UNITED STATES
DEPARTMENT OF TRANSPORTATION

Data Capture and Management (DCM) Program Overview

Dale Thompson, ITS JPO
Gene McHale, FHWA

Mobility Workshop 2012
National Harbor, MD

May 24, 2012

Dale Thompson
ITS-JPO

Overview

- Introduction
  - Key Concepts
  - Data Capture Program Roadmap
- Current Projects and Products
  - Research Data Exchange (RDE)
  - Test Data Sets
  - RDE demonstration
- Critical Issues for DCM
- RDE Next Steps
- Stakeholder Q&A

Data Capture and Management Program (DCM): 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

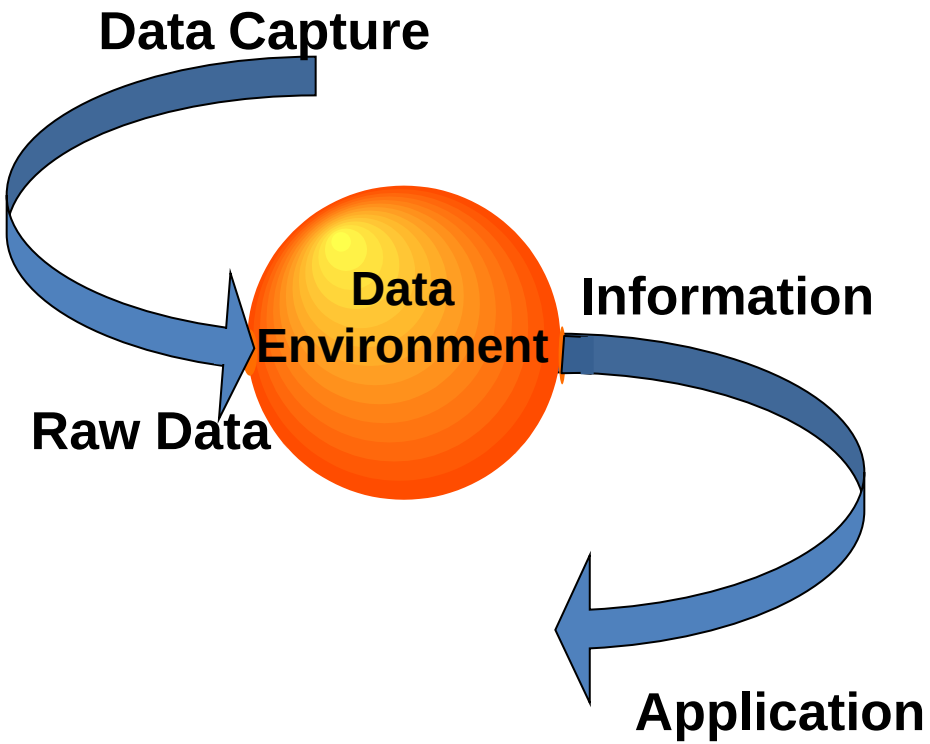
Program Partners

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

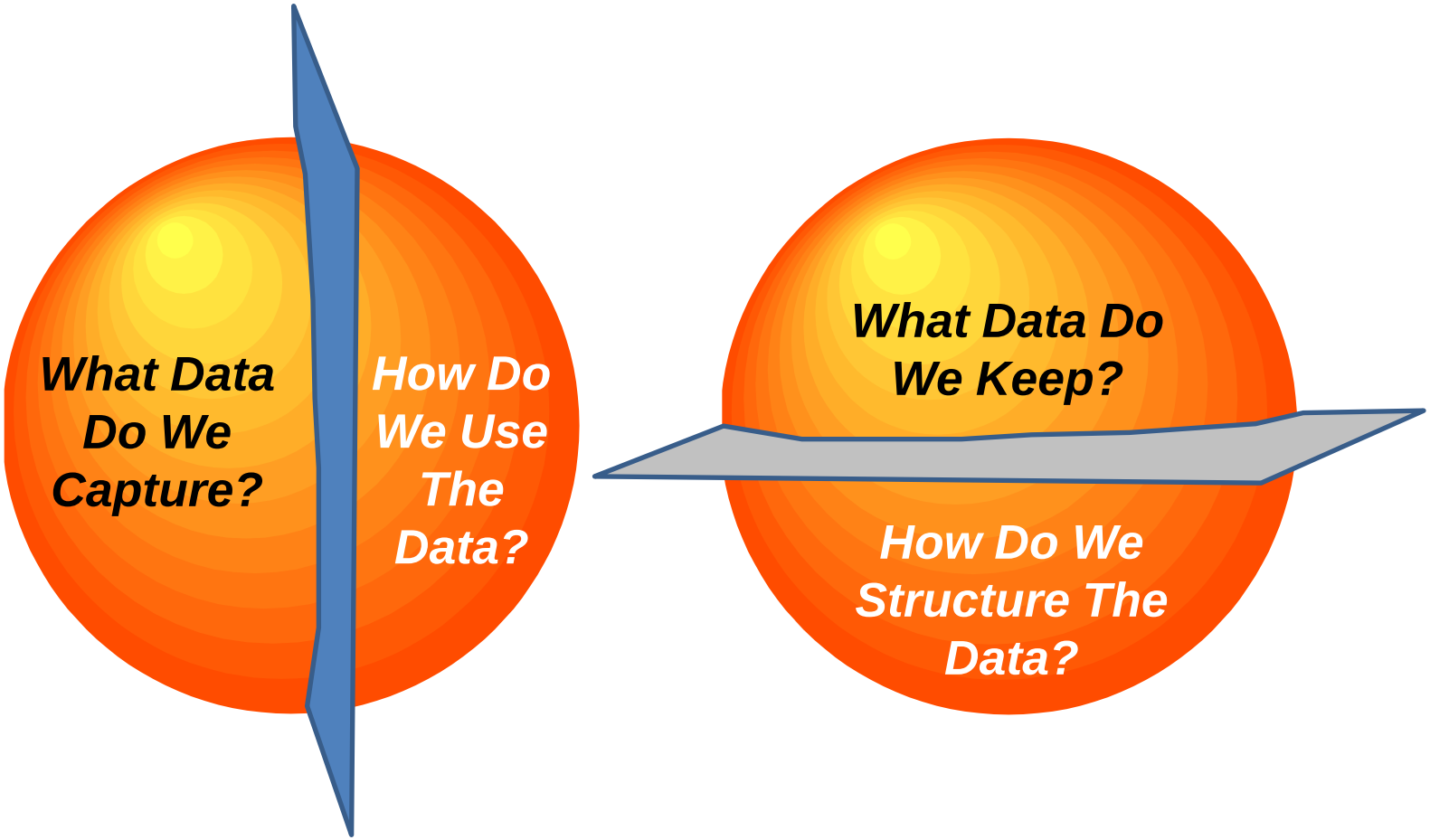
Data Environments

Data environment:

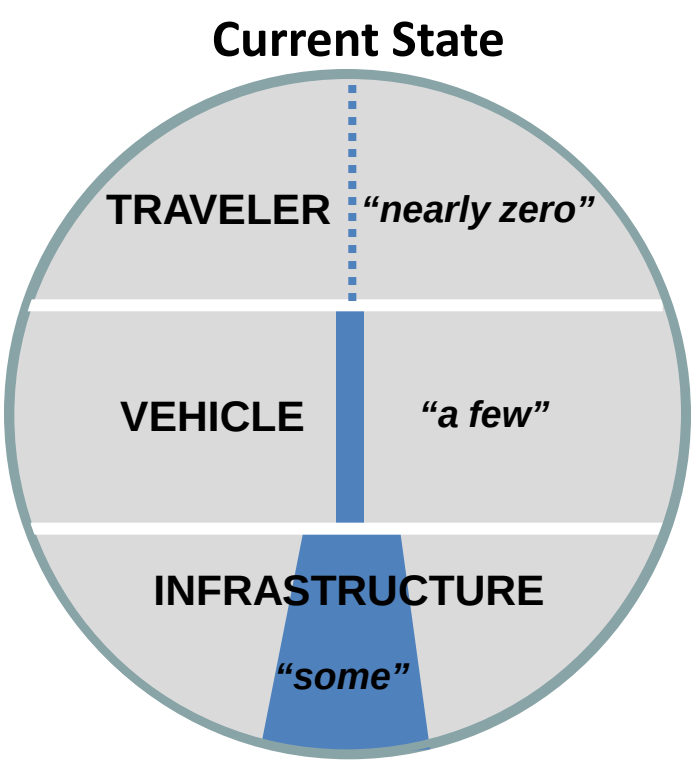
- well-organized collection of data of specific type and quality
- captured and stored at regular intervals from one or more sources
- systematically shared in support of one or more applications

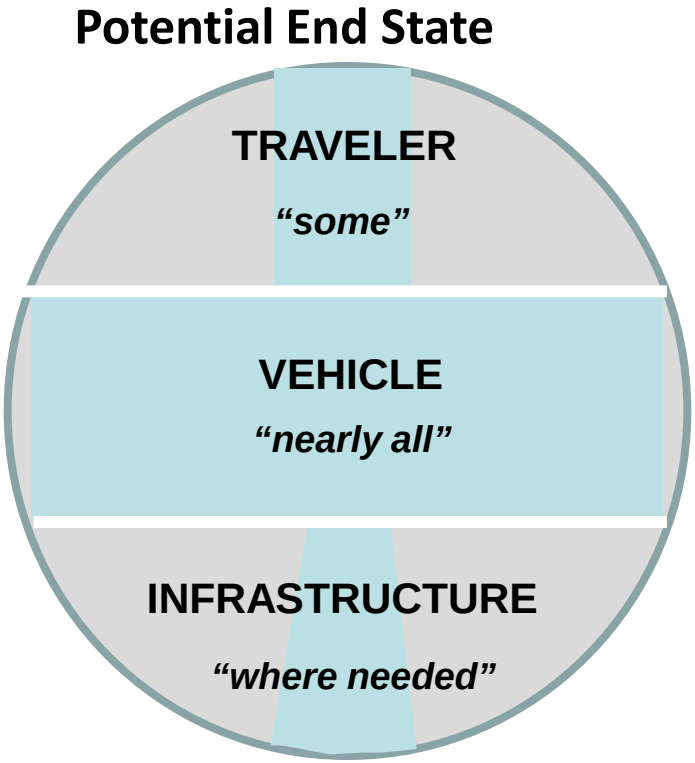


Key Issues in Defining a Data Environment

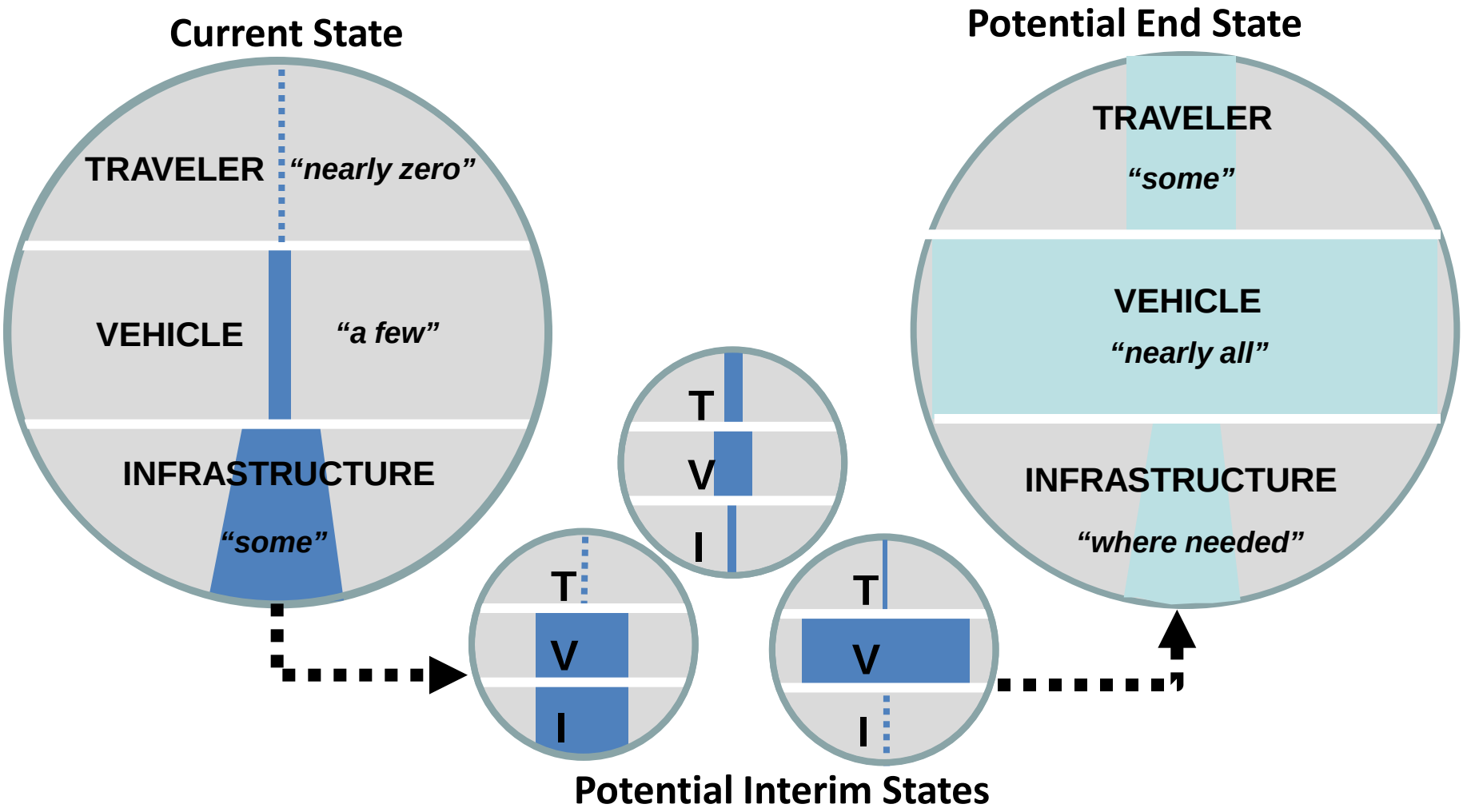


Data Environment Evolution





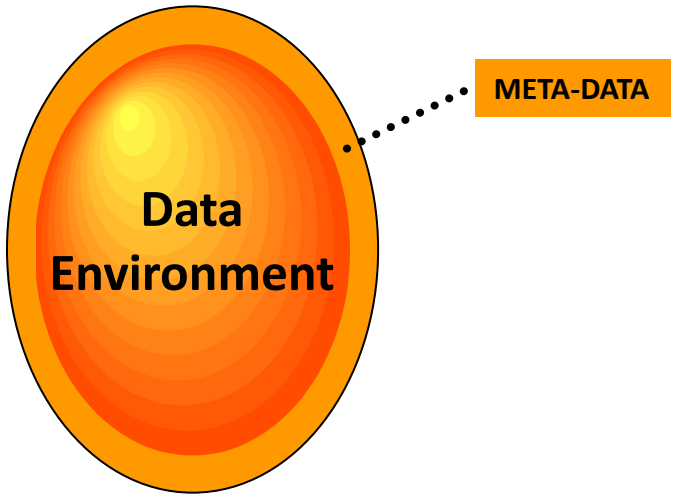
Data Environment Evolution



Elements of Data Capture and Management

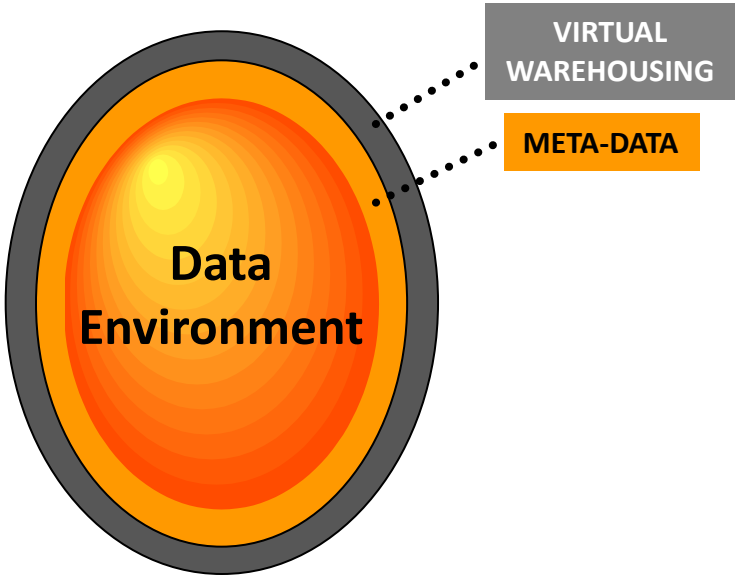
Elements of Data Capture and Management

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



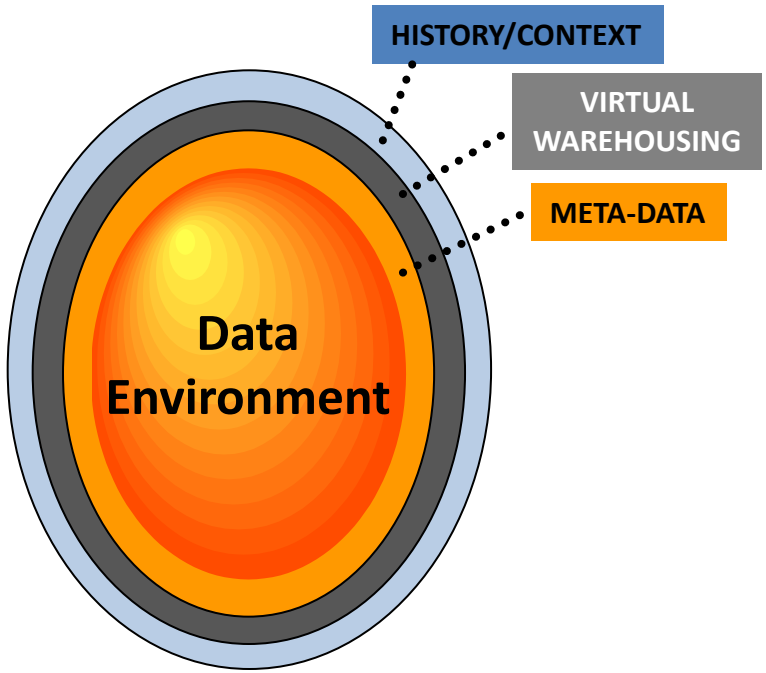
Elements of Data Capture and Management

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



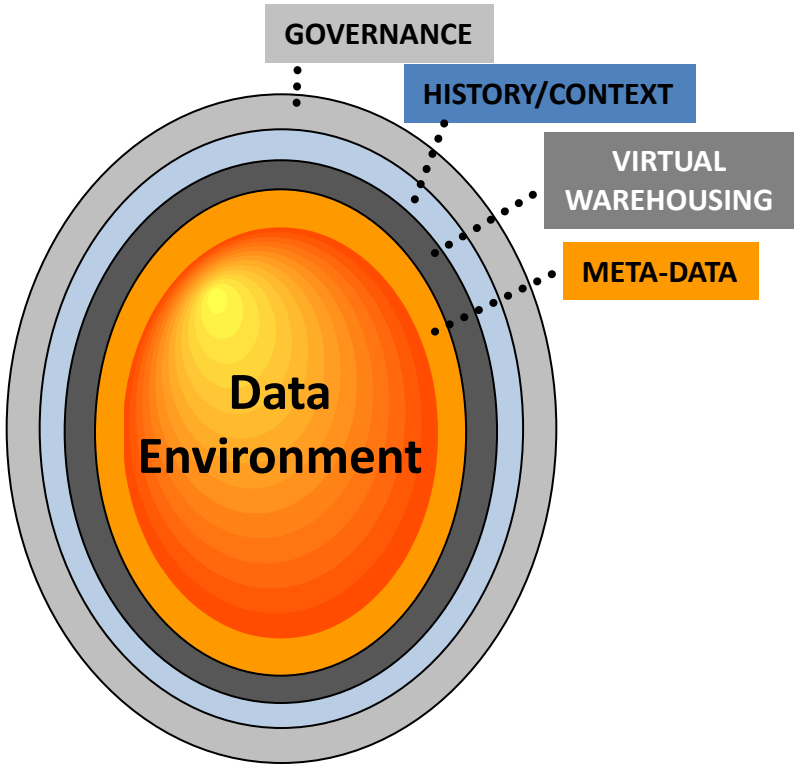
Elements of Data Capture and Management

- **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



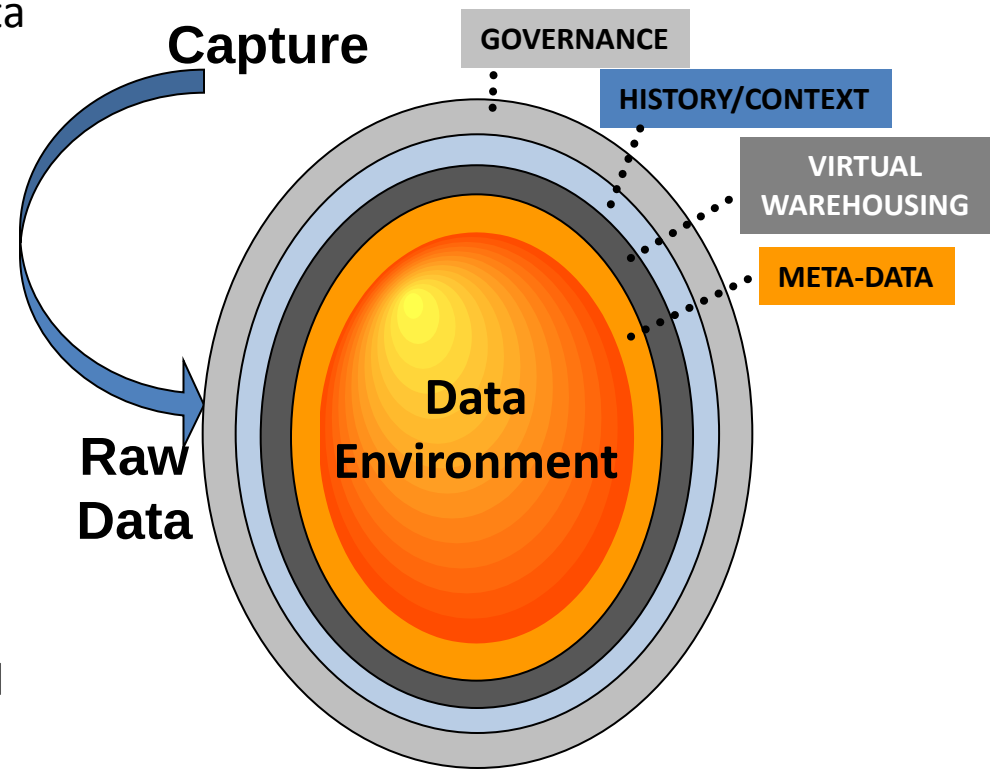
Elements of Data Capture and Management

- **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



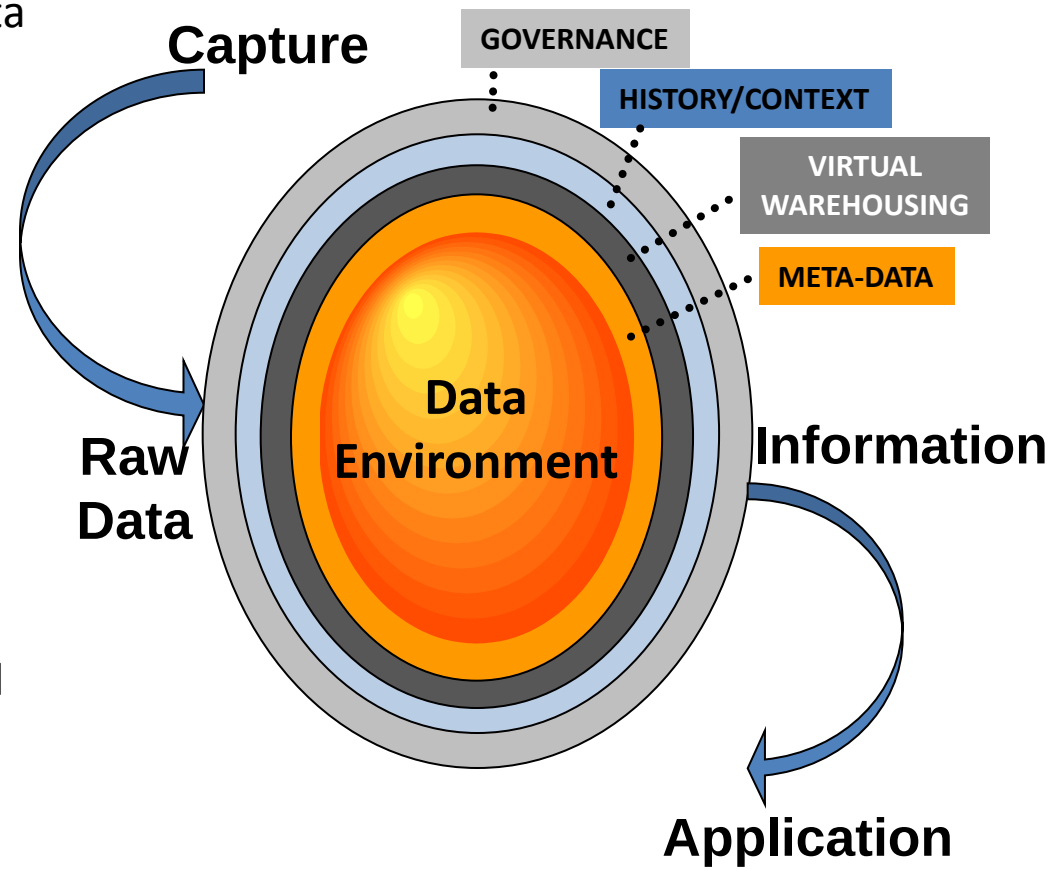
Elements of Data Capture and Management

- **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

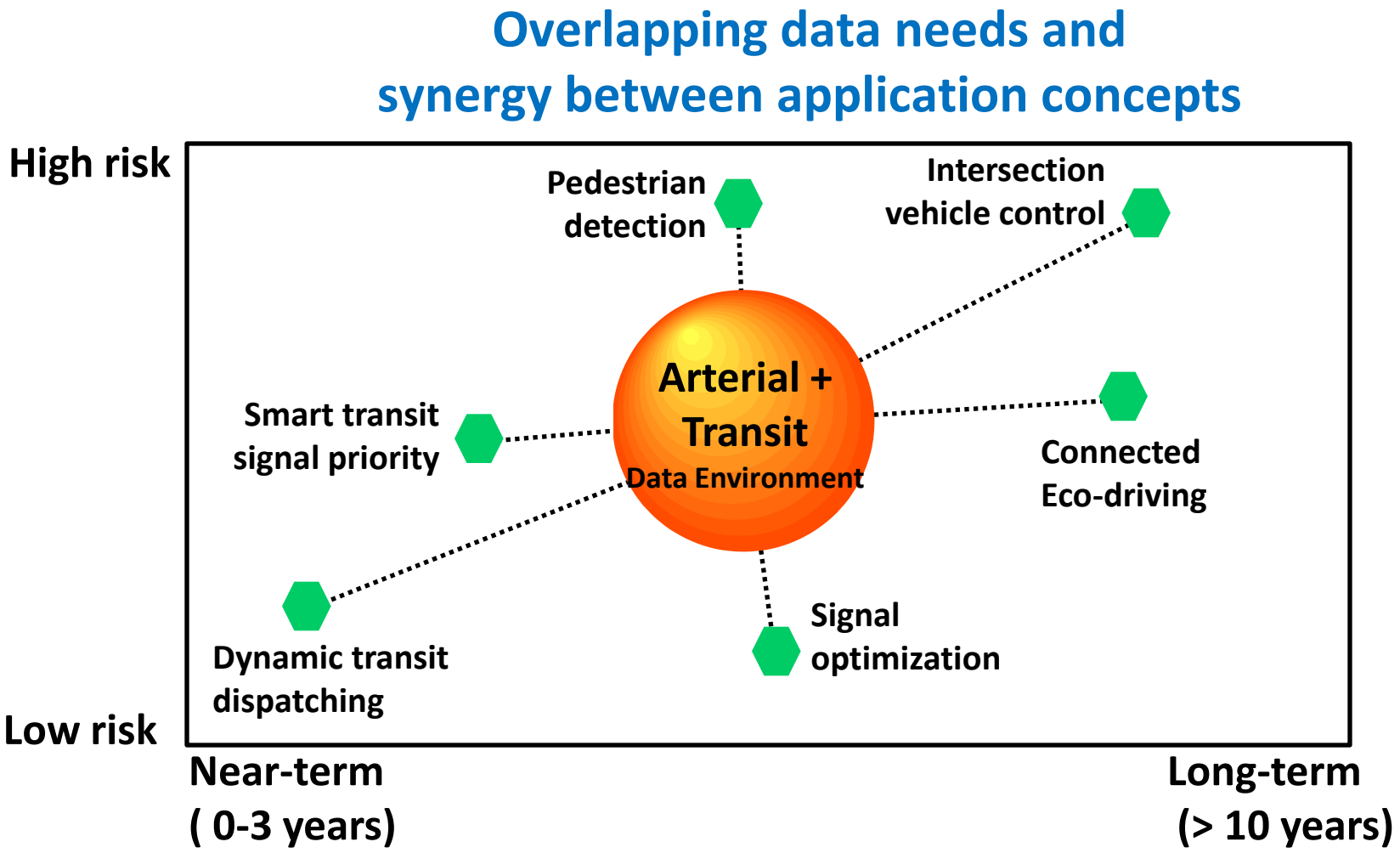


Elements of Data Capture and Management

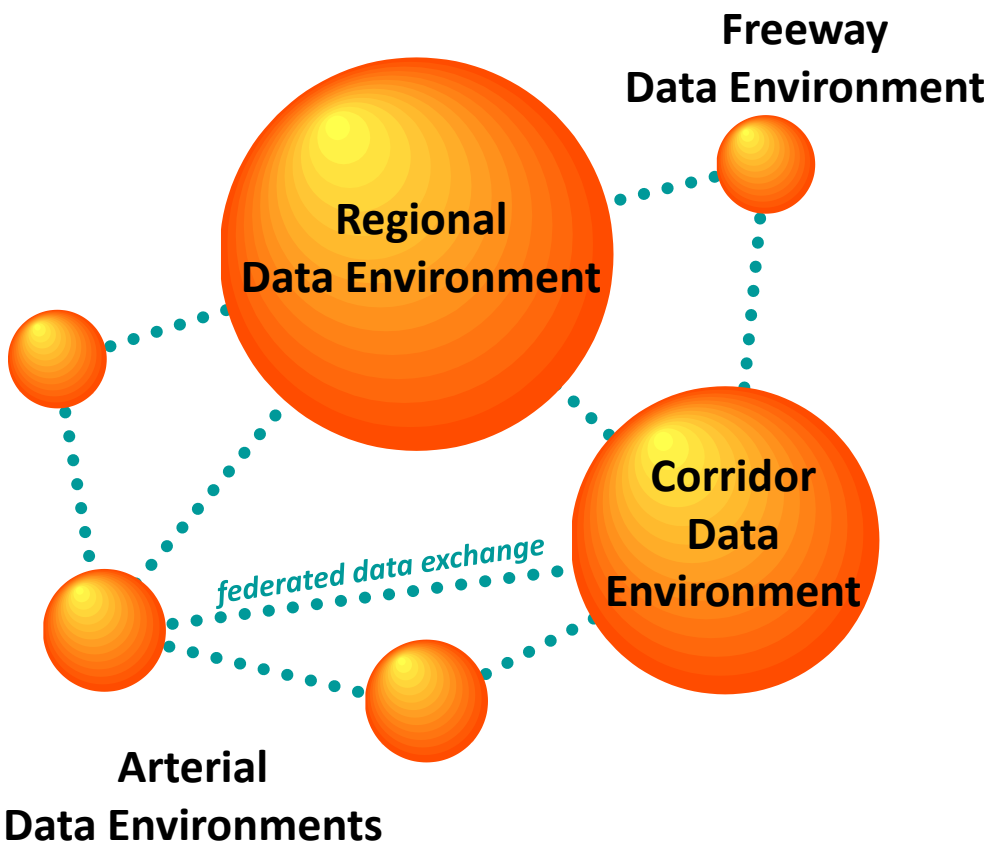
- **Meta data**:
  - Provision of well-documented data environment
- **Virtual warehousing**:
  - Supports access to data environment and forum for collaboration
- **History/context**:
  - Objectives of data assembly
- **Governance**:
  - Rules under which data environment can be accessed and procedures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



Each Data Environment Supports Multiple Apps

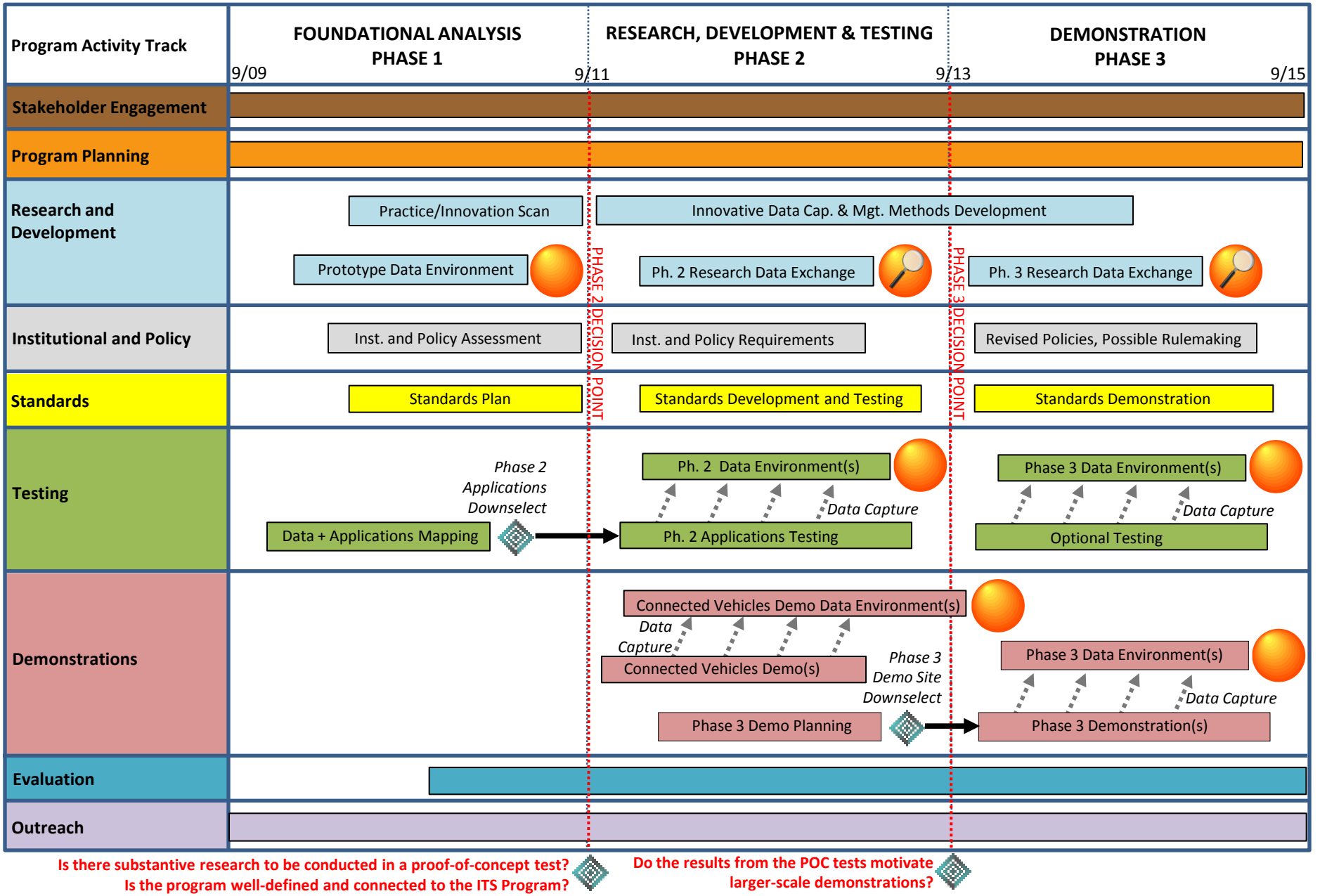


Federated Data Environments



- **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

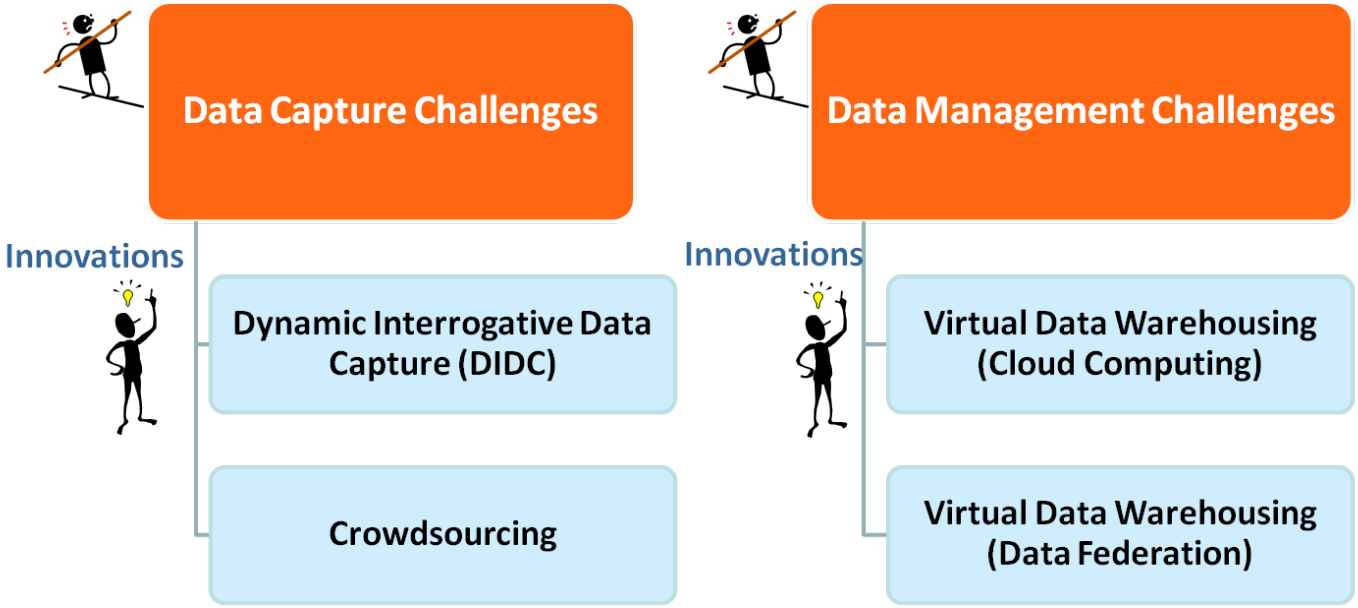
Data Capture and Management Program: High-Level Roadmap



High-Level Roadmap v1.4 (5/17/2011)

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

Innovations Scan Project



- Assessed industry best practices in data capture and management methods and technologies that are applicable to the DCM Program
- Identified four "most promising" emerging concepts and technologies
- USDOT Lead: Mohammed Yousuf (FHWA R&D)
- Contractor: SAIC/Delcan/University of Virginia (PI: Dick Mudge)
- Study Period: 9/22/10 - 11/30/11

Data Capture Challenges

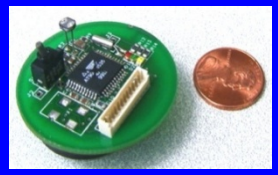
Challenge	Issue	Innovation
Bandwidth Overload	Potential data explosion due to new forms of data will likely over-burden the computational and communication systems	Dynamic Interrogative Data Capture (DIDC)
Data from Travelers	Envisioned transformative applications require new forms of real time and archived data that are extremely costly to obtain, or create possible privacy conflicts if required from all vehicles or travelers	Crowdsourcing
Large Volumes of Data	Large volumes of diverse spatial data call for new methods of data management	Virtual Data Warehousing: • Cloud Computing • Data Federation

Dynamic Interrogative Data Capture (DIDC)

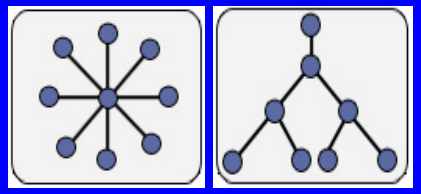
- Each device (in a vehicle, on the infrastructure, or on a person) can set and reset message priorities to different data elements
- Each device can intelligently and dynamically decide on data aggregation levels and transmission frequencies, based on its own state (local conditions) as well as the state of the network (global conditions)
- Each device can query other devices in its vicinity, depending on its data needs, and request certain data aggregation levels

Example of DIDC Concept

Wireless Sensor Networks (WSNs): Adaptive management of data flow in complex systems



Mote Technology



Example WSN Topologies

- **Value to DCM**: Can cut communications and energy costs up to 99% compared with fixed interval approaches -- with same impact (defense applications)
- **Challenges**: Adapting and structuring an analog system for transportation, determining potential benefits

Crowdsourcing

- Practice of tapping into the collective intelligence of the public at large to complete tasks that a company would normally either perform itself or outsource to a known entity (blend of crowd and outsourcing)
  - Improves productivity
  - Minimizes labor and research expenses
  - Consumers involved in creating product

- Crowdsourced Traveler Data
  - **Inrix**: provides traffic information using crowdsourced traffic data)
  - **Waze**: provides 100% crowsourced, free real-time traffic information on mobile devices

- **Value to DCM**: Opt-in crowdsourcing could lead to systematic capture of traveler trip data including itinerary and traveler behavior/response data (transform ATIS and multi-modal corridor management?)
- **Challenges**: Structuring public-private partnerships, privacy

Cloud Computing

- "Model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources that can be rapidly provisioned and released with minimal management effort or service provider interaction" (*Source: National Institute of Standards and Technology (NIST)*)

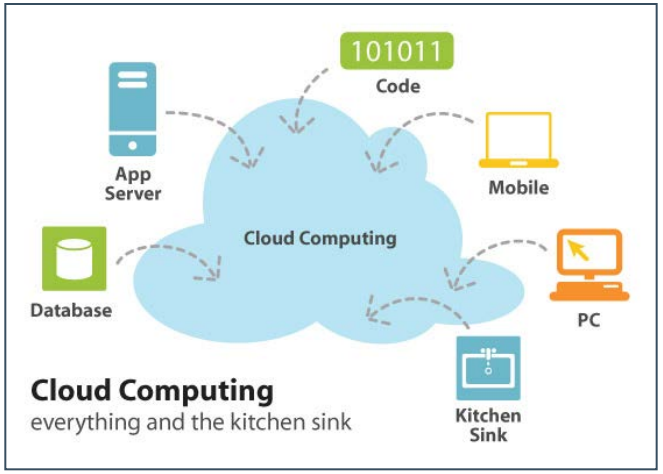


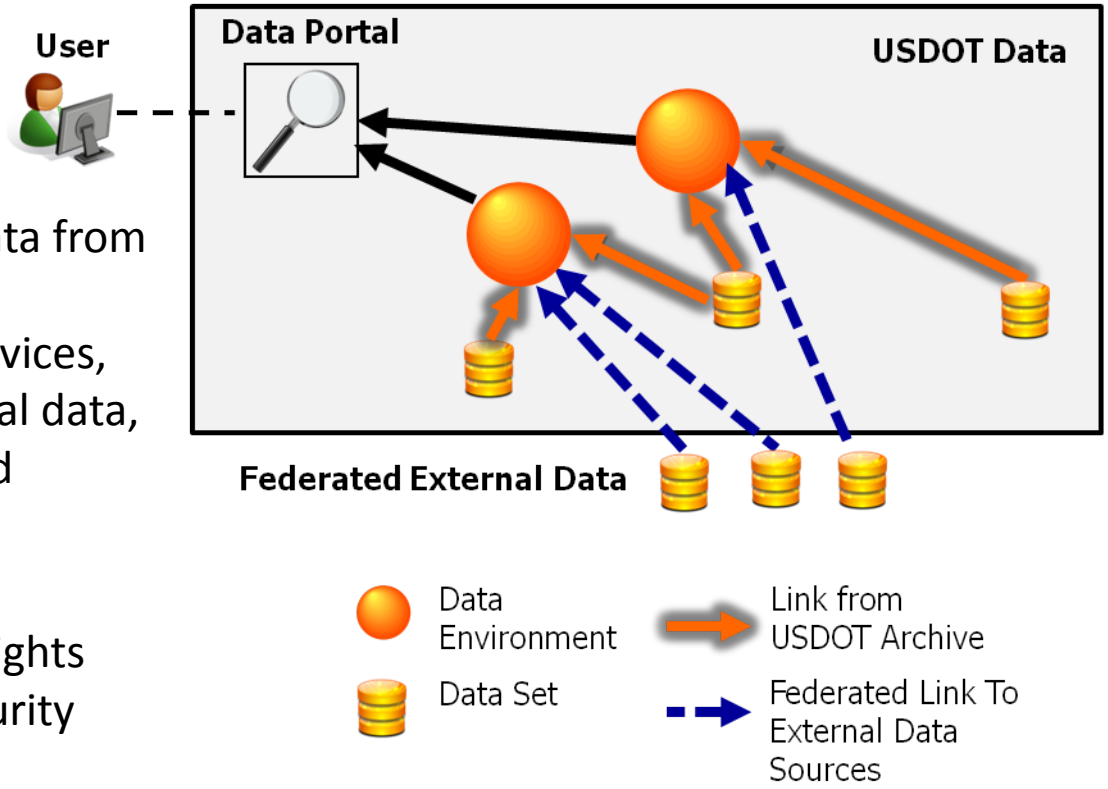
Image Source: http://infreemation.net/cloud-Computing-linear-utility-or-complex-ecosystem

- **Value to DCM**: For RDE, increased flexibility, reduced costs, improved scalability, location independence

- **Challenges**: Security, potential data transfer bottlenecks, data consistency

Data Federation

- Form of data virtualization where data from multiple, heterogeneous, autonomous data sources are made accessible to data consumers as if it is contained in one single relational database, by using on-demand data integration



- **Value to DCM**: incorporates data from many sources without issues associated with centralized services, local data experts maintain local data, brings data to users rapidly and consistently

- **Challenges**: Interfaces and IP rights must be carefully defined, security

Gene McHale
FHWA Operations R&D

Prototype Data Environment: Utilization and Insights

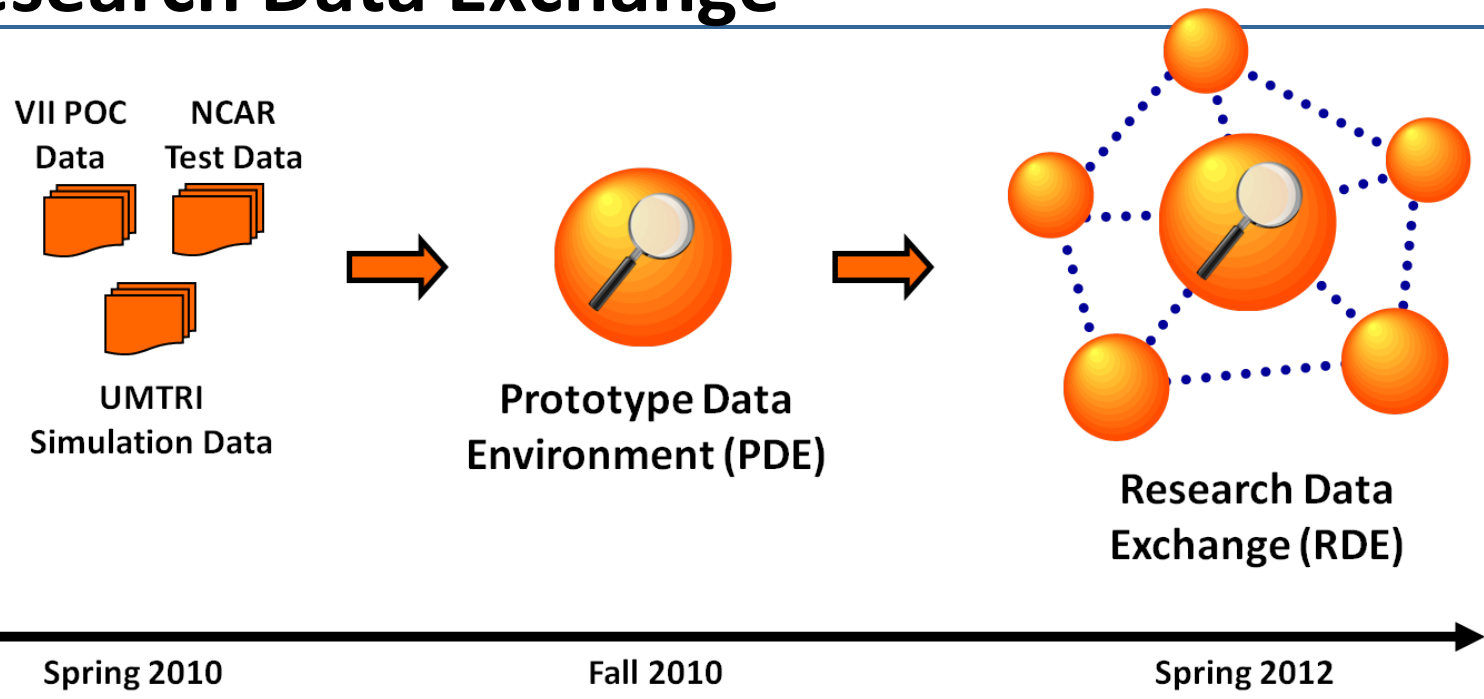
Prototype Data Environment (PDE)

- **Five registered research projects** using PDE 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
  - U. of Washington: Tracking Transit Vehicles in King County

- **2,721 total visits, 1132 unique visitors, 60 countries, 138 registered users**
  - 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 needs 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

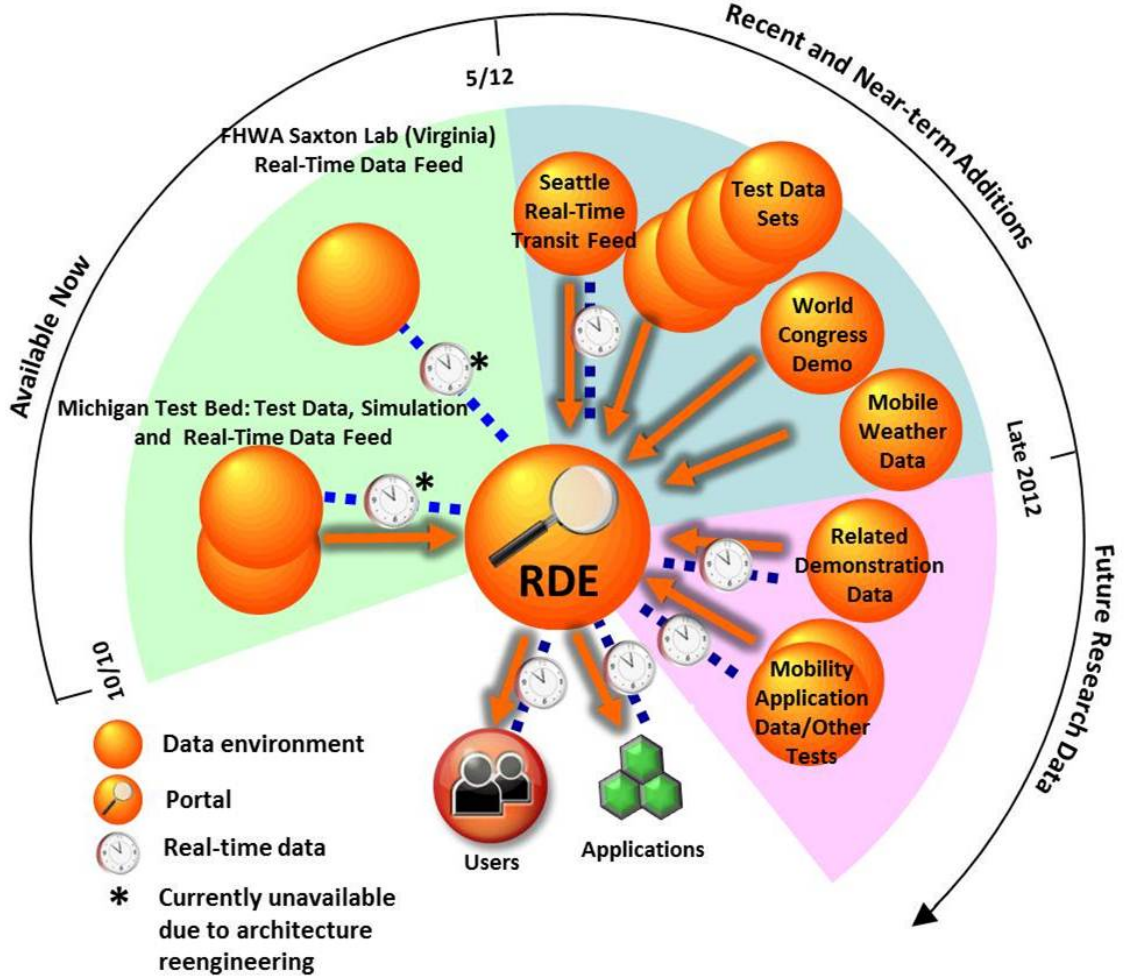


- The Research Data Exchange (RDE) is the connected system of data environments we envision to support application research and development
- The RDE will *not* be a single, centralized repository
  - but rather a *system of systems* linking multiple data management systems
  - some of which will be 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

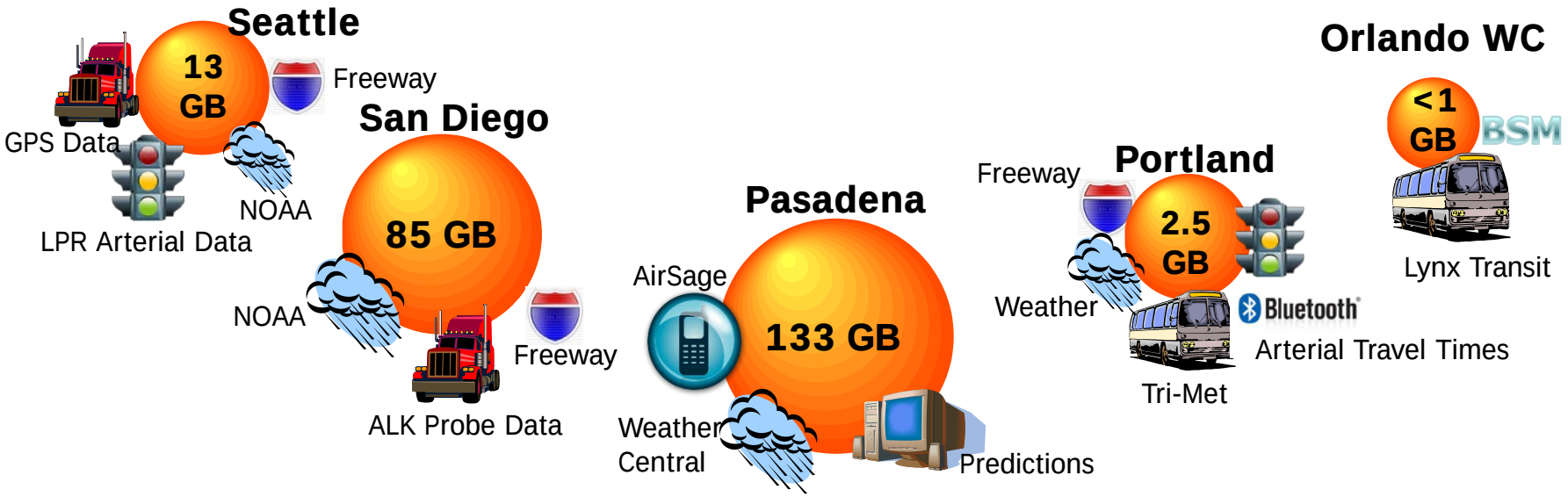
Data Capture and Management: The Road to Deployment

- The Research Data Exchange supports research related to applications enabled by new forms of data
- The RDE does not *itself* represent a prototype operational data environment, however, research supported by the RDE
  - Identifies and characterizes the minimum data set and data characteristics required to realize each application
  - Reveals implications for related standards, IPR, data ownership, and privacy issues
  - Provides lessons learned in terms of balancing data federation and centralization for operational deployments
- Well-formed and described minimum data sets and characteristics can be used to guide the integration of applications into legacy data systems
- In Phase 3 our goal is to demonstrate how new forms of data from wirelessly connected vehicles and data can be incorporated into deployed systems supporting new applications

RDE Data: Current and Near-Term Contents



Potential Research Supported by Near-Term RDE Data Sets



- Primary goal: support development of mobility applications
- The DCM program will investigate the following topics supported by near-term RDE data sets:
  - What are the key differences between current probe data and BSM connected vehicle probe data?
  - How can probe data be used in conjunction with other forms of data to enable new transformative applications?
  - Can multi-modal data be fused and utilized for traveler information and systems management?

Potential Research Topics Using RDE Content

- Quantifying variability of intersection peak hour demands
- Estimating route travel time reliability on arterials
- Analyzing travel time reliability
  - Freight, transit, and integrated multi-modal system perspectives
- Quantifying variability of freeway traffic demands
- Estimating accidents and/or weather impacts on a system
- Verifying accuracy of congestion predictions and accuracy of travel recommendations based on those predictions
- Determining effectiveness of driver information using CMS messages in changing traffic conditions
- Predicting bus delays as a function of prevailing traffic conditions
- Analyzing travel times (delays) and type of incident and incident response for better real-time incident management and travel advisory.

RDE Demonstration

U.S. Department of Transportation
Research Data Exchange

RDE Website: www.its-rde.com

Critical Issues for DCM

- What role can standards play in the RDE?
- How can we manage data privacy?
- How can we deal with intellectual property rights and data ownership issues?
- How do we manage the RDE?
- What are the most valuable data?
- What level of quality is required for the data in the RDE?
- Will other researchers use the RDE?
- What is the value of capturing and making data available?
- What are the most effective ways to provide BSM data to support research?

Next Steps of RDE

- Prototype Research Data Exchange is live
  - Please visit the RDE at: www.its-rde.com
- Stakeholder Engagement Meeting – Summer 2012
- RDE Update 1 – Fall 2012
- Integration with Mobility Applications – TBD

Questions?