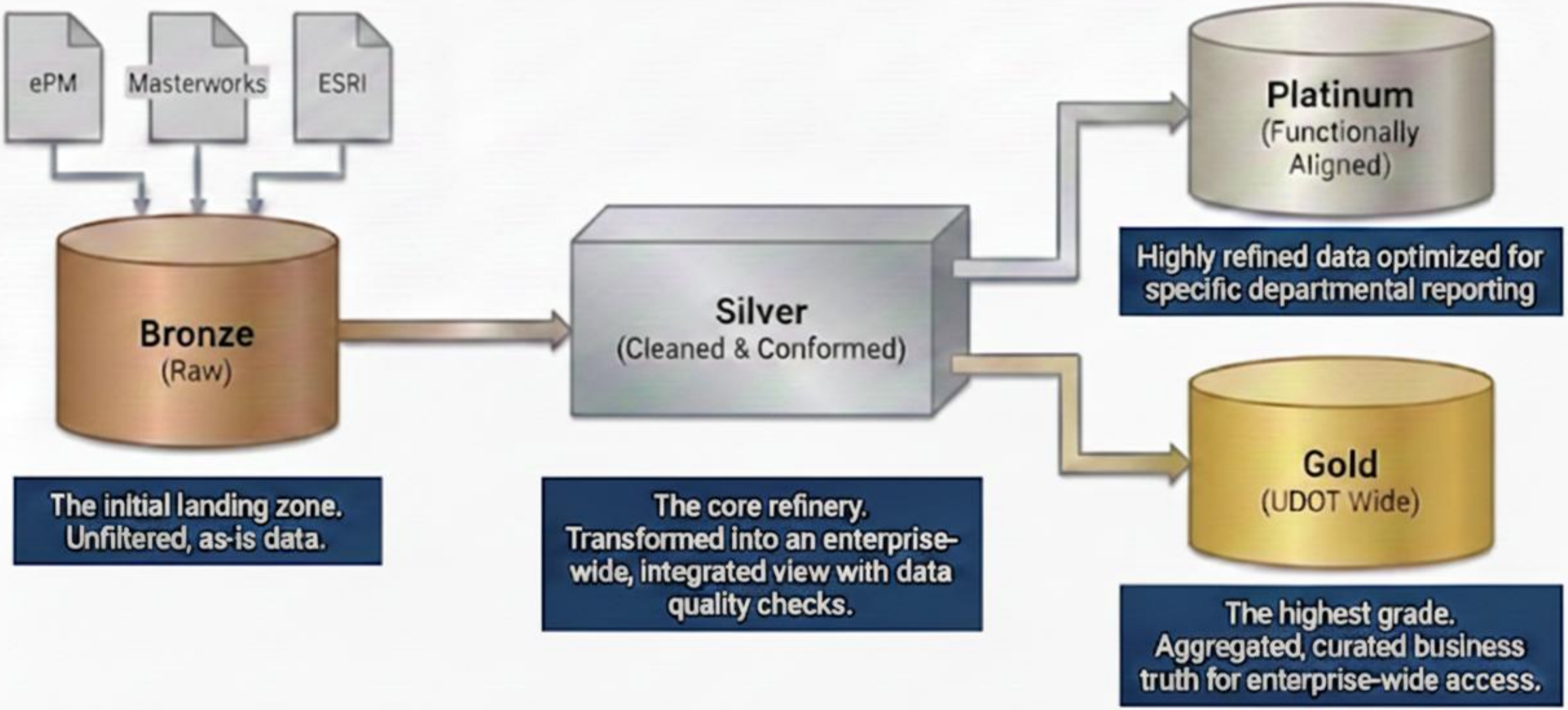
WHAT LEVEL 4/5 LOOKS LIKE

Data processes are integrated into operations with systematic risk management, regular monitoring, and comprehensive training. Governance is embedded in our organizational culture with advanced practices and plans for continuous improvement.

HOW WE GET THERE

Reaching this goal will require focusing on data maturity, training. data management, privacy and security, transparency, and ongoing change management through stakeholder training and communications.

UDOT - Medallion Data Architecture



ADAP Medallion Architecture and Standards Overview – Progressive Refinement

Goal: Evaluate layer criteria, determine rules and naming conventions, and identify data quality checks for the Silver layer.
Architecture: Data is progressively refined through four layers, each serving a distinct purpose.

Bronze Layer
(Raw)

- **Purpose:** Initial landing zone for raw data from source systems. Data is ingested "as-is" with minimal changes.
- **Dataset:** bronze_zone
- **Table Naming:** Keep source table names largely intect, often with a source system prefix (e.g., src_TableName or epm_epm_pins).
- **Data Formatting:** Stored in original format. Minimal cleaning (e.g., handling system-specific nulls, basic type casting). Schema should largely mirror the source.

Silver Layer
(Cleaned & Conformed)

- **Purpose:** Transform Bronze data into an enterprise-wide, conformed, and integrated view. Serves as a reliable source for multiple downstream uses.
- **Dataset:** silver_zone
- **Table Naming:** Standardize names to reflect business entities (e.g., mw_project_info, epm_project_info). Names are in lower snake_case.
- **Data Formatting:** Cleaned (handle missing values, correct errors, remove deplicates), Conformed (standardize units, codes, and terminologies), integrated (loin tables from different sources).
- **Data Quality Checks:** Implemented to conform values (check missing values, data types, outliers, test values).

Platinum Layer
(Functionally Aligned)

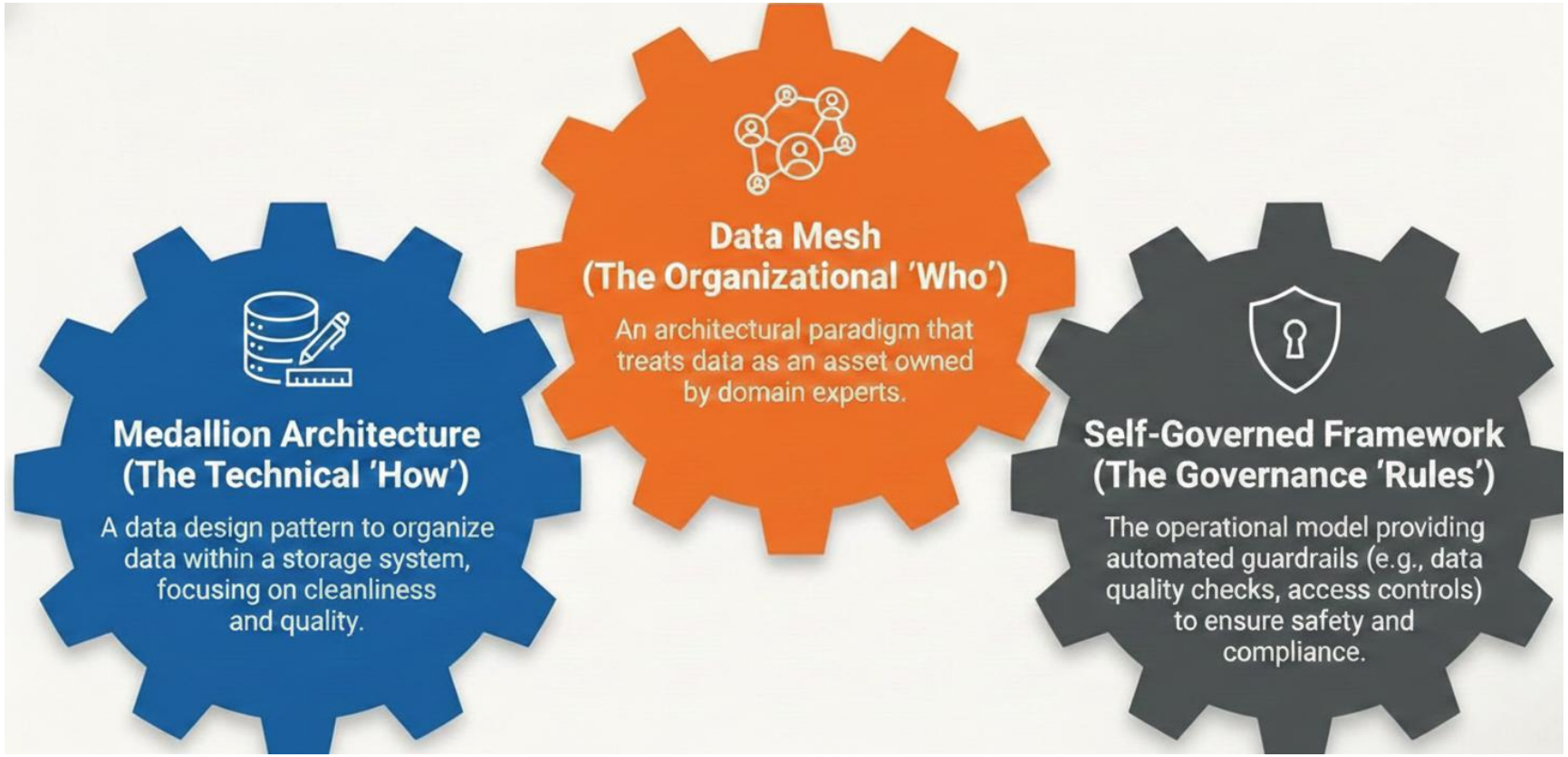
- **Purpose:** Highly refined, aggregated, and often denormalized data optimized for a specific functional purpose (e.g., reporting, dashboard, system ingestion).
- **Dataset:** platinum_zone
- **Table Naming:** Highly descriptive and infuitive for and-users; named according to the report, dashboard, or analytical use case (e.g., MonthlySalesReport_Agg).
- **Data Formatting:** Aggregated and Denormalized for performance. Project-specific and tailored to the needs of the intended audience.

Gold Layer
(UDOT Wide)

- **Purpose:** Highly refined, aggregated, and often denormalized data meant to represent the business for UDOT-wide access with potential for public availability.
- **Dataset:** gold_zone
- **Table Naming:** Highly descriptive and intuitive; named according to the department, divisitor, or sector of the business (WIP).
- **Data Formatting:** Aggregated and Denormalized for performance. Tailored to the needs of a particular project or business domain.
- **Field Names/Descriptions:** Field names can be renamed for end-users (lower snake_case). Descriptions should contain relevant transformations and formulas.

Balanced Framework

- Data Architecture
- Ownership
- Governance



The Perfect Implementation: Data Mesh gives ownership, Medallion Architecture refines data, and a Self-Governed Framework that ensures compliance and security.

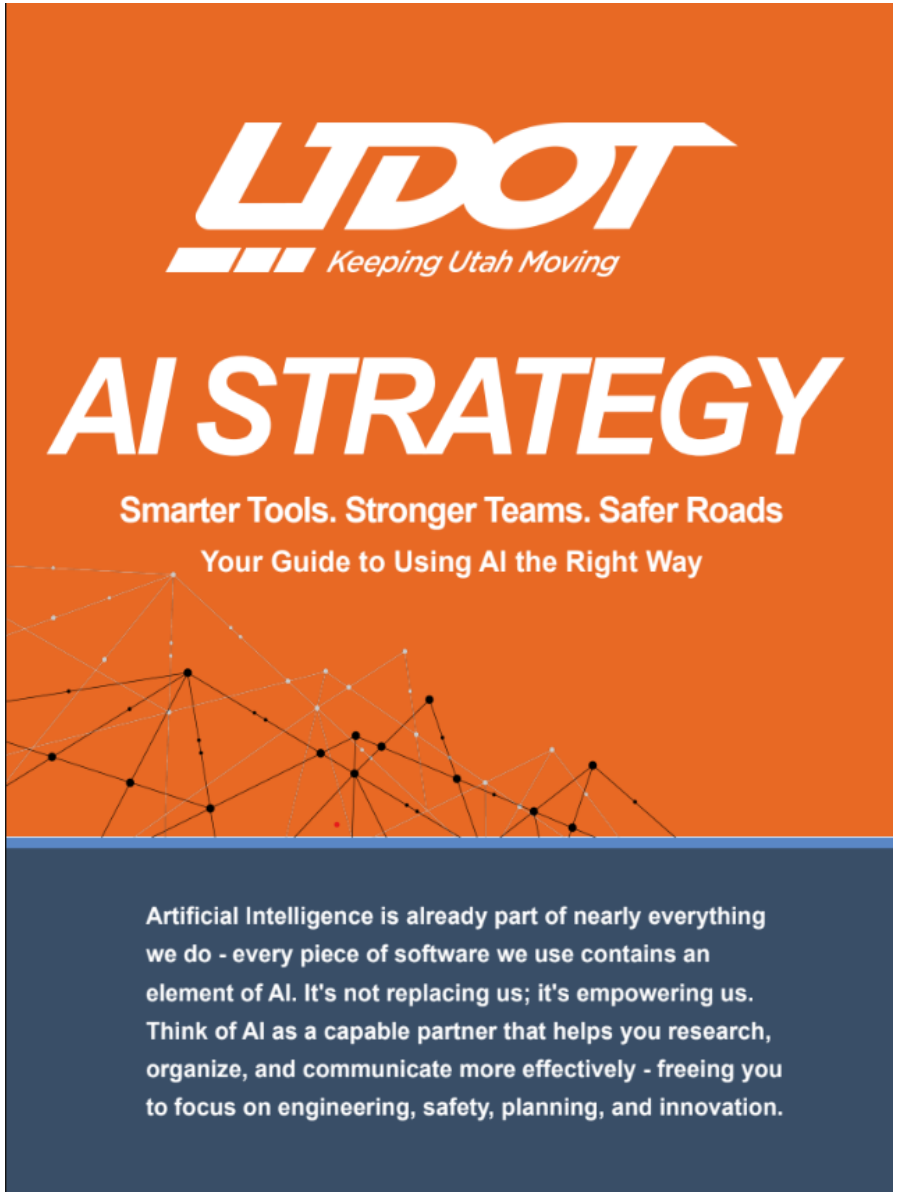
Accelerating Change | Continuous Improvement | Organizational Development

Guiding Principles

- Trust as the Cornerstone
- Privacy and Security
- Balance between AI and Human Judgement
- Consider Risk
- Accuracy and Reliability
- Transparancy and Accountability
- Constructive Use

Governance & Responsible Use

Responsible AI use centers on transparency, data protection, fairness, and human oversight to ensure safe, ethical, and trustworthy outcomes



Today's Presenters

Jesse Newberry
HNTB
jnewberry@hntb.com

Joel Jundt
South Dakota Department of Transportation
joel.jundt@state.sd.us

Jennifer Volkening
Utah Department of Transportation
jvolkening@utah.gov

Gregory Ciparelli
Connecticut Department of Transportation
gregory.ciparelli@ct.gov

Benmjain McCulloch
Texas Department of Transportation
benjamin.mcculloch@txdot.gov

NATIONAL ACADEMIES Sciences Engineering Medicine
TRB TRANSPORTATION RESEARCH BOARD

ACRP Insight Event
Exploring the Impact of AI on the Airport Industry

May 19 – 20, 2026

Washington, D.C.

NATIONAL ACADEMIES Sciences Engineering Medicine

Upcoming events for you

July 12-15, 2026

65th Annual Workshop on Transportation Law

https://www.nationalacademies.org/events/1089

https://www.nationalacademies.org/trb/events

NATIONAL ACADEMIES Sciences Engineering Medicine

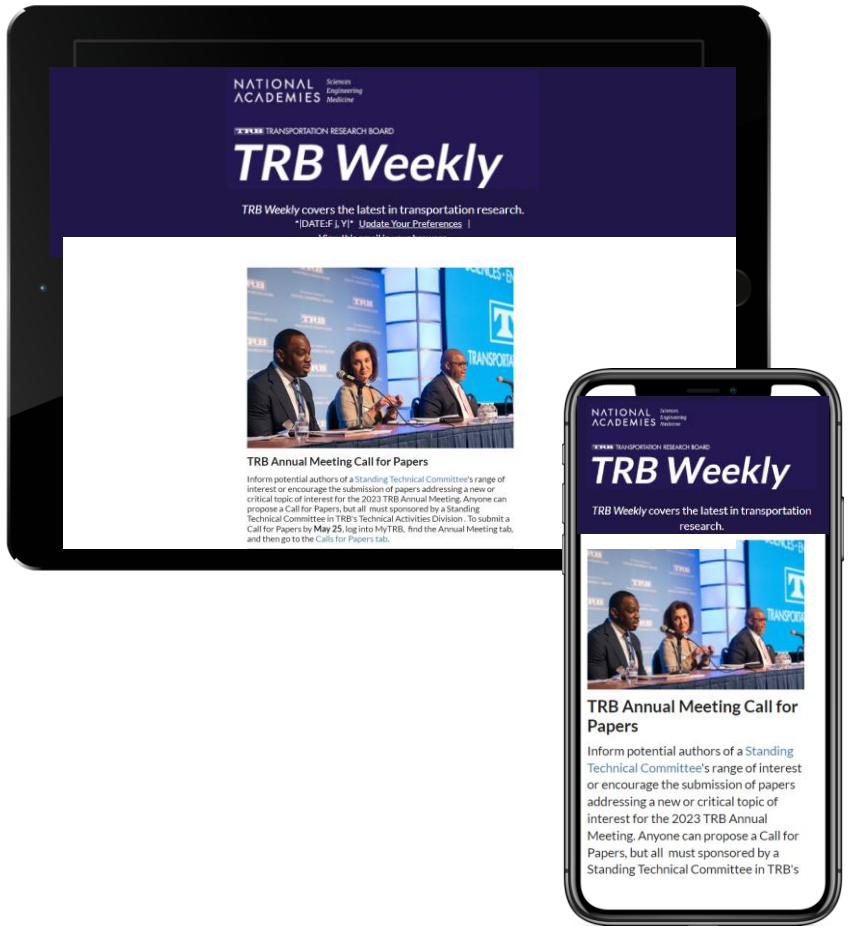
TRB TRANSPORTATION RESEARCH BOARD

Subscribe to *TRB Weekly*

If your agency, university, or organization perform transportation research, you and your colleagues need the *TRB Weekly* newsletter in your inboxes!

Each Tuesday, we announce the latest:

- RFPs
- TRB's many industry-focused webinars and events
- 3-5 new TRB reports each week
- Top research across the industry



Spread the word and subscribe!
https://bit.ly/ResubscribeTRBWeekly

Discover new TRB Webinars weekly

Set your preferred topics to get the latest listed webinars and those coming up soon every Wednesday, curated especially for you!

https://mailchi.mp/nas.edu/trbwebinars

And follow #TRBwebinar on social media

NATIONAL ACADEMIES Sciences Engineering Medicine

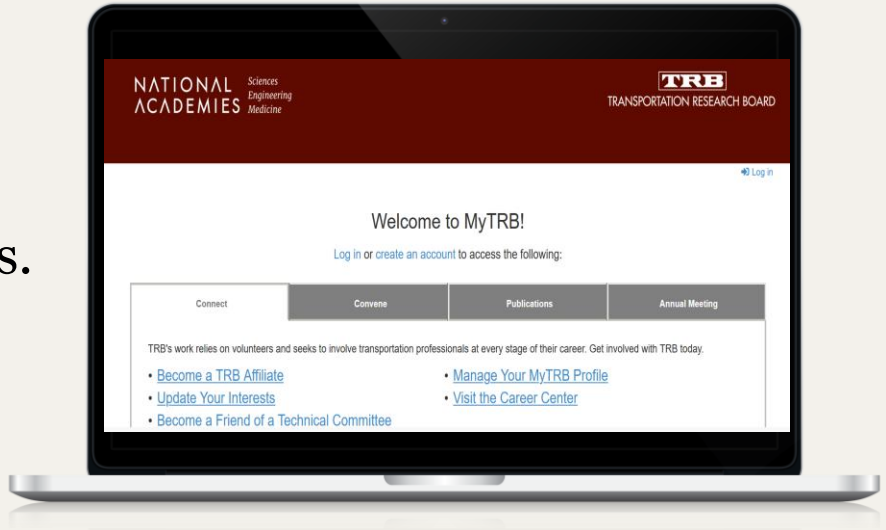
TRB TRANSPORTATION RESEARCH BOARD

Get involved

TRB mobilizes expertise, experience, and knowledge to anticipate and solve complex transportation-related challenges.

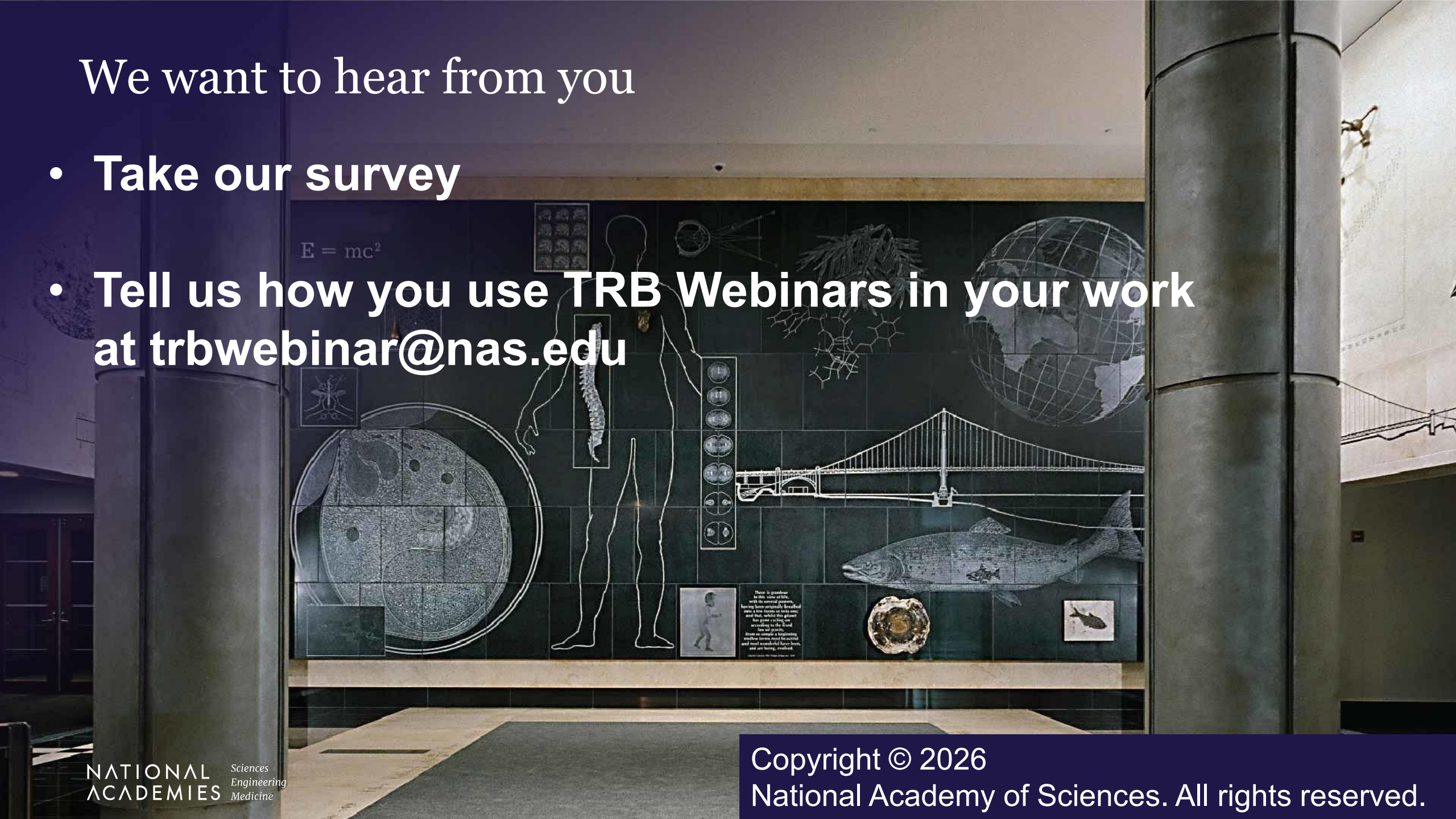
TRB's mission is accomplished through the hard work and dedication of **thousands of volunteers**.

https://www.nationalacademies.org/trb/get-involved



We want to hear from you

- **Take our survey**
- **Tell us how you use TRB Webinars in your work at trbwebinar@nas.edu**



NATIONAL ACADEMIES Sciences Engineering Medicine