UNITED STATES
DEPARTMENT OF TRANSPORTATION

# Real-Time Data Capture and Management Program Research Data Exchange Concept of Operations Review

DCM/DMA Summer Webinar Series

Gene McHale

August 10, 2011

# Overview of Webinar

- Purpose
  - Provide an update on DCM program status,  an introduction to Research Data Exchange Concept
- Agenda
  - Introduction to the DCM Program
  - Research Data Exchange Concept
  - DCM Program Next Steps
  - Discussion

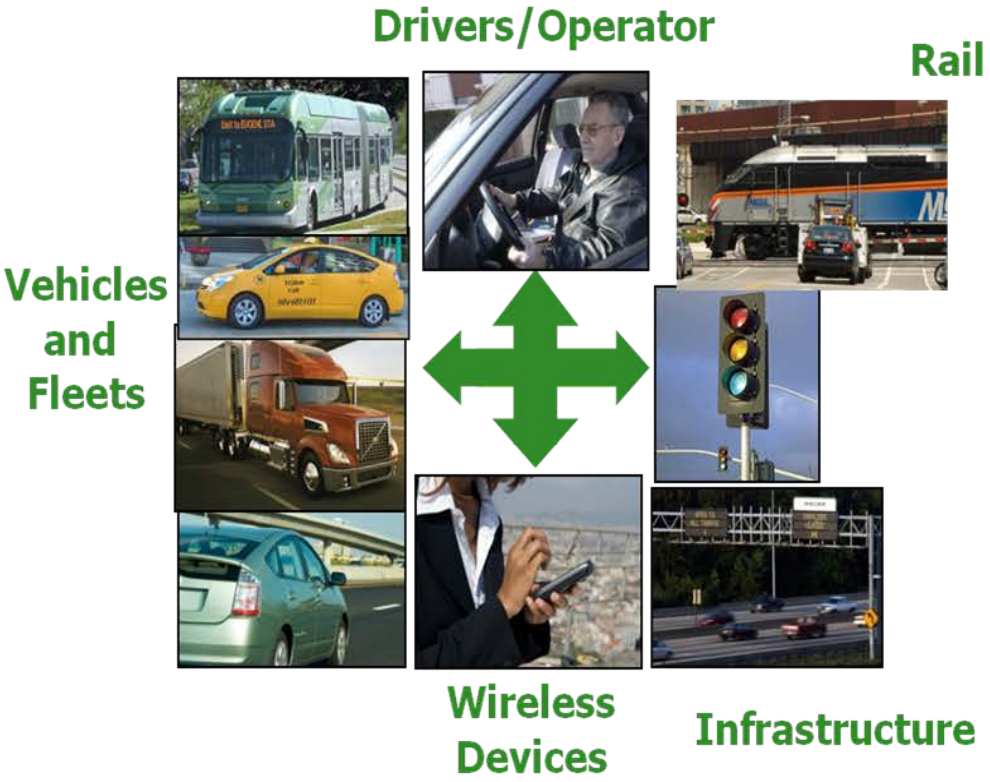
# ITS Research = Multimodal and Connected

**To Improve Safety, Mobility and Environment**

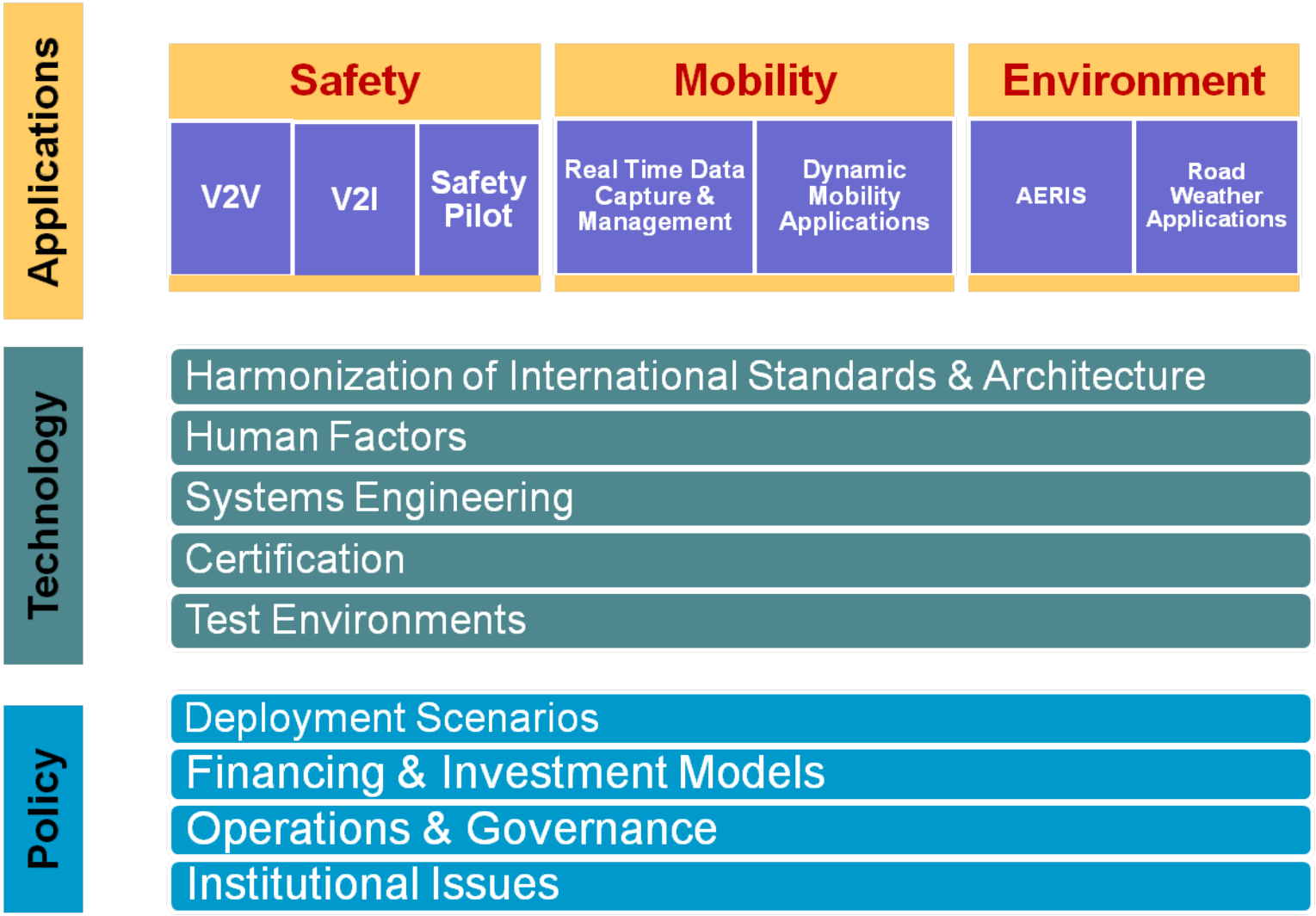
**Research of technologies and applications that use wireless communications to provide connectivity:**

- Among vehicles of all types
- Between vehicles and roadway infrastructure
- Among vehicles, infrastructure and wireless consumer devices

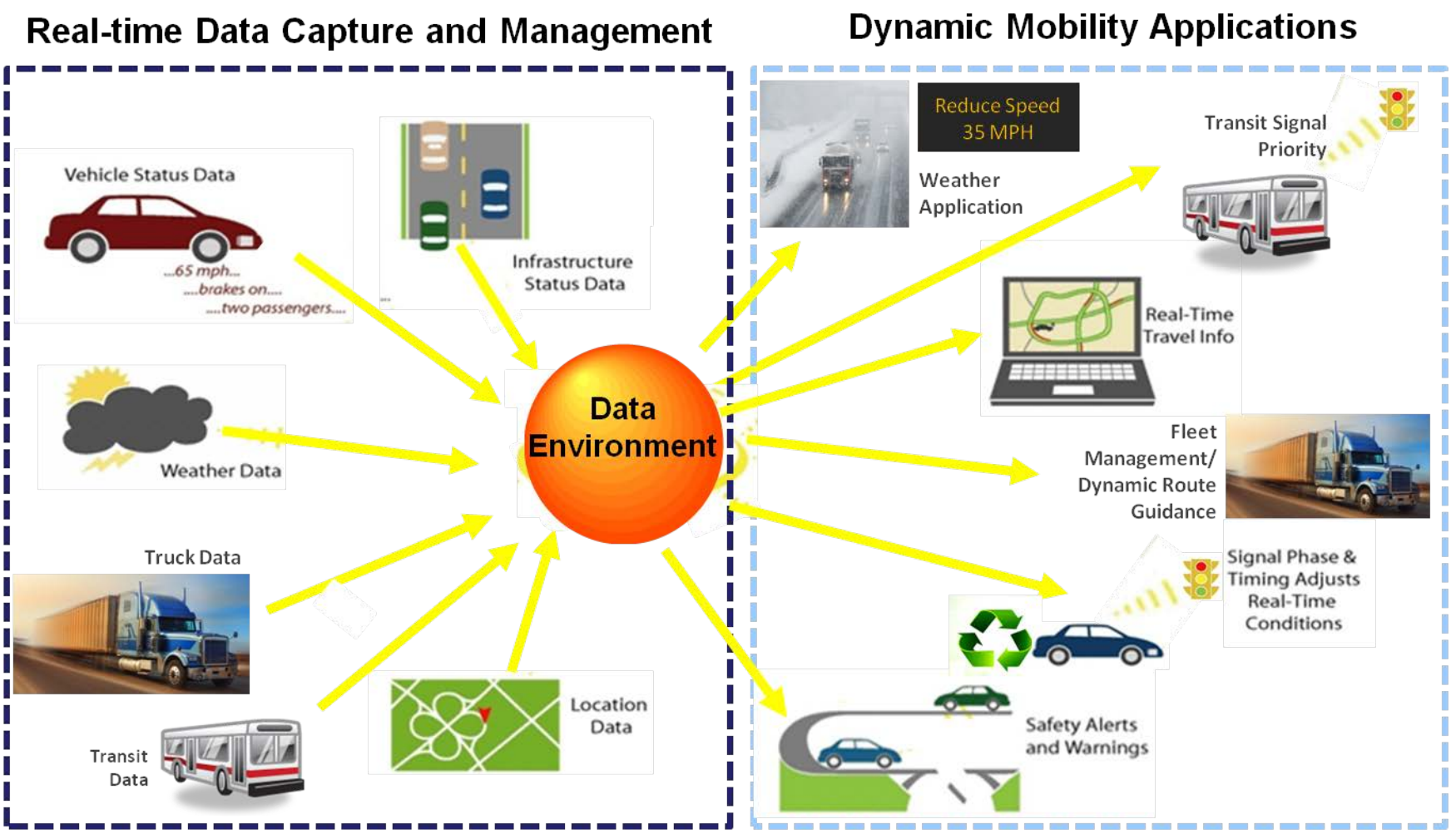
FCC Allocated Spectrum at 5.9 GHz for Transportation Safety (known as DSRC)



# ITS Research Program Components



# Mobility Program



# Data Capture and Management Program: Vision and Program Objectives

**Vision**

- Active acquisition and systematic provision of integrated, multi-source data to enhance current operational practices and transform future surface transportation systems management

**Objectives**

- Enable systematic data capture from connected vehicles (automobiles, transit, trucks), mobile devices, and infrastructure
- Develop data environments that enable integration of data from multiple sources for use in transportation management and performance measurement
- Reduce costs of data management and eliminate technical and institutional barriers to the capture, management, and sharing of data
- Determine required infrastructure for transformative applications implementation, along with associated costs and benefits

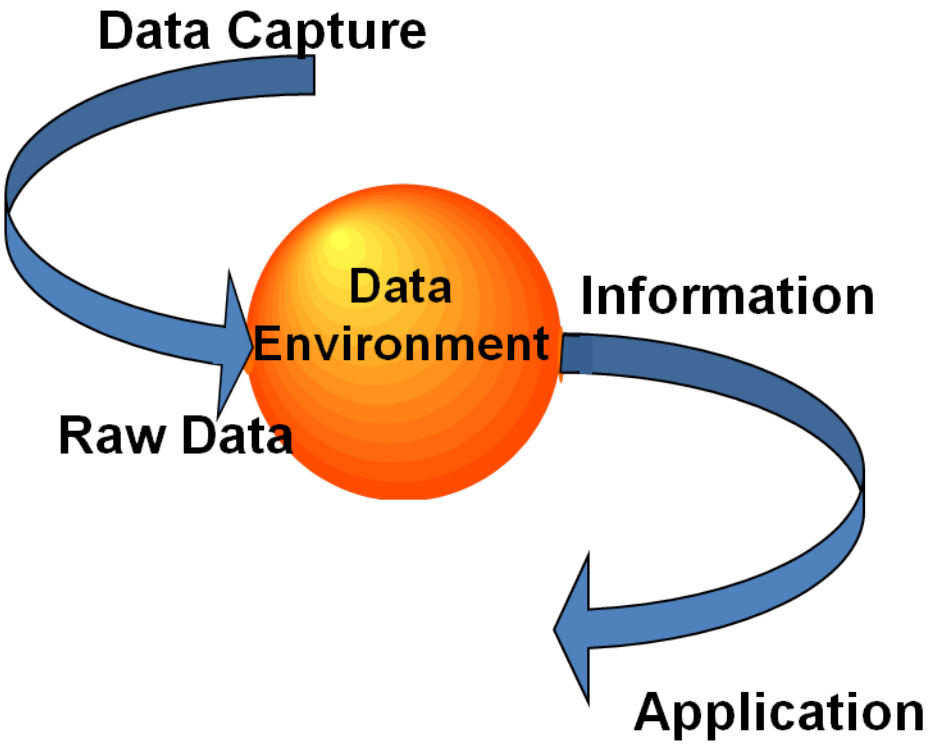
**Program Partners**

- ITS JPO, FTA, FHWA R&D, FHWA Office of Operations
  BTS, FMCSA

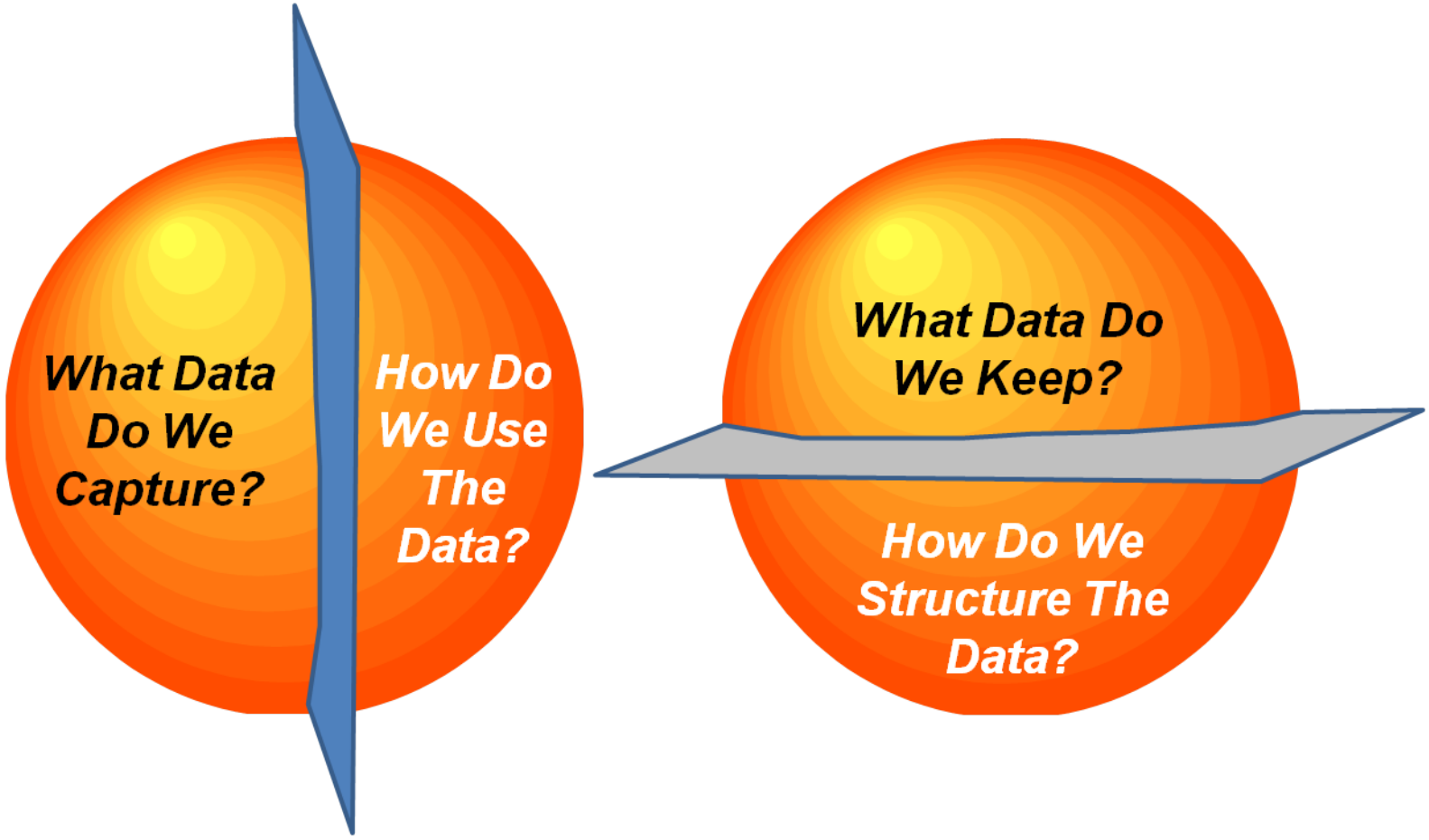
# Data Environments

**Data environment**:

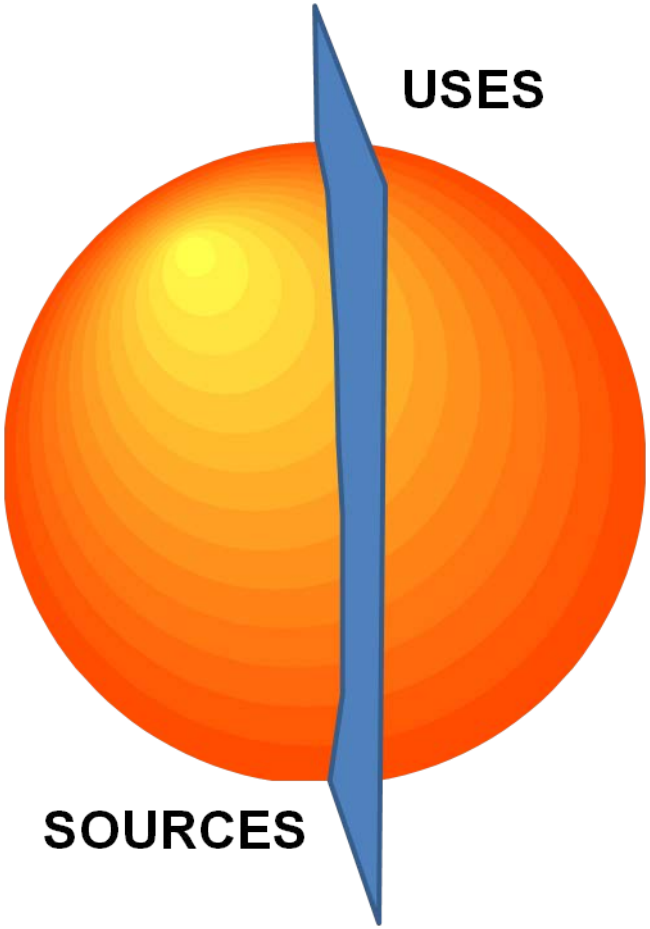
- well-organized collection of data of specific type and quality
- captured and stored at regular intervals from one or more sources
- systematically shared in support of one or more applications



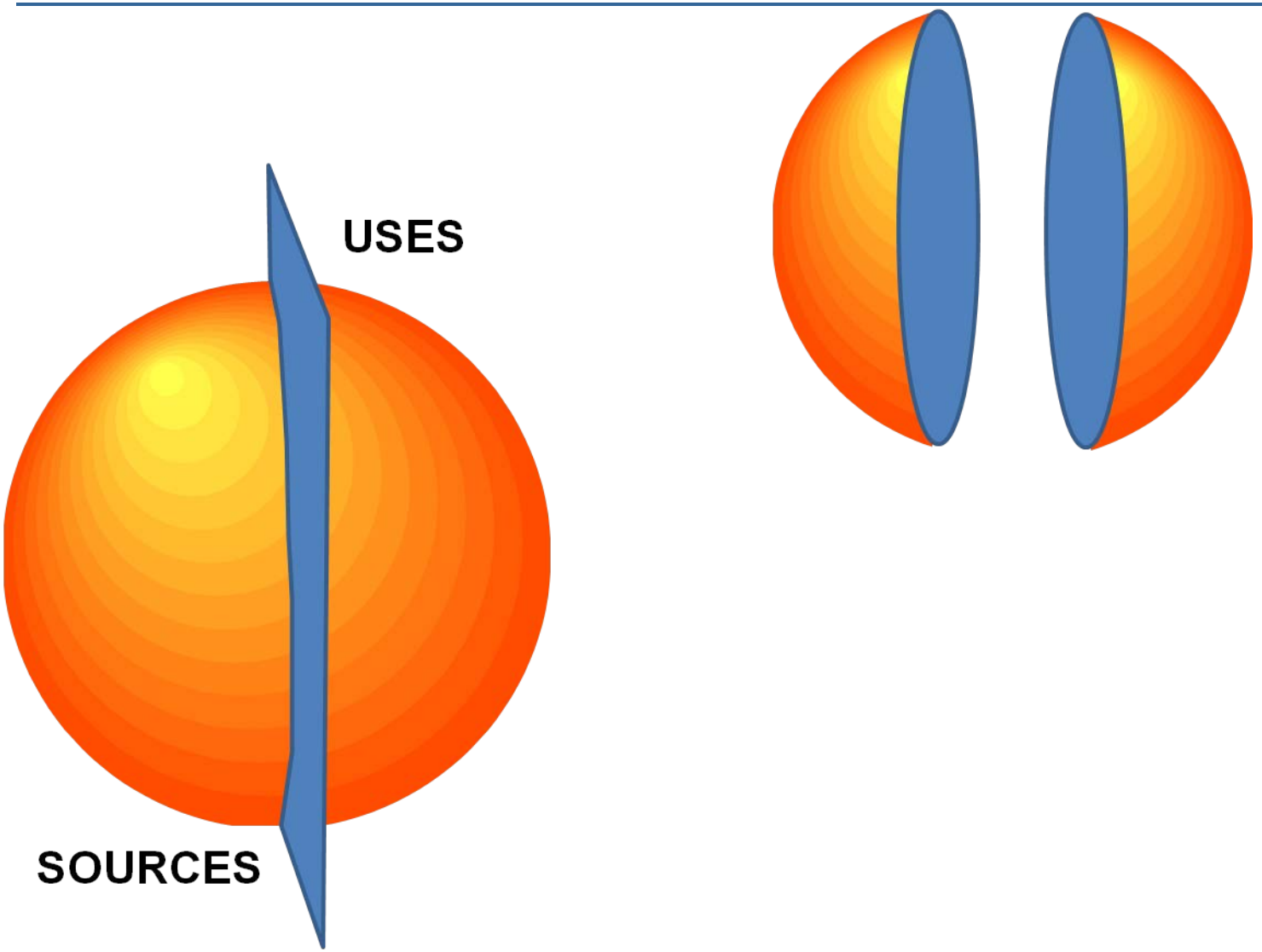
# Key Issues in Defining A Data Environment



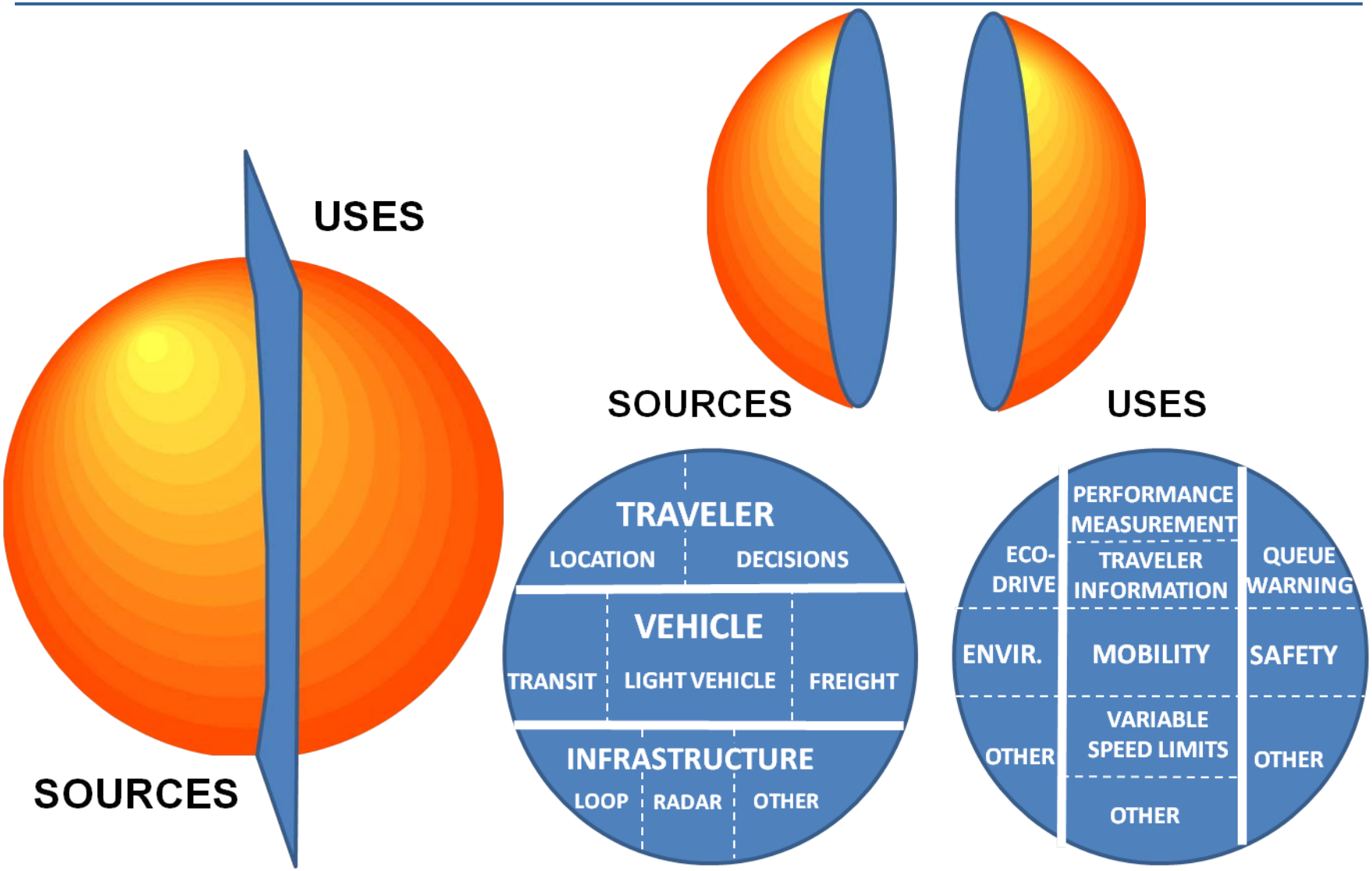
# Data Sources and Uses



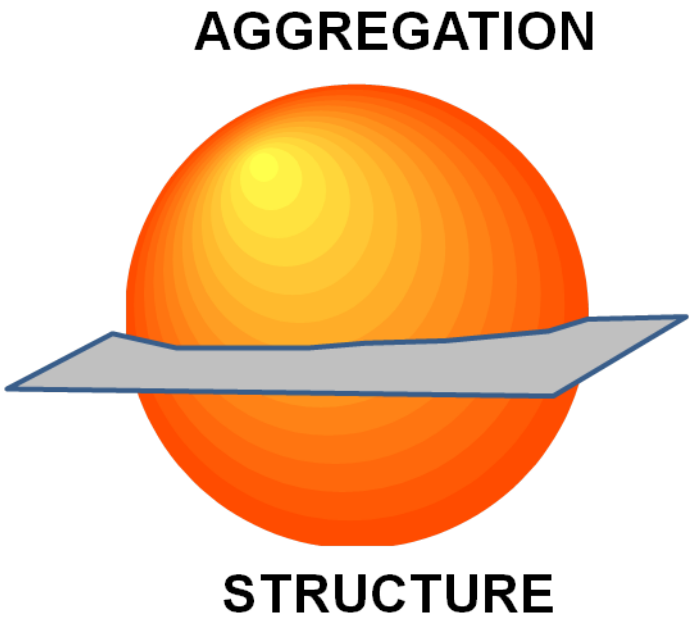
# Data Sources and Uses



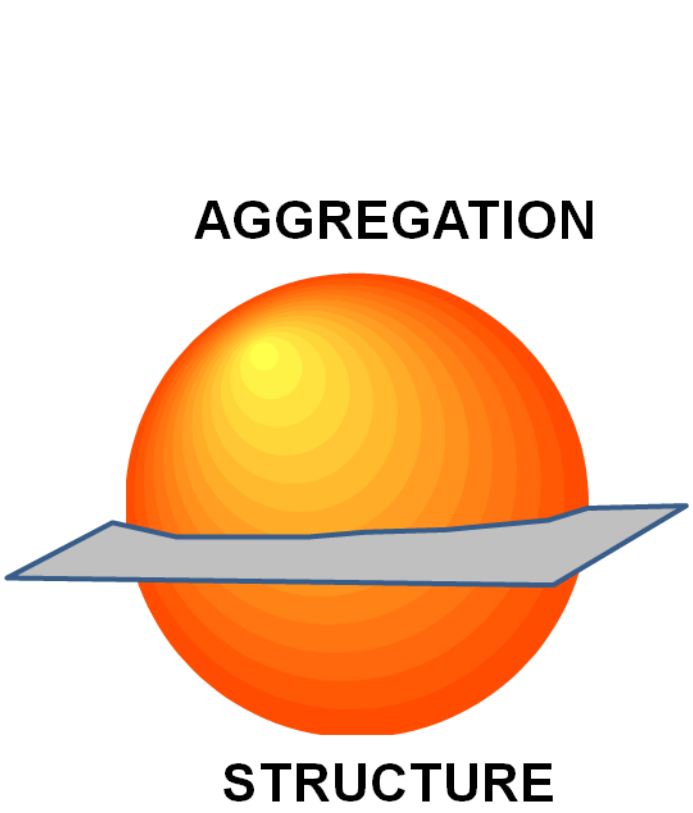
# Data Sources and Uses

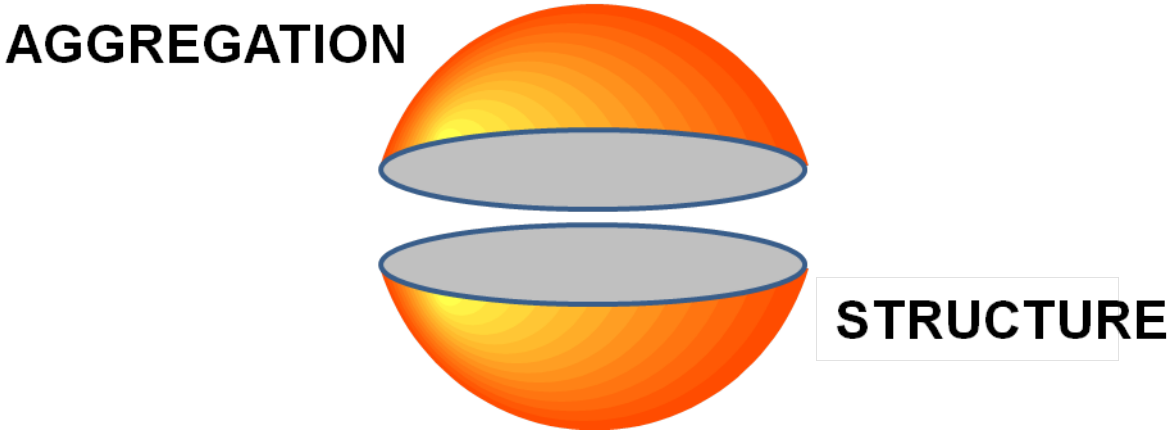


# Data Aggregation and Structure

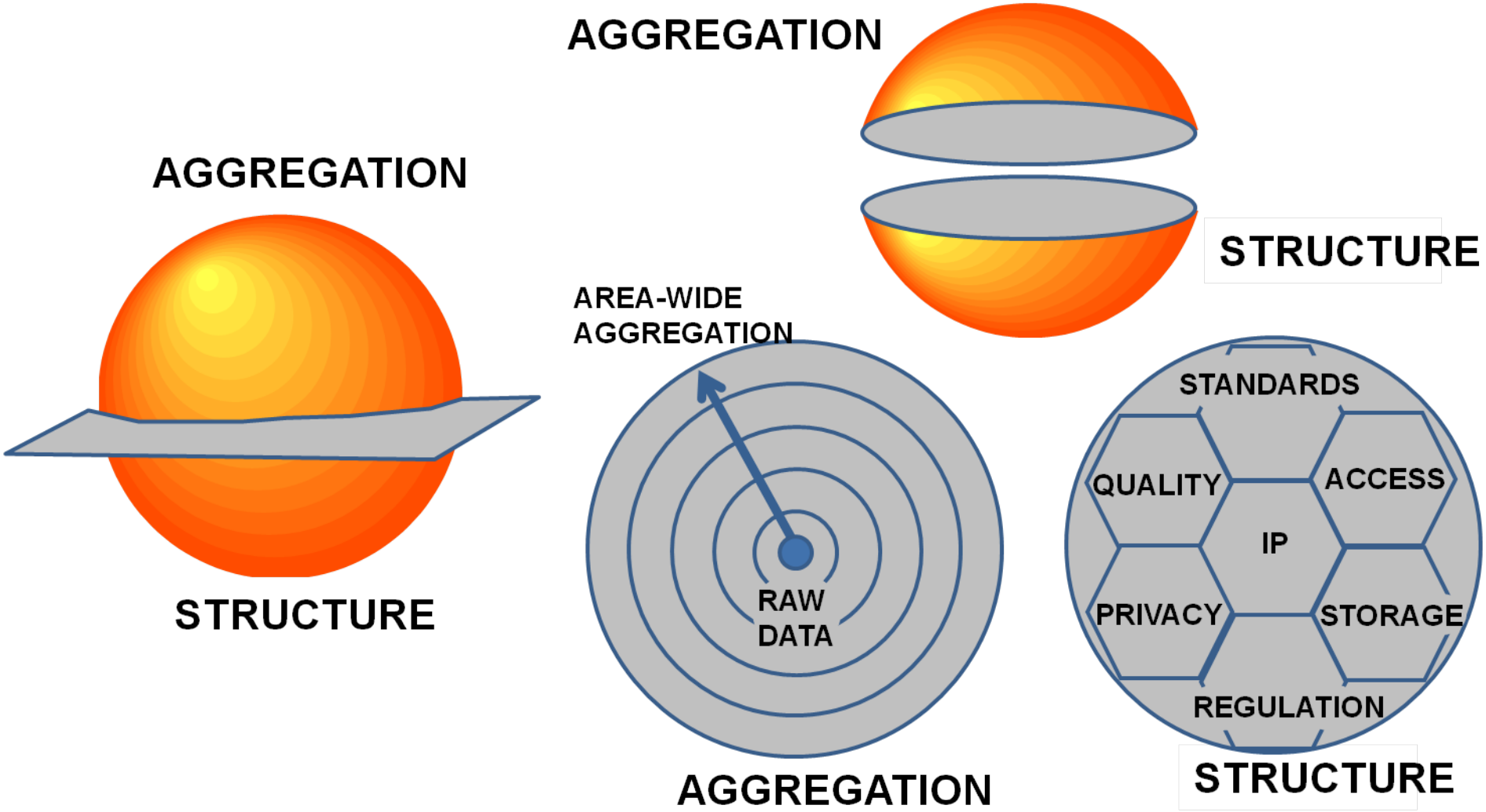


# Data Aggregation and Structure



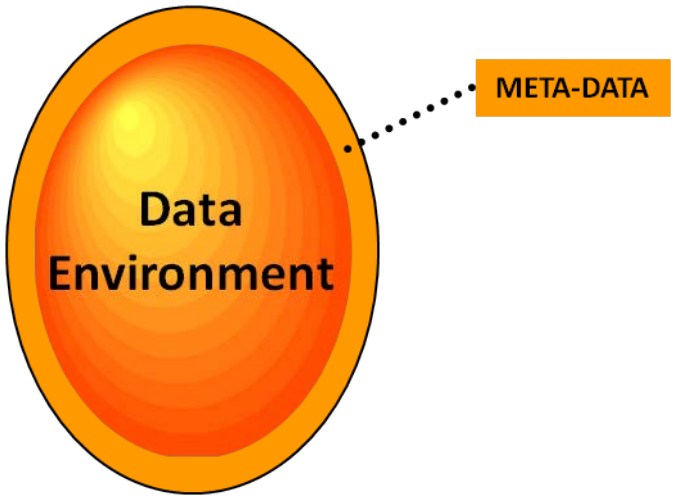
# Data Aggregation and Structure



# *Elements of Data Capture and Management*

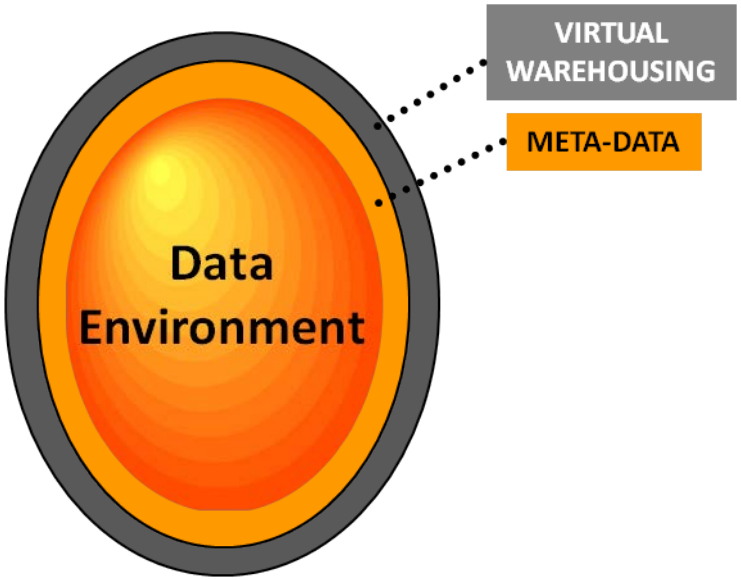
# *Elements of Data Capture and Management*

- **Meta data**:
  - Provision of well-documented data environment



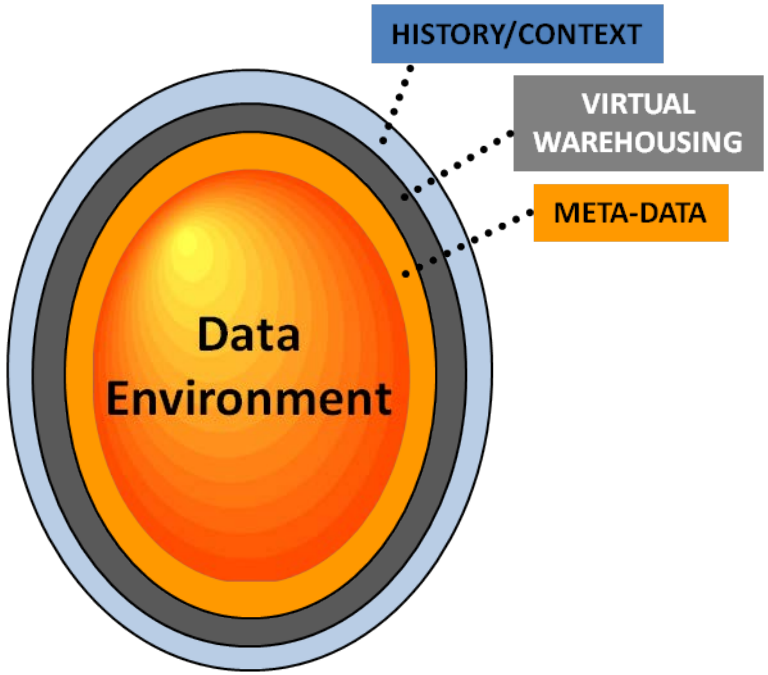
# *Elements of Data Capture and Management*

- **Meta data**:
  - Provision of well-documented data environment
- **Virtual warehousing**:
  - Supports access to data environment and forum for collaboration



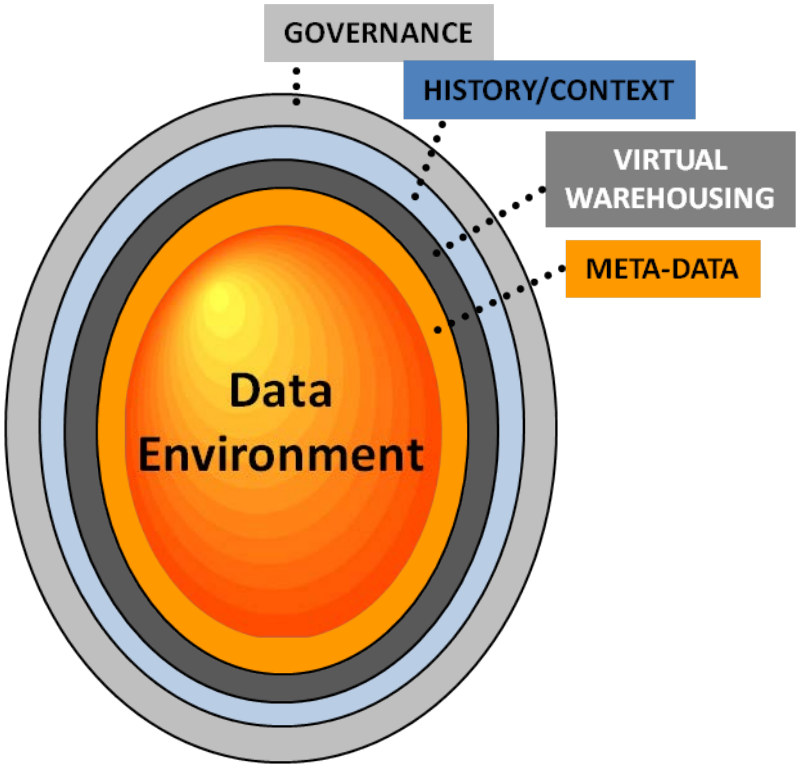
# *Elements of Data Capture and Management*

- **Meta data**:
  - Provision of well-documented data environment
- **Virtual warehousing**:
  - Supports access to data environment and forum for collaboration
- **History/context**:
  - Objectives of data assembly



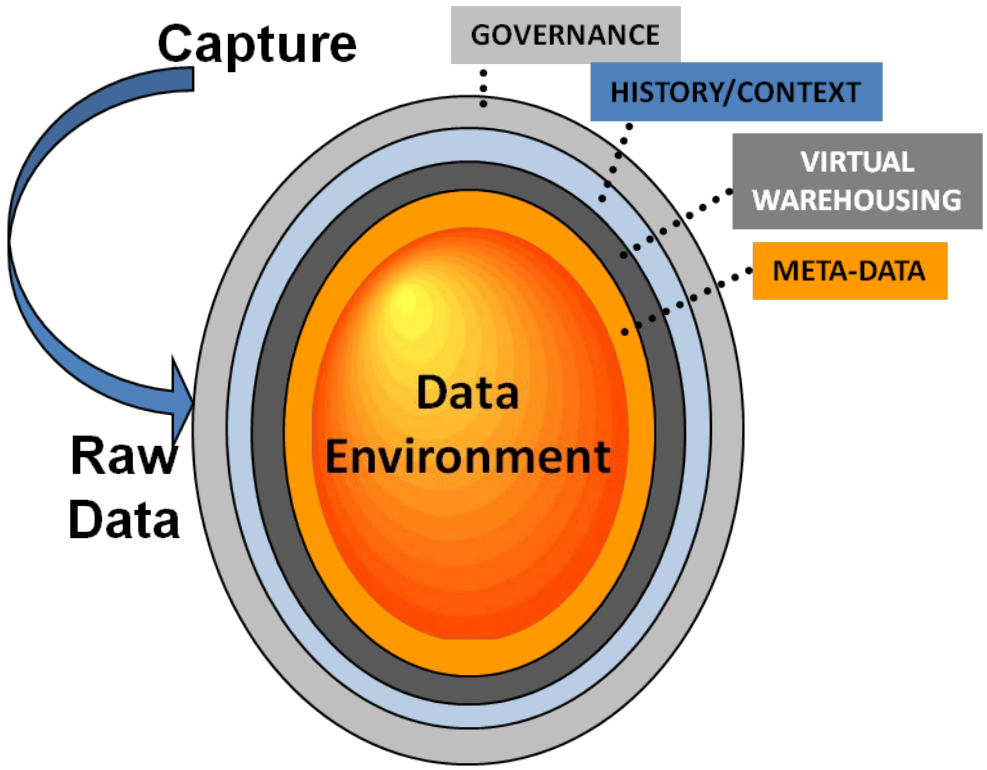
# *Elements of Data Capture and Management*

- **Meta data**:
  - Provision of well-documented data environment
- **Virtual warehousing**:
  - Supports access to data environment and forum for collaboration
- **History/context**:
  - Objectives of data assembly
- **Governance**:
  - Rules under which data environment can be accessed and procedures for resolving disputes



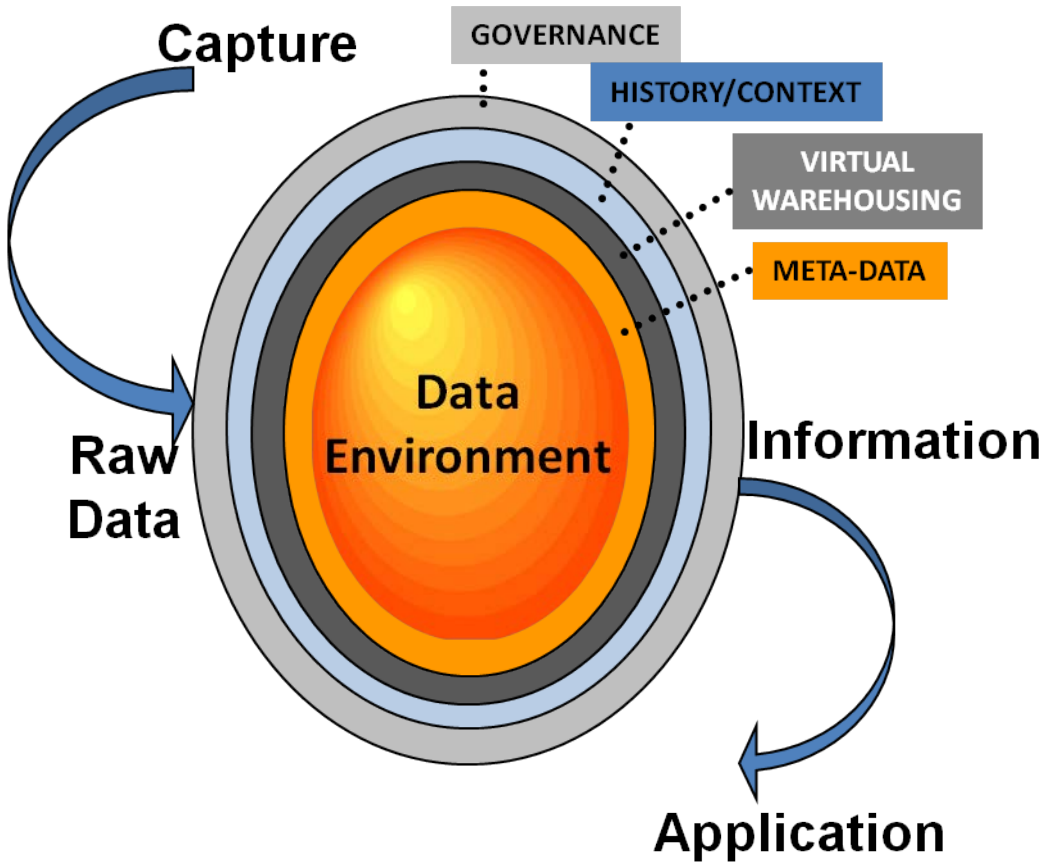
# *Elements of Data Capture and Management*

- **Meta data**:
  - Provision of well-documented data environment
- **Virtual warehousing**:
  - Supports access to data environment and forum for collaboration
- **History/context**:
  - Objectives of data assembly
- **Governance**:
  - Rules under which data environment can be accessed and procedures for resolving disputes

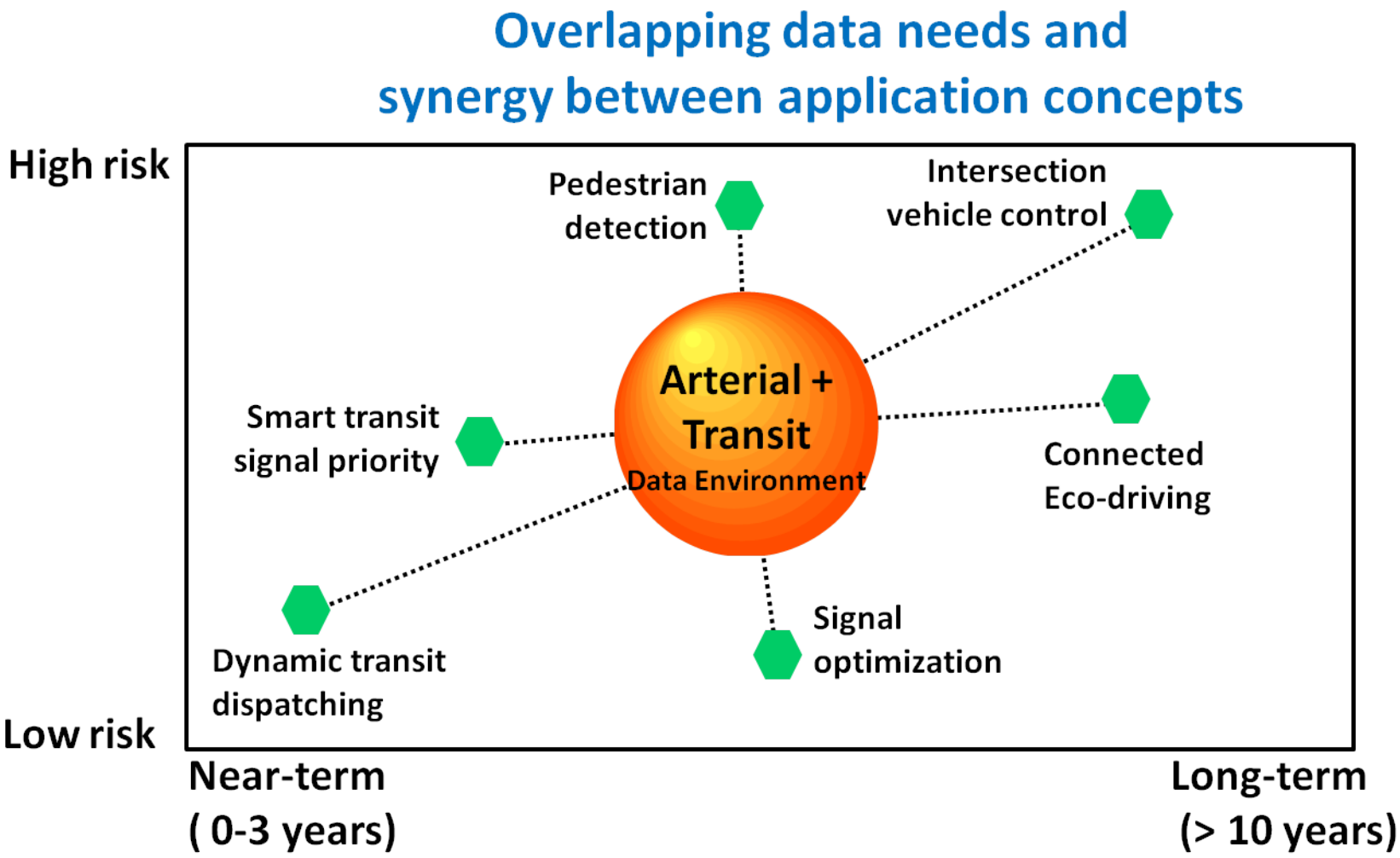


# *Elements of Data Capture and Management*

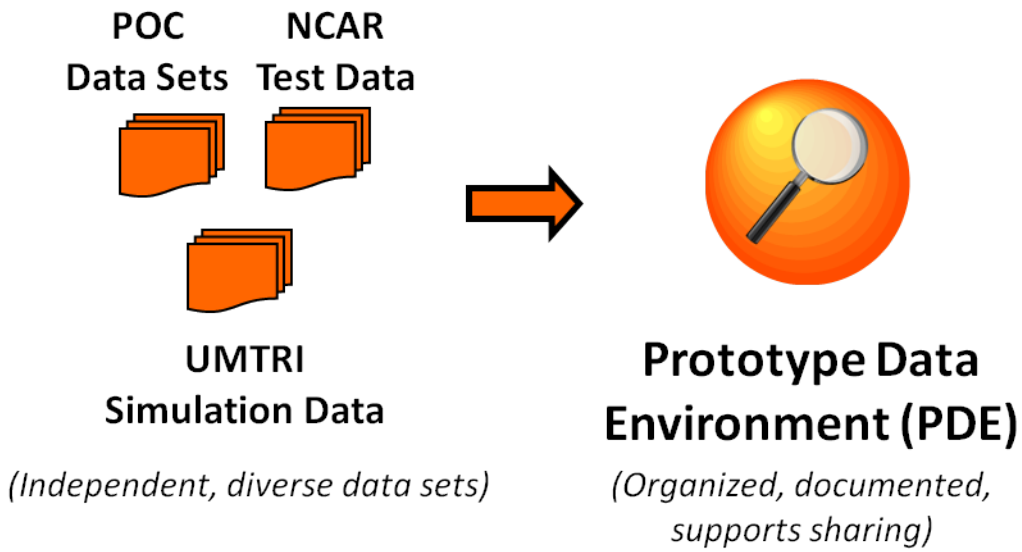
- **Meta data**:
  - Provision of well-documented data environment
- **Virtual warehousing**:
  - Supports access to data environment and forum for collaboration
- **History/context**:
  - Objectives of data assembly
- **Governance**:
  - Rules under which data environment can be accessed and procedures for resolving disputes



# Each Data Environment Supports Multiple Apps

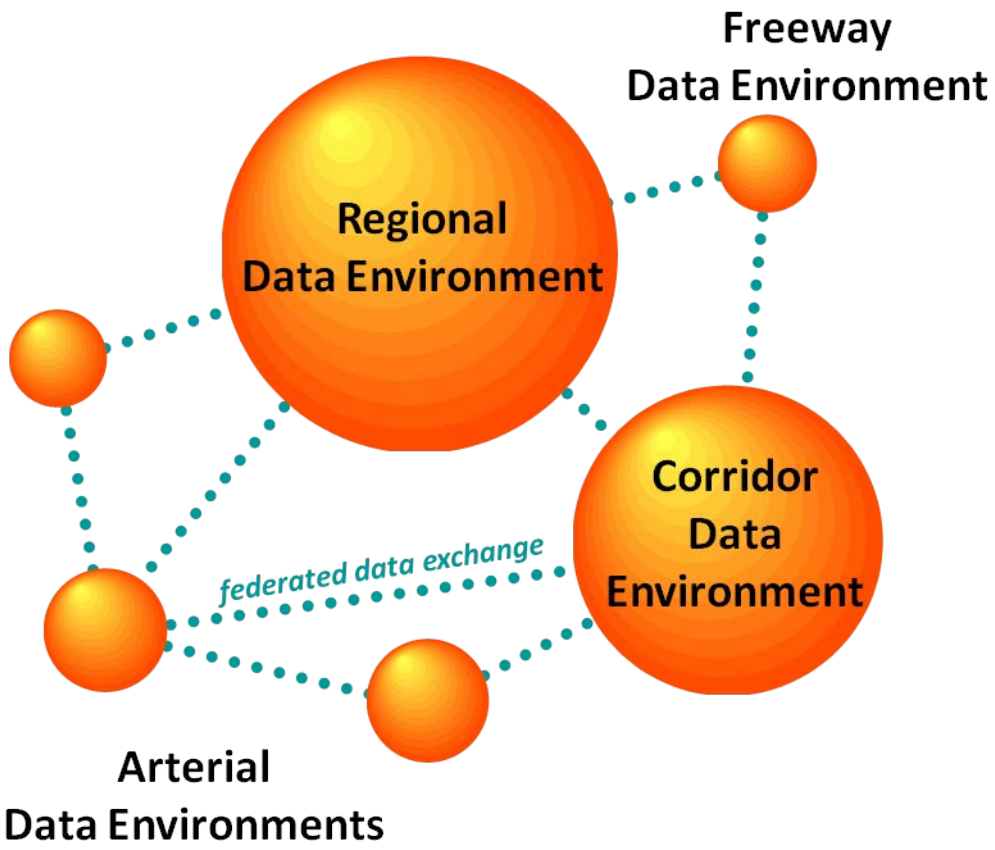


# The Need for a Research Data Exchange



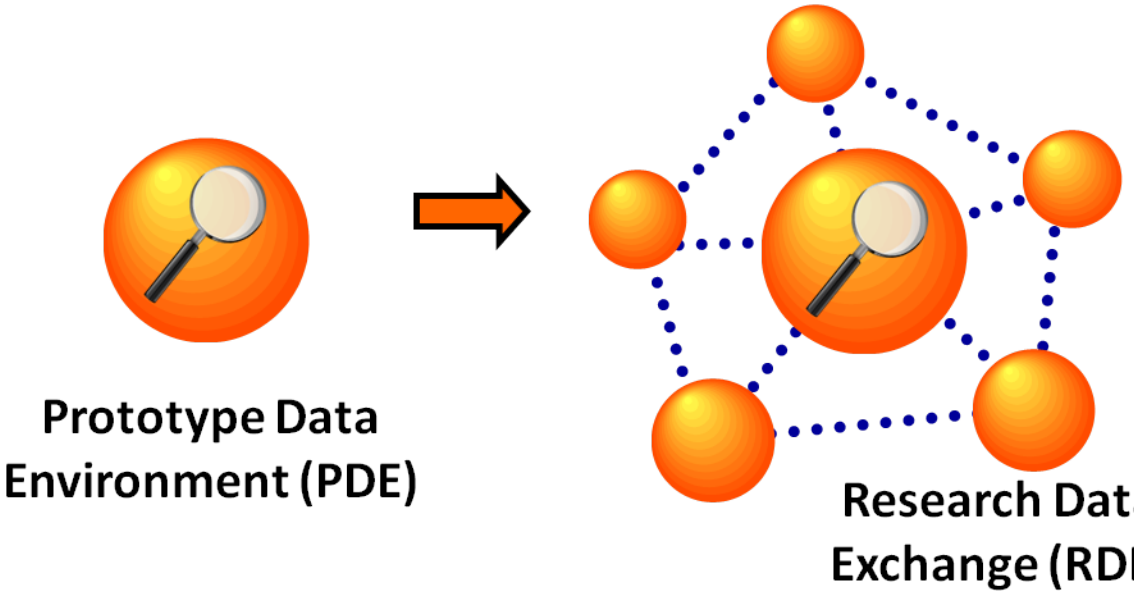

- Data Environment prototyping revealed **need for a complex system of multiple data environments to support application development**
  - Researchers interested in organizing data in different ways for different purposes – need to support multiple collections
  - Both archived and real-time data provision (static archive not sufficient)
  - Value of local control and documentation of data
  - Single mega-archive (*a.k.a., the Death Star*) has high technical risk
  - Potential role of data ***federation*** – virtual, decentralized collections

# Federated Data Environments



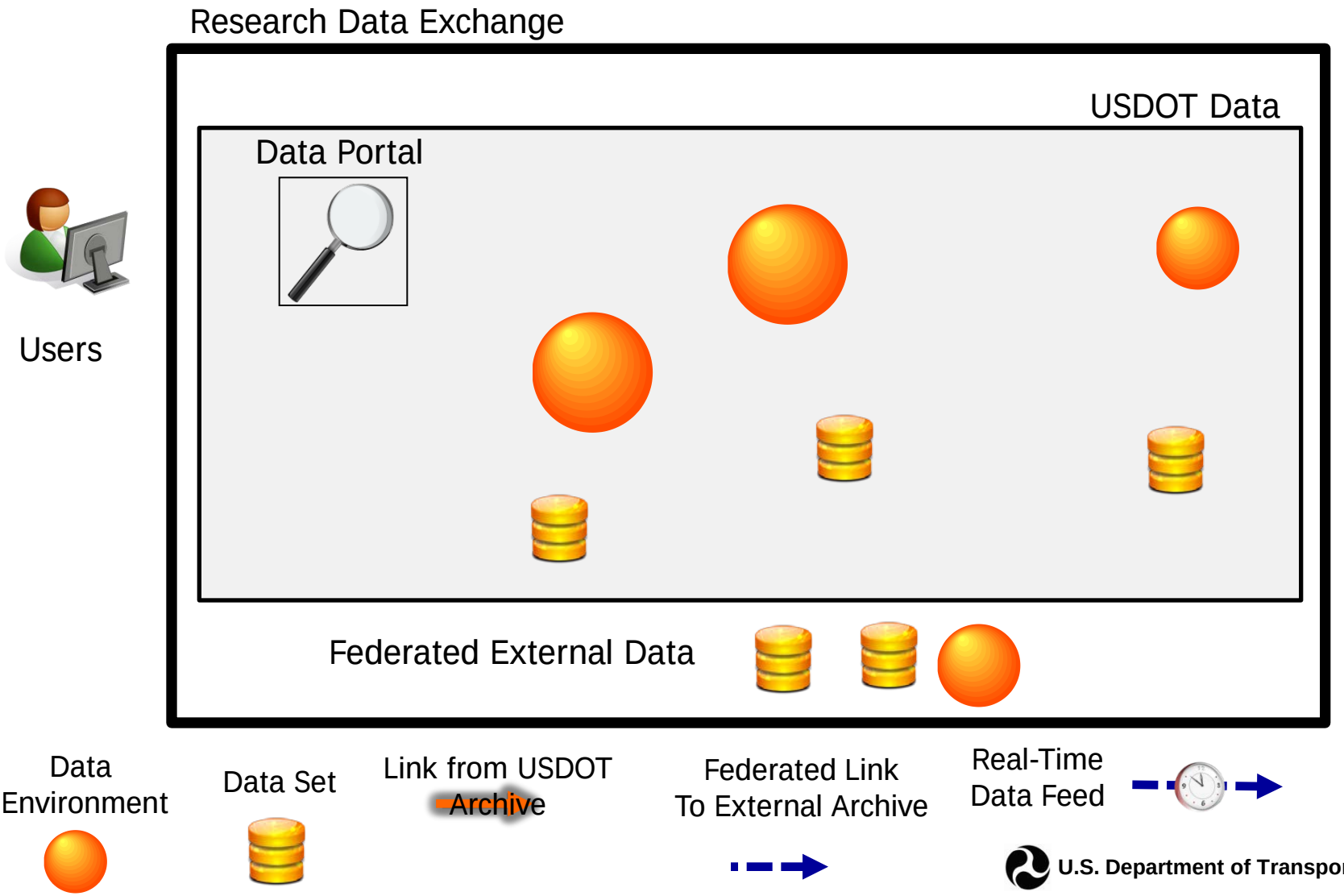

- **Federated Data Systems**
  - Decentralized
  - Virtual
  - Independent
  - Heterogeneous
  - Systematic data exchange among federated environments
- **Each data environment supports a specific level of system control/decision**
  - For example, geographic (figure)
  - Might also be functional or jurisdictional, other

# The Research Data Exchange Concept

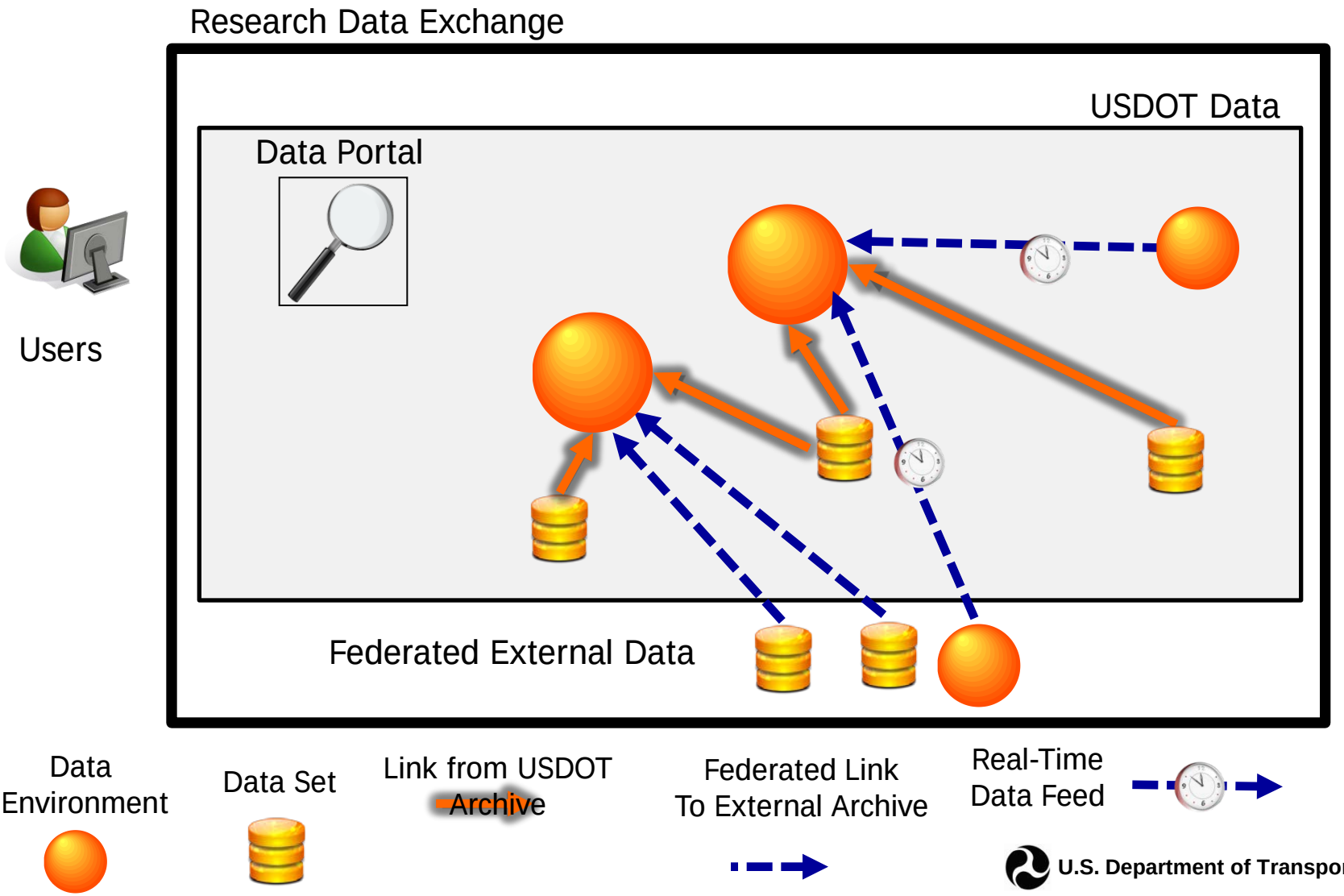



- The Research Data Exchange (RDE) is the connected system of data environments we envision to support application research and development
- The RDE will *not* be a single, centralized repository
  - but rather a *system of systems* linking multiple data management systems
  - some of which will be maintained and controlled outside of the USDOT, through a common web-based Data Portal
- Some data will be archived at USDOT within the RDE, other data will be archived outside of USDOT and federated with the RDE

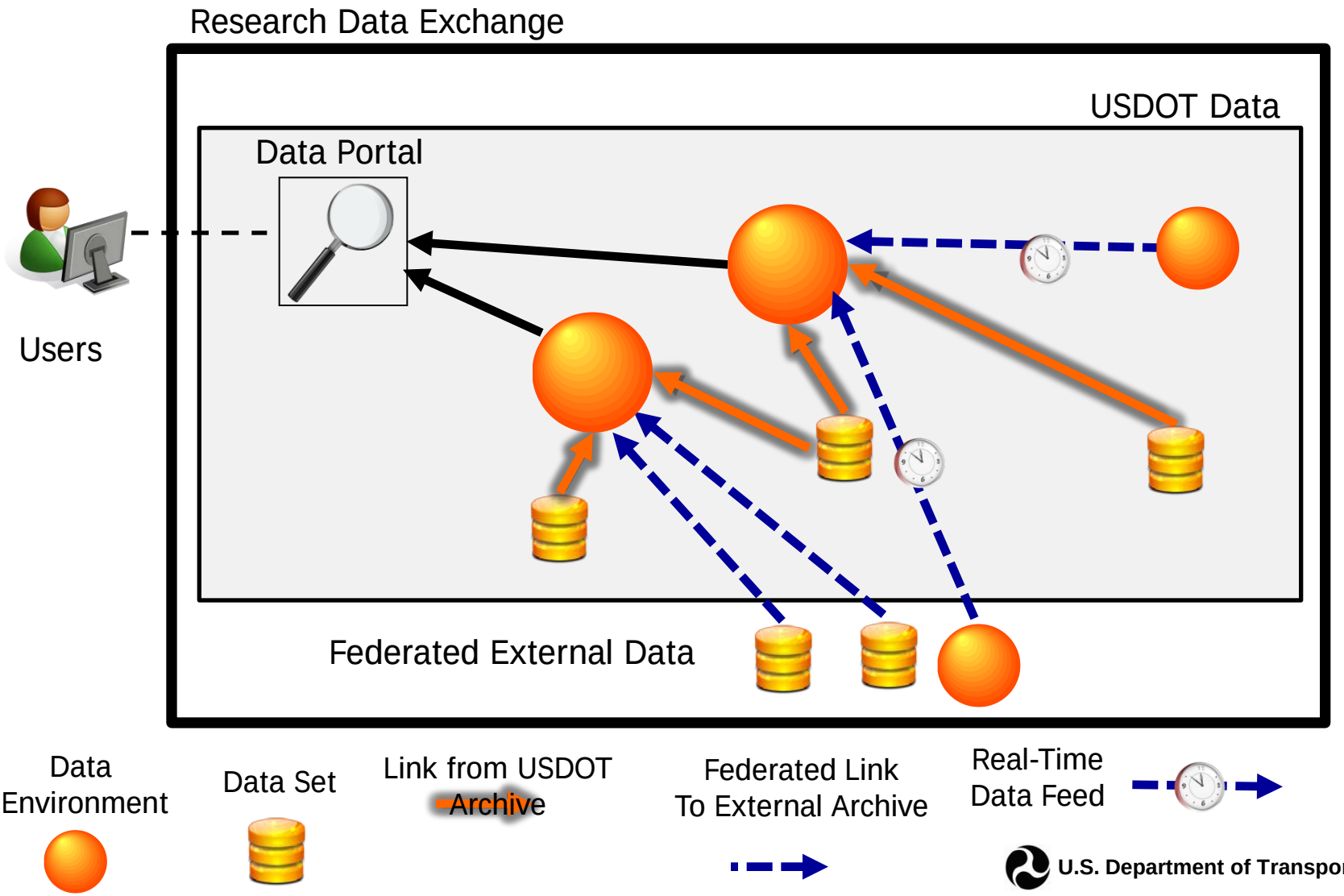
# Research Data Exchange



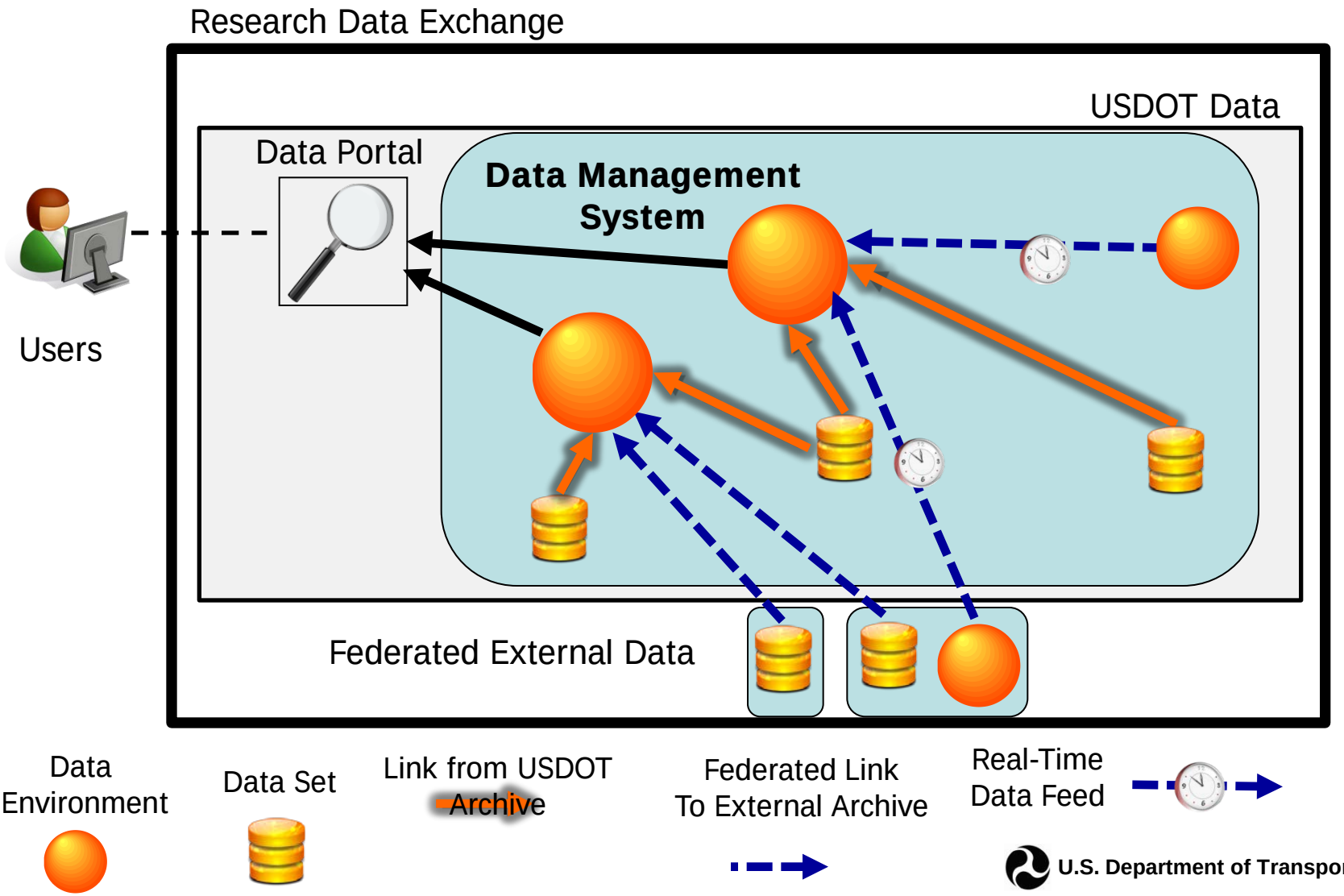
# Research Data Exchange



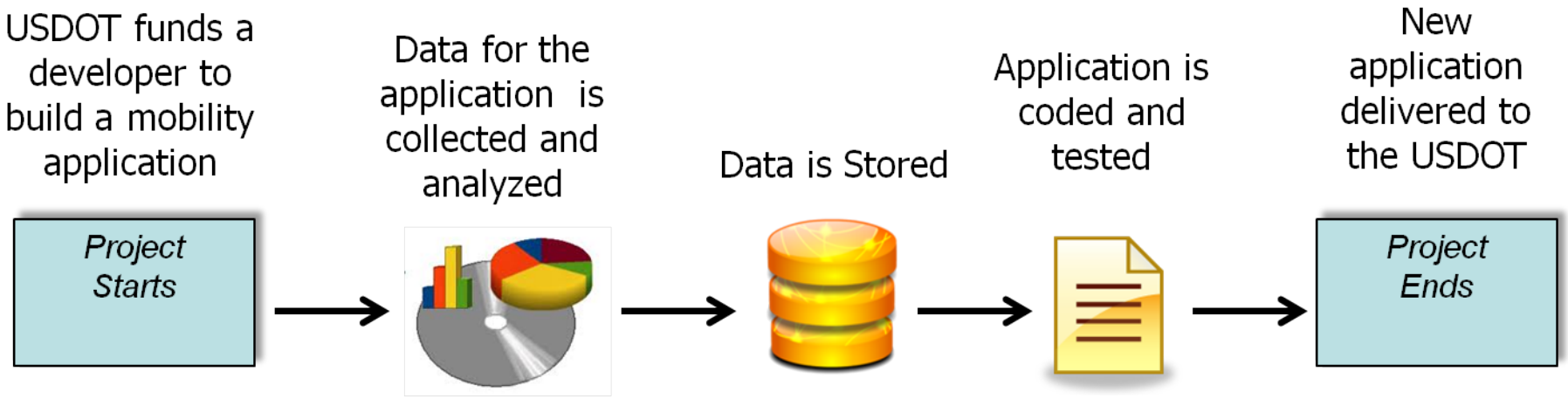
# Research Data Exchange



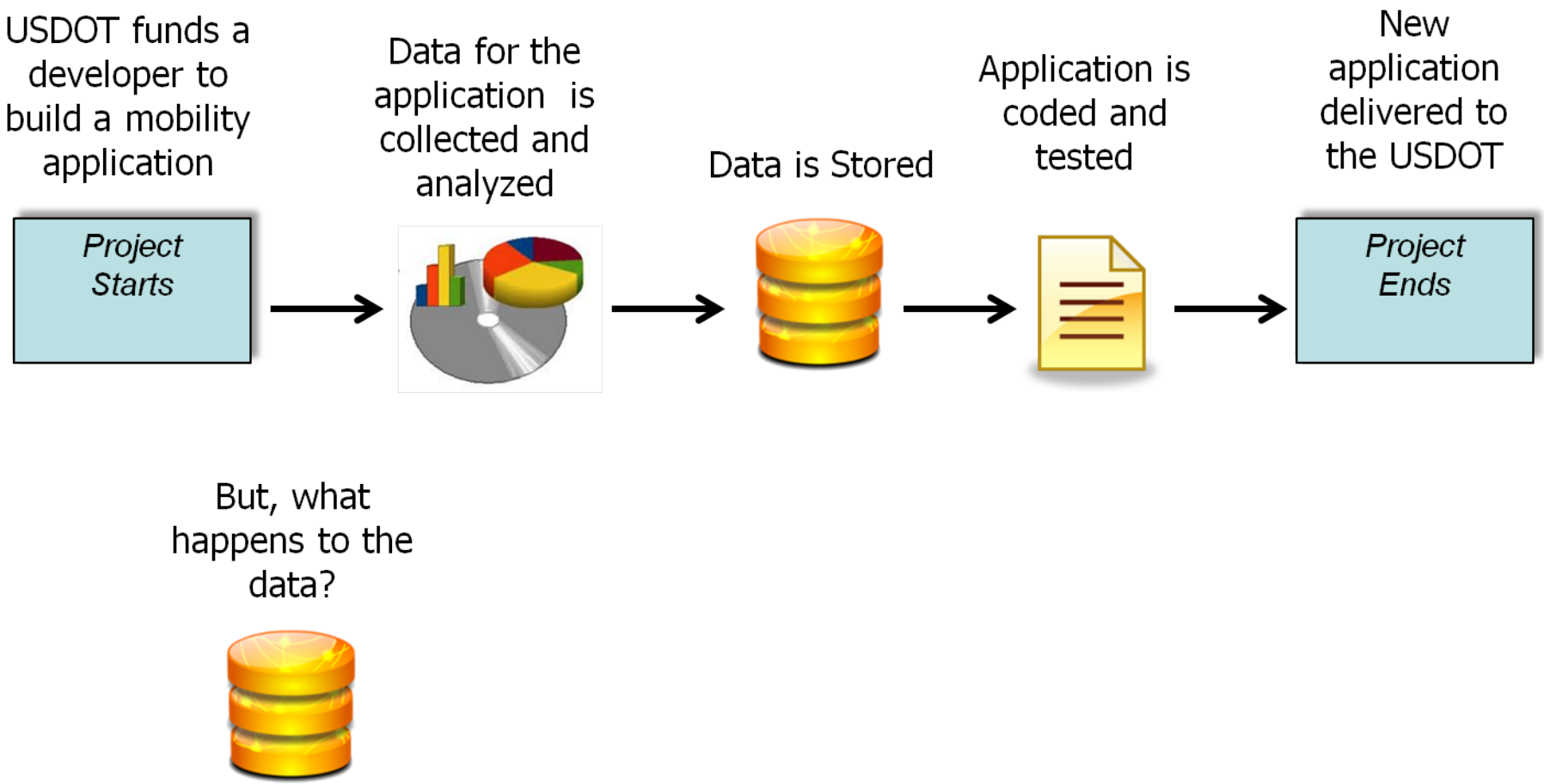
# Research Data Exchange



# Application Developer Example (without RDE)



# Application Developer Example (without RDE)

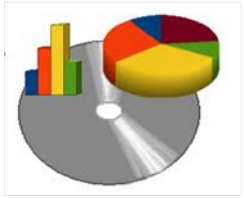


# Application Developer Example (without RDE)

USDOT funds a developer to build a mobility application

Project Starts

Data for the application is collected and analyzed



Data is Stored

Application is coded and tested

New application delivered to the USDOT

Project Ends

But, what happens to the data?

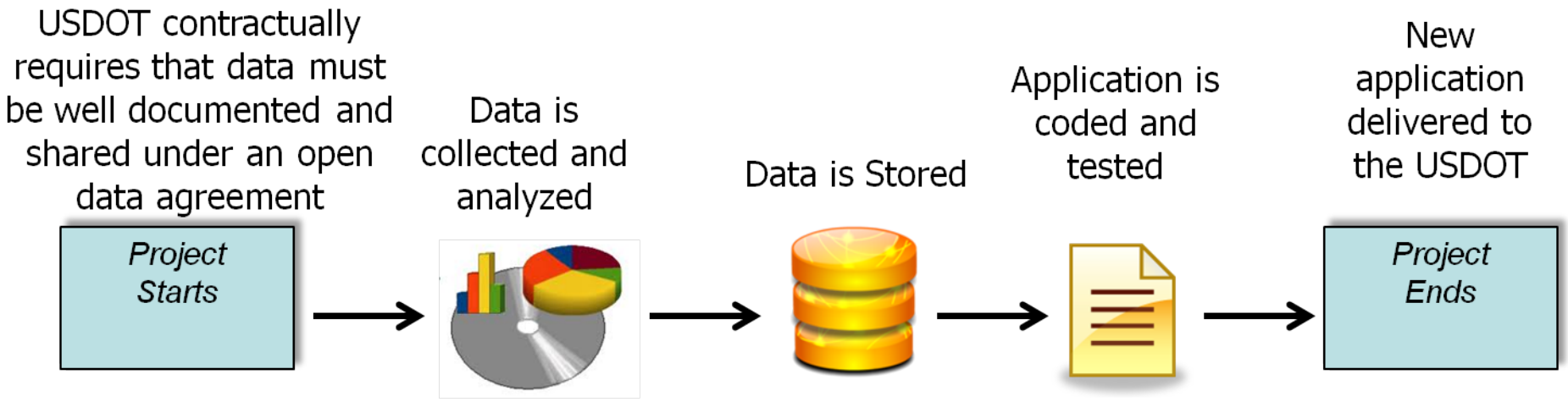
It becomes orphaned data

Sometimes the developer doing the work holds on to the data

Sometimes the data is given back to the government and stored in someone's desk

Sometimes the data is just deleted and often times the data can not be reused

# Application Developer Example (with RDE)

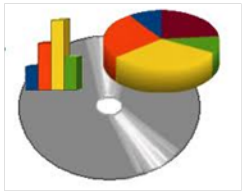


# Application Developer Example (with RDE)

USDOT contractually requires that data must be well documented and shared under an open data agreement

Project Starts

Data is collected and analyzed



Data is Stored

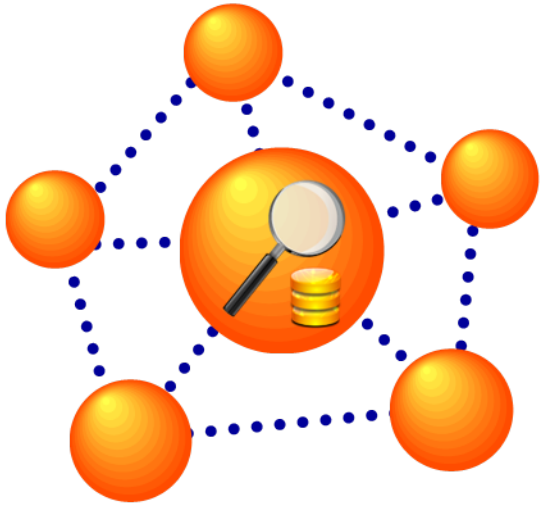
Application is coded and tested

New application delivered to the USDOT

Project Ends

But, what happens to the data?

The data are made available through the RDE while the project is ongoing and when the project is completed



RDE

# Research Example (without RDE)

Researcher from a UTC is developing and testing a real-time transit scheduler algorithm

Researcher searches multiple sites and data warehouses looking for available data sets


The researcher may find some data sets but still needs to clean and combine data to conduct the research


Researcher publishes results in a journal

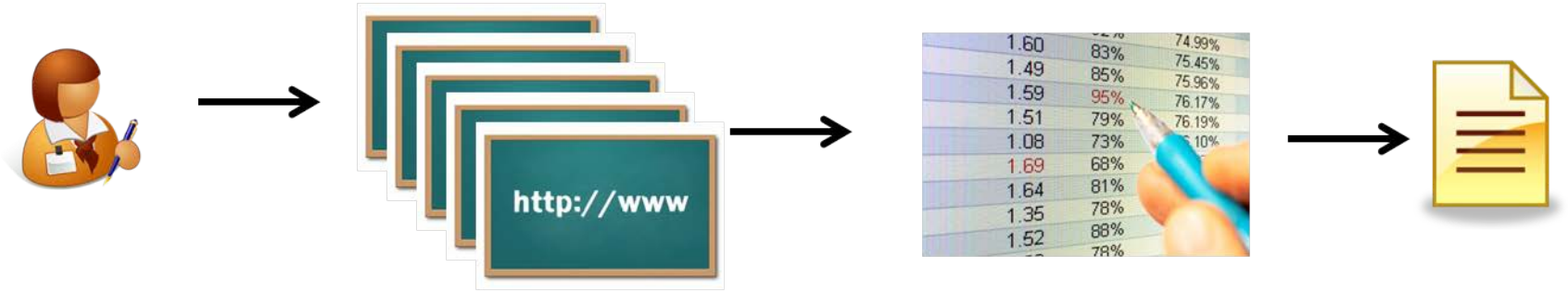
# Research Example (without RDE)

Researcher from a UTC is developing and testing a real-time transit scheduler algorithm

Researcher searches multiple sites and data warehouses looking for available data sets

The researcher may find some data sets but still needs to clean and combine data to conduct the research

Researcher publishes results in a journal




**How long does this process take?**

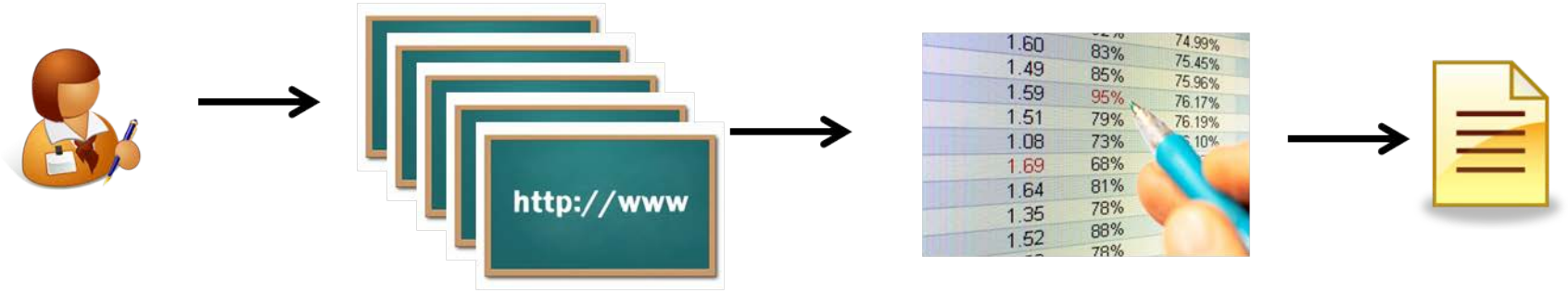
# Research Example (without RDE)

Researcher from a UTC is developing and testing a real-time transit scheduler algorithm

Researcher searches multiple sites and data warehouses looking for available data sets

The researcher may find some data sets but still needs to clean and combine data to conduct the research

Researcher publishes results in a journal




**How long does this process take?**

Searching for the data sets and getting access could take almost all of the time for the project. Most likely the researcher will have to collect their own data.

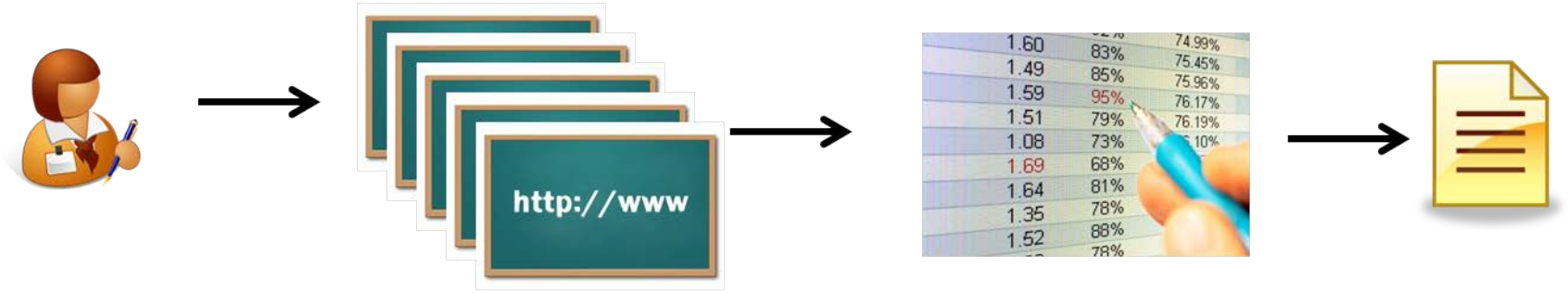
# Research Example (without RDE)

Researcher from a UTC is developing and testing a real-time transit scheduler algorithm

Researcher searches multiple sites and data warehouses looking for available data sets

The researcher may find some data sets but still needs to clean and combine data to conduct the research

Researcher publishes results in a journal




**How long does this process take?**

Searching for the data sets and getting access could take almost all of the time for the project. Most likely the researcher will have to collect their own data.

**What happens to the new data set the researcher cleaned and combined?**

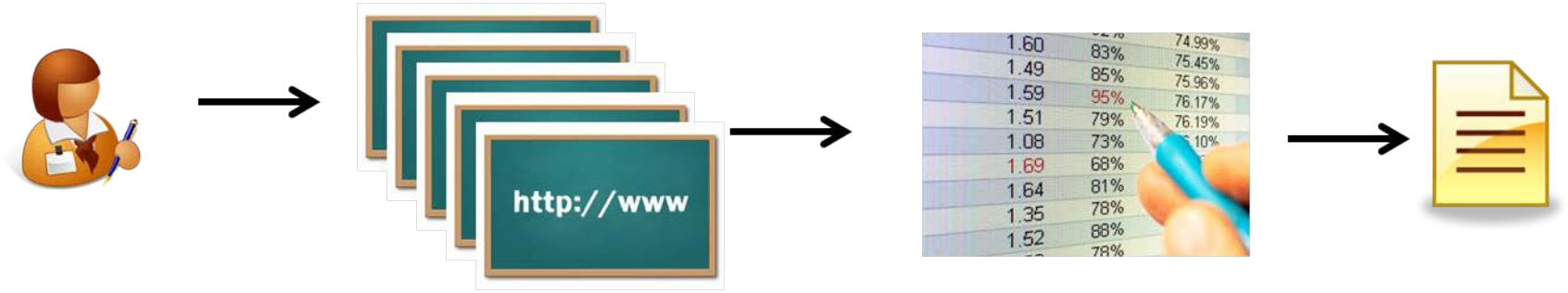
# Research Example (without RDE)

Researcher from a UTC is developing and testing a real-time transit scheduler algorithm

Researcher searches multiple sites and data warehouses looking for available data sets

The researcher may find some data sets but still needs to clean and combine data to conduct the research

Researcher publishes results in a journal




**How long does this process take?**

Searching for the data sets and getting access could take almost all of the time for the project. Most likely the researcher will have to collect their own data.

**What happens to the new data set the researcher cleaned and combined?**

All the work the researcher has done with the data is lost because the researcher has no place to store the new data set

# Research Example (with RDE)

Researcher from a UTC is developing and testing a real-time transit scheduler algorithm

Researcher searches RDE Portal for data sets and quickly finds data she needs

Data Portal

The researcher combines and reformats the data


Researcher publishes results in a journal

# Research Example (with RDE)

Researcher from a UTC is developing and testing a real-time transit scheduler algorithm

Researcher searches RDE Portal for data sets and quickly finds data she needs

The researcher combines and reformats the data

Researcher publishes results in a journal


**How long does this process take?**

# Research Example (with RDE)

Researcher from a UTC is developing and testing a real-time transit scheduler algorithm

Researcher searches RDE Portal for data sets and quickly finds data she needs

The researcher combines and reformats the data

Researcher publishes results in a journal


**How long does this process take?**

Having a one-stop search location for all USDOT data sets and other federated data sets saves time and gives the researcher a community to work with to help with the research

# Research Example (with RDE)

Researcher from a UTC is developing and testing a real-time transit scheduler algorithm

Researcher searches RDE Portal for data sets and quickly finds data she needs

The researcher combines and reformats the data

Researcher publishes results in a journal


**How long does this process take?**

Having a one-stop search location for all USDOT data sets and other federated data sets saves time and gives the researcher a community to work with to help with the research

**What happens to the new data set the researcher reformatted and combined?**

# Research Example (with RDE)

Researcher from a UTC is developing and testing a real-time transit scheduler algorithm

Researcher searches RDE Portal for data sets and quickly finds data she needs

The researcher combines and reformats the data

Researcher publishes results in a journal


**How long does this process take?**

Having a one-stop search location for all USDOT data sets and other federated data sets saves time and gives the researcher a community to work with to help with the research

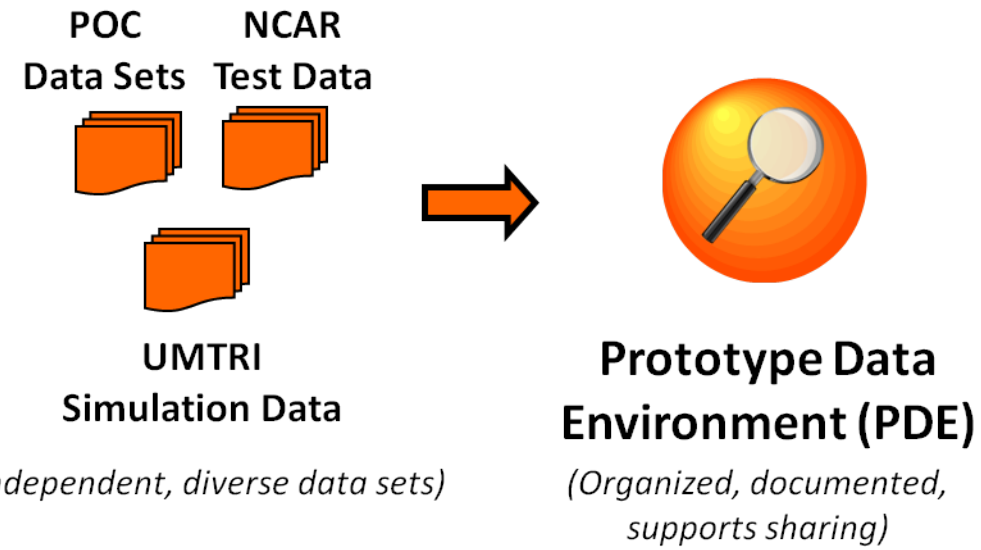
**What happens to the new data set the researcher reformatted and combined?**

The researcher then can resubmit their updated data set to the RDE for others to use in future projects.

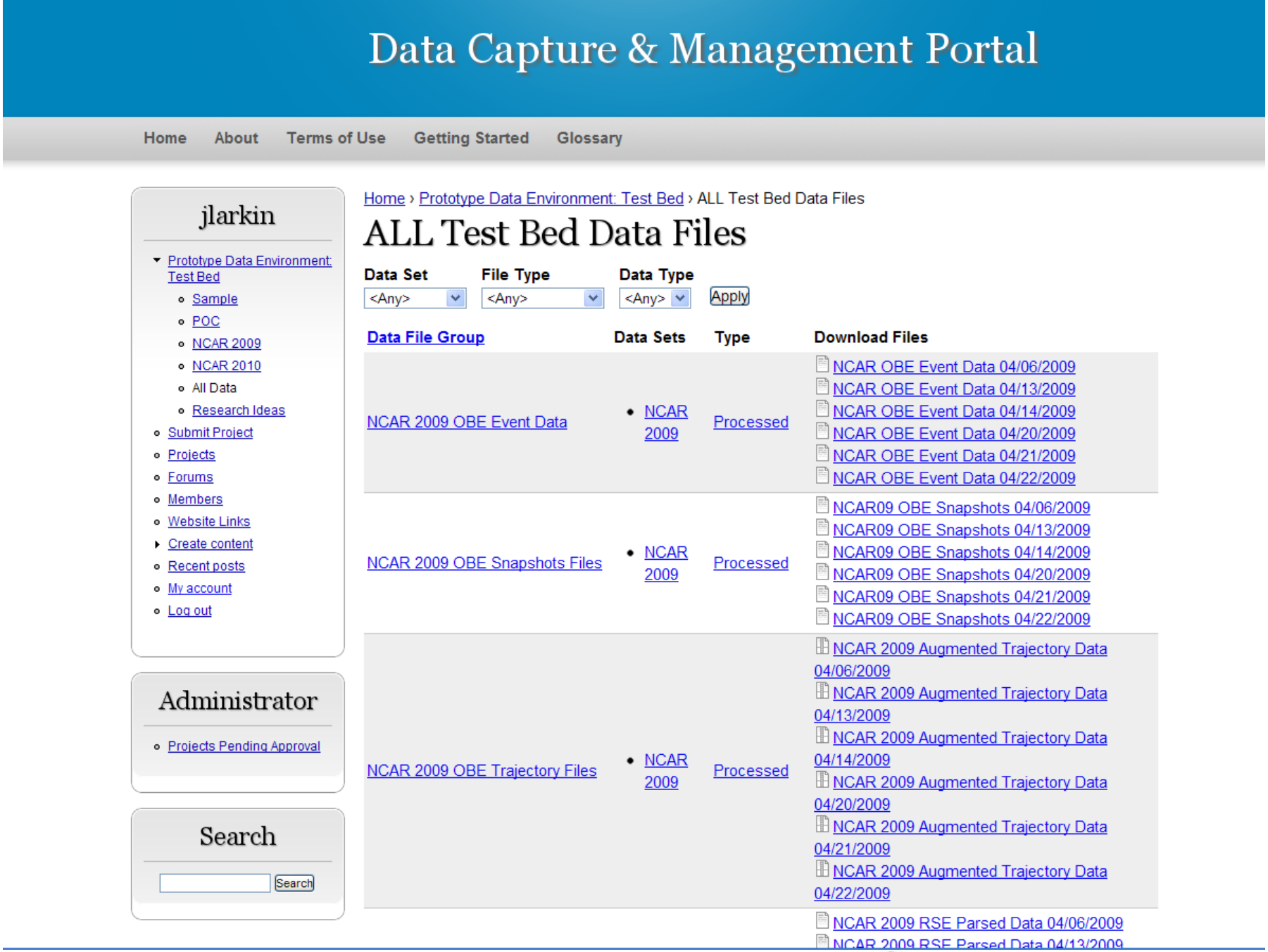
# Prototype Data Environment

- https://datacapture.noblis.org/

- Prototypes one component of a federated system
  - Refine the Data Environment concept
  - Test key hypotheses about governance and user collaboration




- Data (and meta-data) regarding the V2V/V2I Technology Test Bed (MI)
  - Documented probe data samples from recent tests (POC/NCAR)
  - Simulated 100% market penetration data for the test bed environs
    - *contributed by the University of Michigan Transportation Research Institute (UMTRI)*
  - Open source analytical tools
  - Forums for researchers to register projects, flag erroneous data, contribute analyses and data views




*(Screenshot of Prototype Data Environment download page)*

# Prototype Data Environment: Current Utilization and Insights

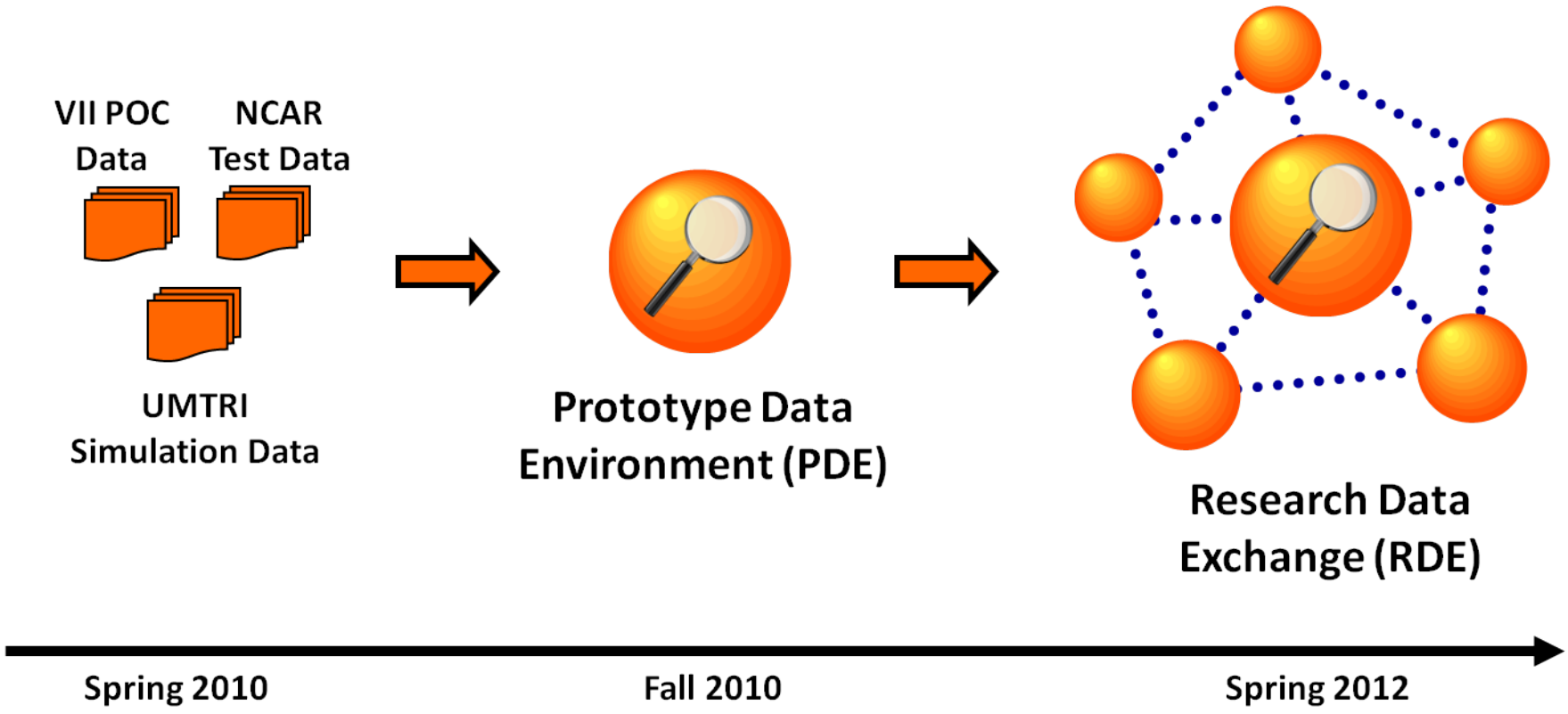
- **Four registered research projects** using the POC/NCAR data
  - U. of Washington:  Adaptive Vehicle Routing on Arterial Networks
  - U. of Virginia: Traffic Signal Control and Performance Measures
  - PATH: Advanced Traffic Signal Algorithms
  - Virginia Tech: Traffic Responsive Signal Control

Prototype Data Environment (PDE)

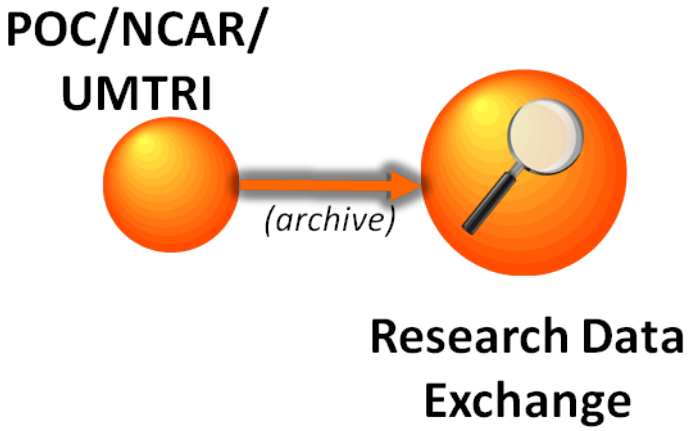
- **1870 total visits, 723 unique visitors, 38 countries**
  - Most popular features are home page and data download page
  - Steady monthly utilization, number of unique visitors continues to rise

- Insights from the Prototype Data Environment
  - Revealed **need for a complex system of multiple data environments**
    - Research Data Exchange (RDE) Concept of Operations
  - Enabled internal collaboration, engaged stakeholders
    - Focal point for discussions on Open Data concepts

# Evolution from Independent Data Sets to Research Data Exchange

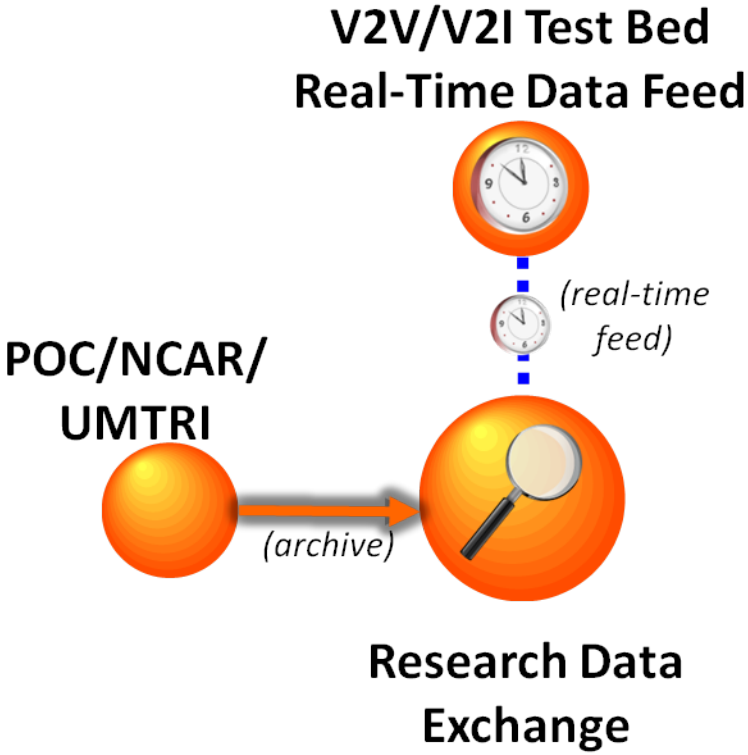


# Incrementally Constructing the RDE: Transition Prototype Data Environment



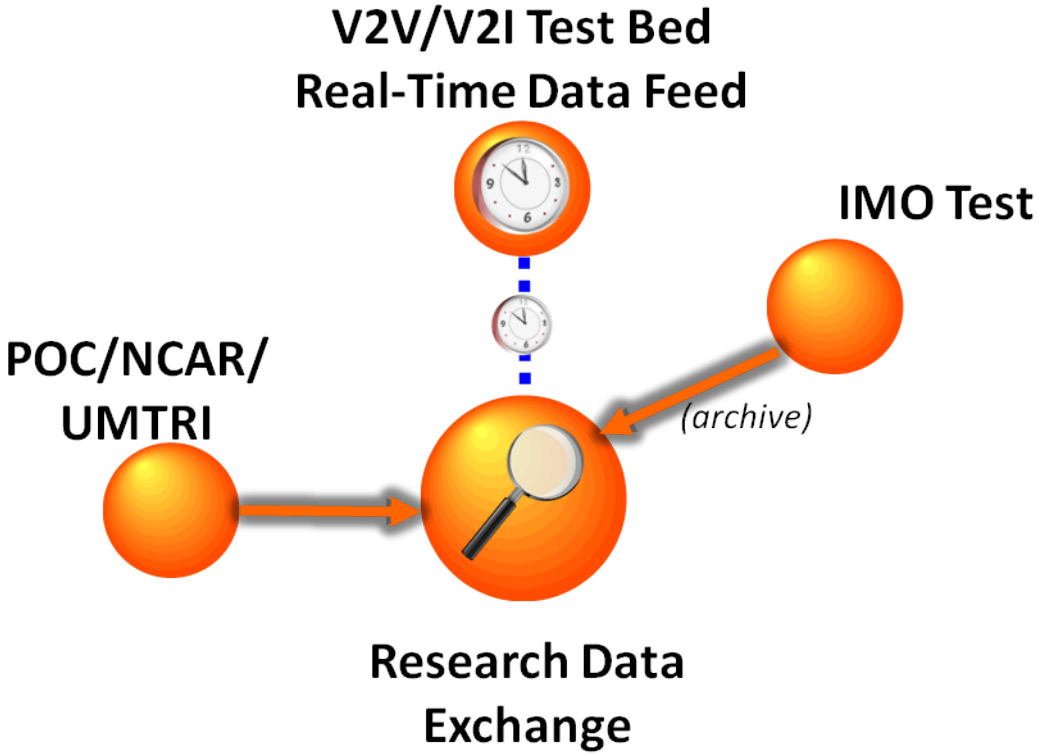

- Transition POC/NCAR/UMTRI data from Prototype Data Environment to RDE

# Incrementally Constructing the RDE: Establish Real-Time Data Feed from Test Bed



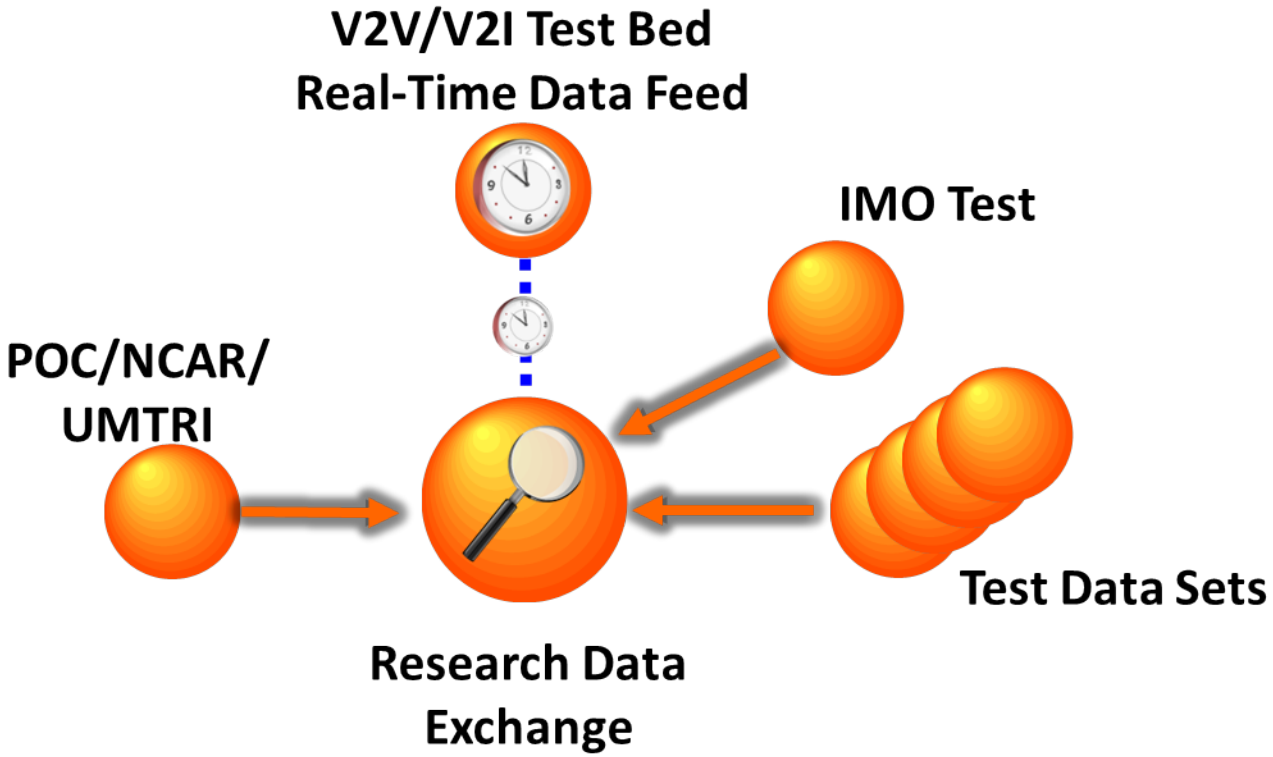

- Maintain capability to feed data directly in real-time from tests at the V2V/V2I Test Bed (Michigan) and tests at the FHWA Turner-Fairbank R&D facility
- This same capability will be re-used for other real-time feeds connected to the RDE
  - E.g., World Congress Demo

# Incrementally Constructing the RDE: Incorporate Weather Program IMO Test Data



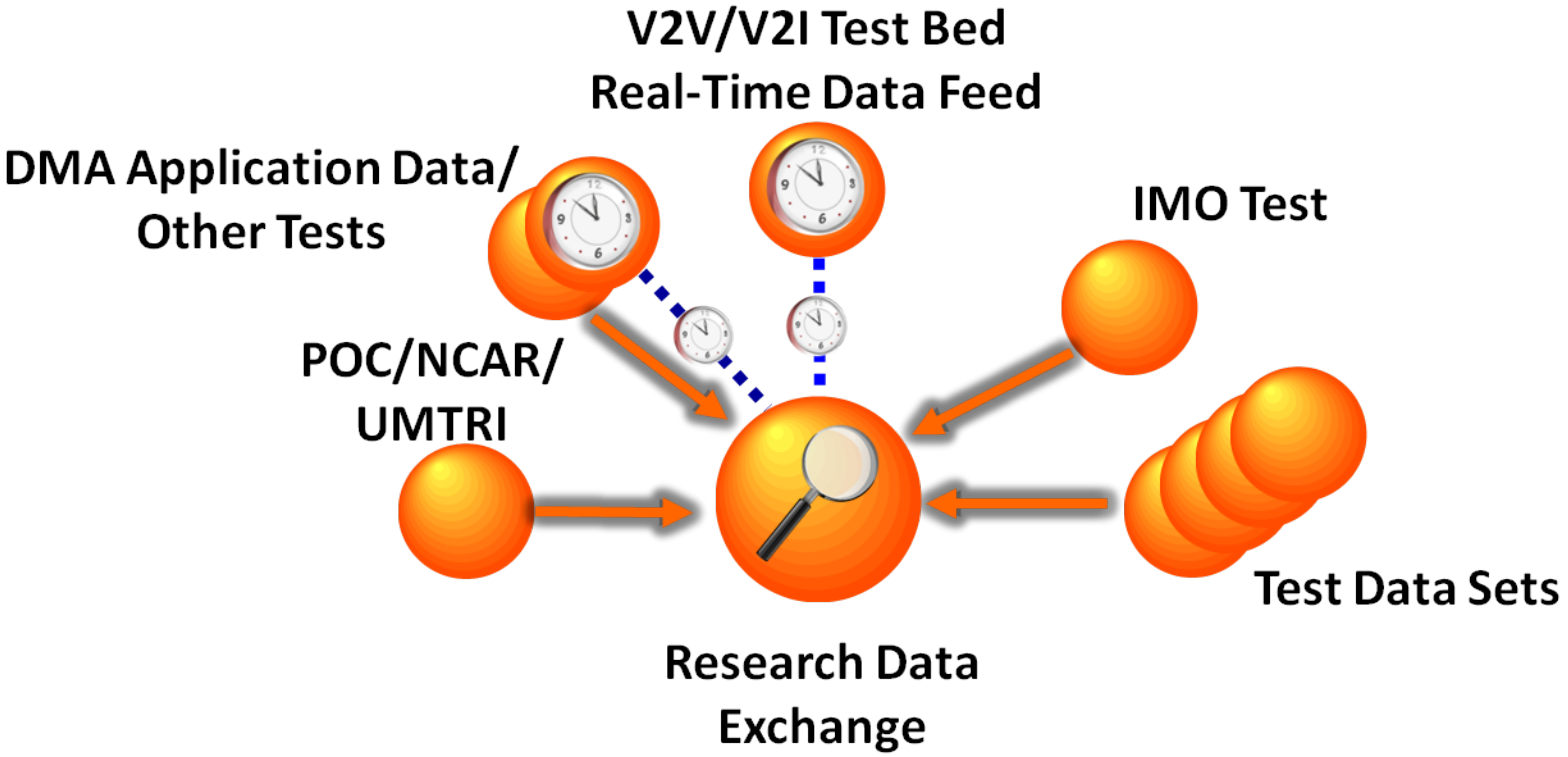

- Integrated Mobility Observations (IMO) Test ongoing
  - Snow plow connected vehicle data from Nevada and Minnesota
  - NCAR coordinating the preparation of data from participating agency vehicles to be archived, with meta-data, within the RDE

# Incrementally Constructing the RDE: Incorporate Test Data Sets




- Test Data Set RFA/RFP
  - Assemble and document high-value data from already-conducted tests
  - Four awards have been made (Summer 2011)
    - Data sets will be incorporated into RDE archives (Seattle, Portland, Pasadena, San Diego)

# Incrementally Constructing the RDE: Data from Application Development/Other Tests




- The RDE can serve as a repository for archived data from other tests
  - Data generated as part of transformative application development and testing
  - Other programs/volunteer data
- The RDE will not serve as a junk closet, however

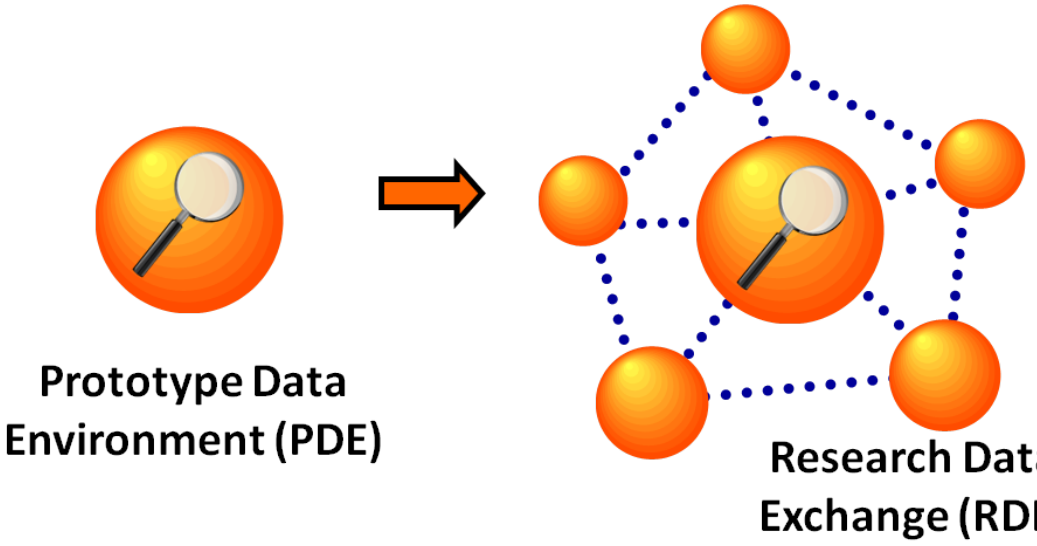
# RDE Concept of Operations Document

- Document purpose is to communicate an understanding of user needs and to describe how the system will operate to fulfill those needs.
- The audience for the document includes:
  - System developers who will create and support the RDE
  - USDOT mobility program stakeholders
  - Analysts, researchers, and mobility application developers requiring access to research data  for analysis and application development

# RDE Concept of Operations Document

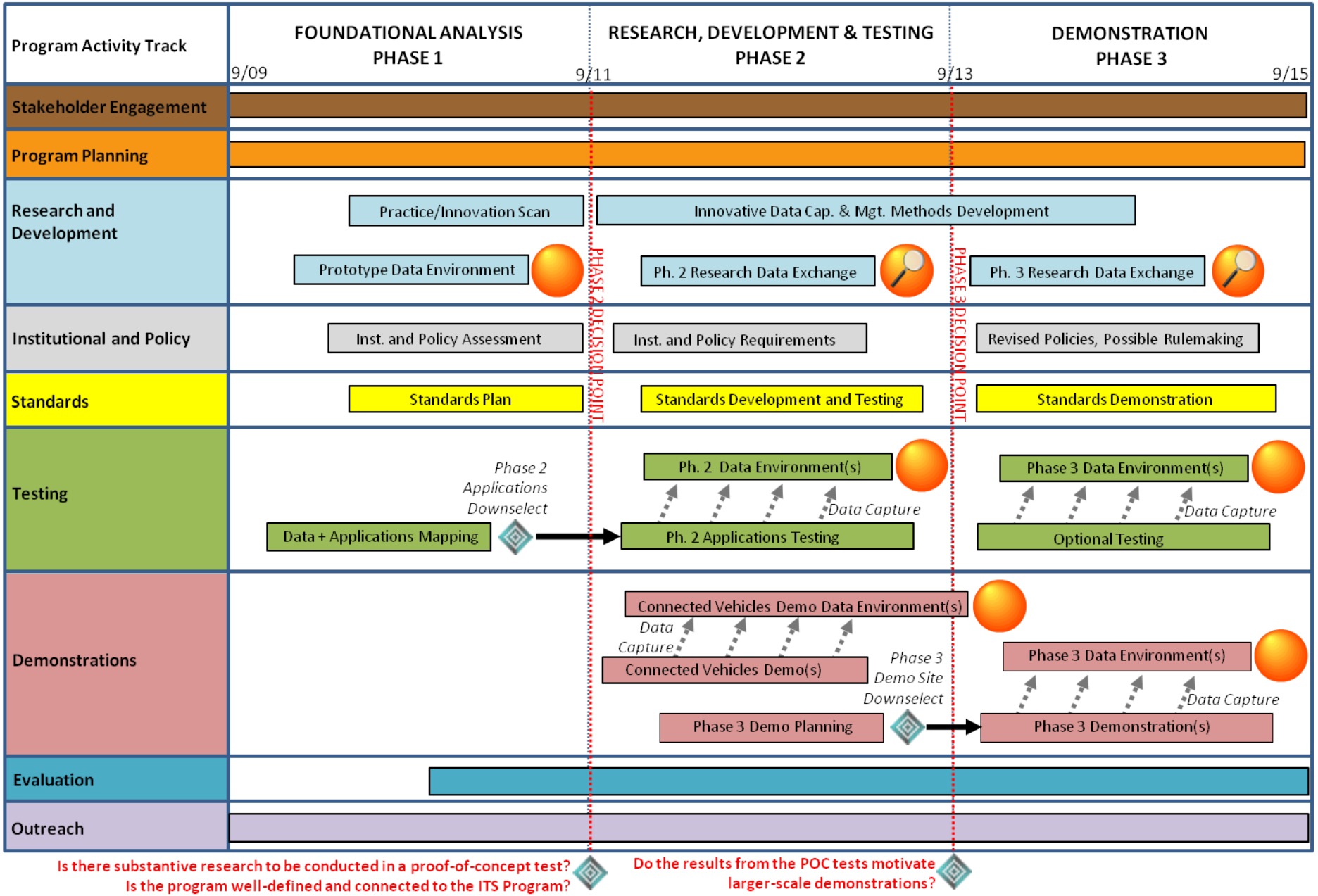
- Content follows IEEE Guide for Concepts of Operations, including:
  - Current System or Situation
  - Justification for and Nature of Changes
  - Concepts for the Proposed System
  - Operational Scenarios
  - Summary of Impacts
  - Analysis of the Proposed System
- Draft RDE ConOps document is posted on….
- We welcome your comments on the document or the concepts presented during today's webinar
- Comments will be addressed and a revised draft of the document will be posted on the ITS JPO website

# Research Data Exchange (RDE): Key Goals, 12-Month Outlook




- Initiate RDE Development and Management procurement (Fall 2011)
- Establish real-time data feed with V2V/V2I Test Bed (Fall 2011)
- Incorporate new data sets
  - IMO Test Data (Fall 2011)
  - World Congress Demonstration Data (Fall 2011)
  - Test Data Sets (Winter/Spring 2012)
- PDE-to-RDE transition (March 2012)
- **These data are our initial steps in providing a data-rich environment for transformative applications development**

# Data Capture and Management Program: High-Level Roadmap




Is there substantive research to be conducted in a proof-of-concept test? Is the program well-defined and connected to the ITS Program?

Do the results from the POC tests motivate larger-scale demonstrations?

High-Level Roadmap v1.3 ( 5/9/2011)

LEGEND: *Data Capture* Data Feed; Decision point; Data Environment; Research Data Exchange; Program Activity

# For more information …

**US DOT ITS JPO Website**

www.its.dot.gov

**Dale Thompson**
US DOT ITS Joint Program Office

Dale.Thompson@dot.gov

**Gene McHale**
FHWA Office of Operations R&D

Gene.McHale@dot.gov