

# Project Update Webinar

## Customer Experience

### Projects Covered

- Federated Data Usage Platform
- Models for a Data Concierge Service for a National Secure Data Service

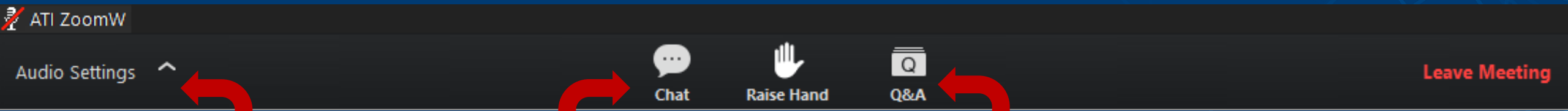


AMERICA'S DATAHUB  
CONSORTIUM



## Housekeeping Items:

- Please stay on mute unless speaking during Q&A.
- Please use the “chat” function for technical difficulties only.
- Place all questions in the Q&A Box.
- Please check your audio settings if you are having difficulties hearing us.



Check **Audio Settings** if you can't hear us

Click **Chat** to ask for help.

Use **Q&A** for direct questions



AMERICA'S DATAHUB  
CONSORTIUM



# Agenda

- Overview of the National Secure Data Service, Dr. Heather Madray
- Overview of Models for a Data Concierge Service, Dr. Heather Madray
- Project Update on Models for a Data Concierge Service, NORC
- Q&A
- Overview of Federated Data Usage Platform, Dr. May Aydin
- Project Update on Federated Data Usage Platform, Mathematica
- Project Update on Federated Data Usage Platform, NORC
- Q&A
- Closing remarks



# Overview of the National Secure Data Service (NSDS)

**Dr. Heather Madray**

**NCSES, Program Director for Data Access, Confidentiality, and Quality Assessment (DACQA)**

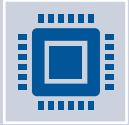


AMERICA'S DATAHUB  
CONSORTIUM



# CHIPS and Science Act Requirements

---



Section 10375 of the 2022 CHIPS and Science Act calls for a 5-year demonstration project to develop, refine, and test models to inform the full implementation of a National Secure Data Service (NSDS).



The NSDS is envisioned as set of shared services and a government-wide data linkage and access infrastructure to support evidence building.



CHIPS and Science calls for consultation with the director of OMB, the National Artificial Intelligence Research Resource (NAIRR), and alignment with the Advisory Council on Data for Evidence Building (ACDEB) recommendations.



The NSDS Demonstration will be implemented by the National Center for Science and Engineering Statistics (NCSES).

# Why an NSDS?

---

Novice and non-expert data users have difficulty navigating the complex data ecosystem.

Data users don't have a place to ask open-ended questions and learn about data options based on their topic of interest.

Accessing data is burdensome, time-consuming, and often expensive.

Linking data requires lengthy processes to determine data ownership, requirements, and limitations on use.



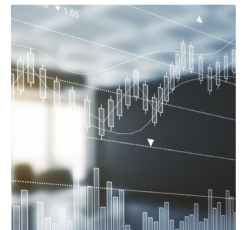
Health data

Climate data



Economic data

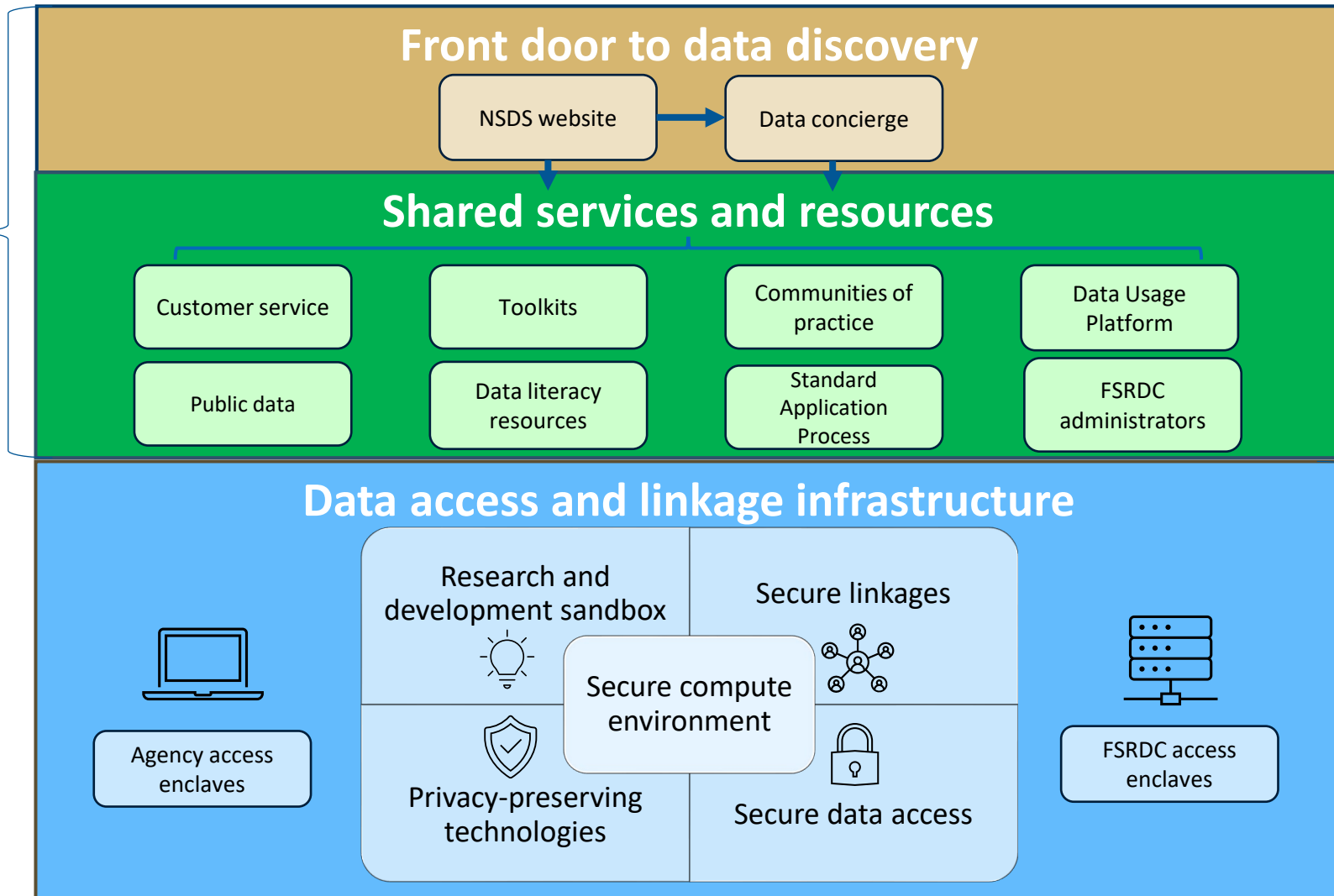
And more...



# What could an NSDS look like?

## Anyone can access

Users can navigate the website on their own to discover services or public data. Users can also engage the data concierge if they aren't sure where to start or have questions.



## Requires secure access

The data concierge guides users to the appropriate secure access modality, directs them to attain needed security credentials, and helps initiate linkages or leverage tools.

# Questions an NSDS can help answer



How much federal funding is given to minority-serving institutions?



An NSDS data concierge can guide users to public data.



How has government data on workforce education been used?



An NSDS data usage platform identifies published products using government data assets.



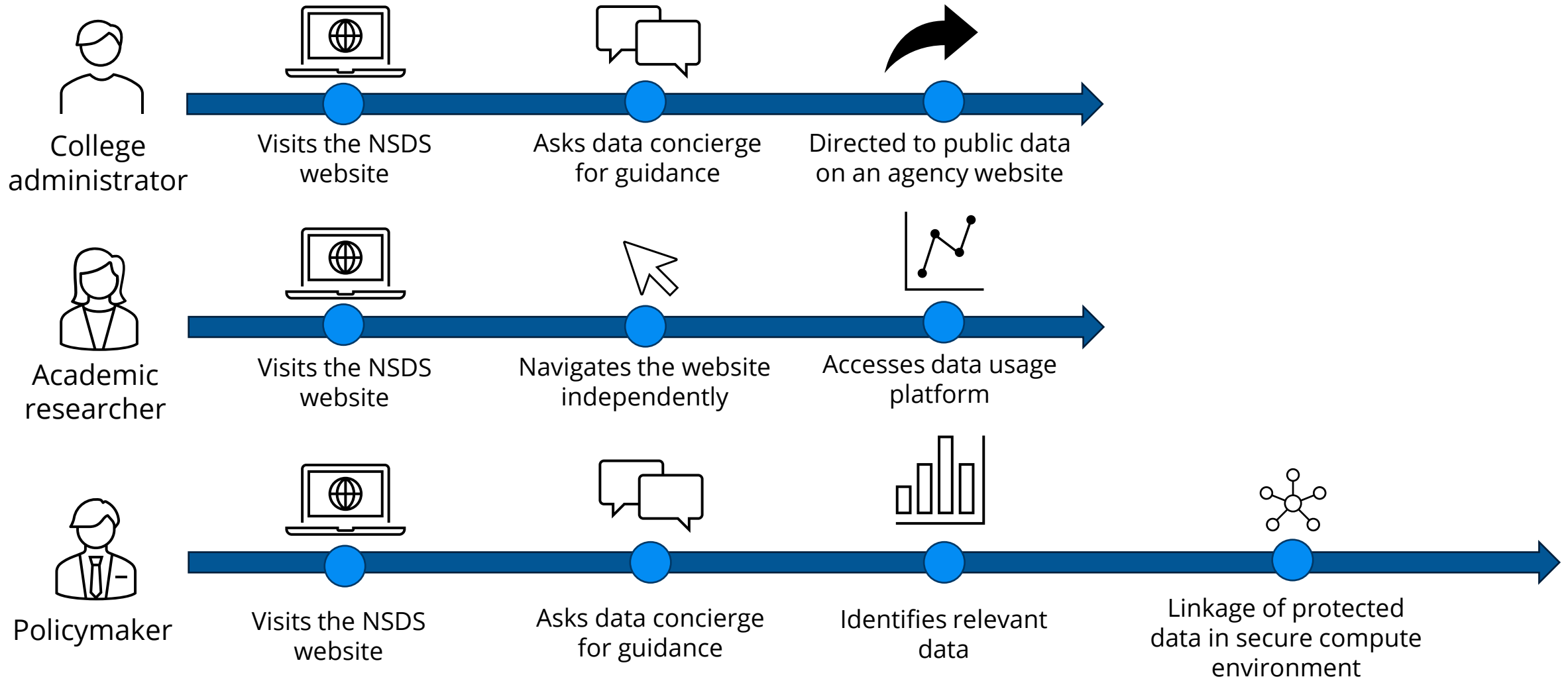
How are health outcomes influenced by access to transportation services?



An NSDS can securely link data across two federal agencies (e.g., transportation and health data).



# NSDS User Journey Examples



# Project Update: Models for a Data Concierge Service

**Dr. Heather Madray**

**NCSES, Program Director for Data Access, Confidentiality, and Quality Assessment (DACQA)**



AMERICA'S DATAHUB  
CONSORTIUM



# Models for a Data Concierge Service

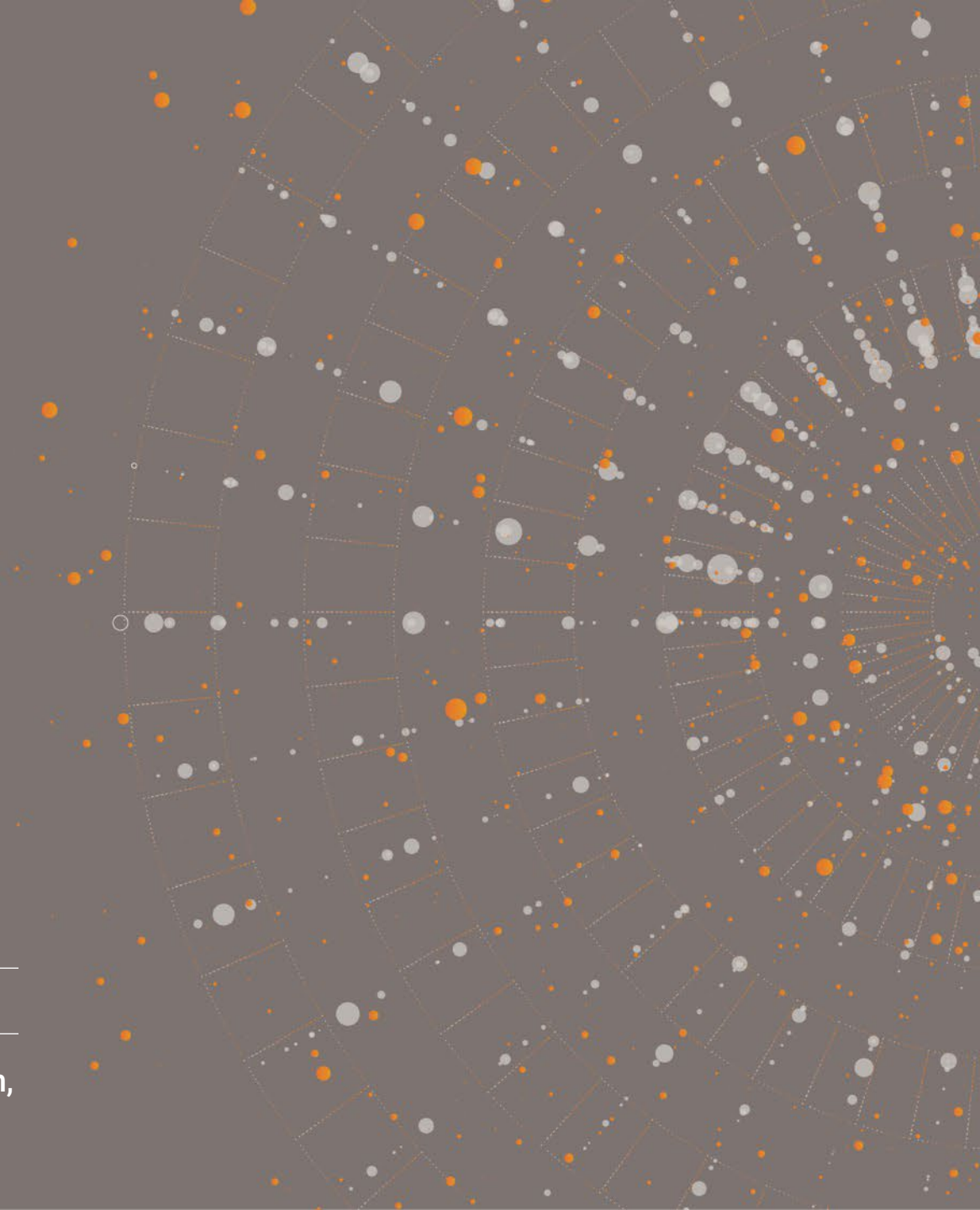
America's Datahub Consortium (ADC) Webinar

---

August 27, 2024

---

Team: Seth Brohinsky, Mike Berning, Brandon Sepulvado, Martha Stapleton,  
Sara Lafia



# Agenda

---

**01** Project Background

---

**02** Interim Findings

---

**03** Lessons Learned

---

**04** Next Steps

---



---

# Project Background



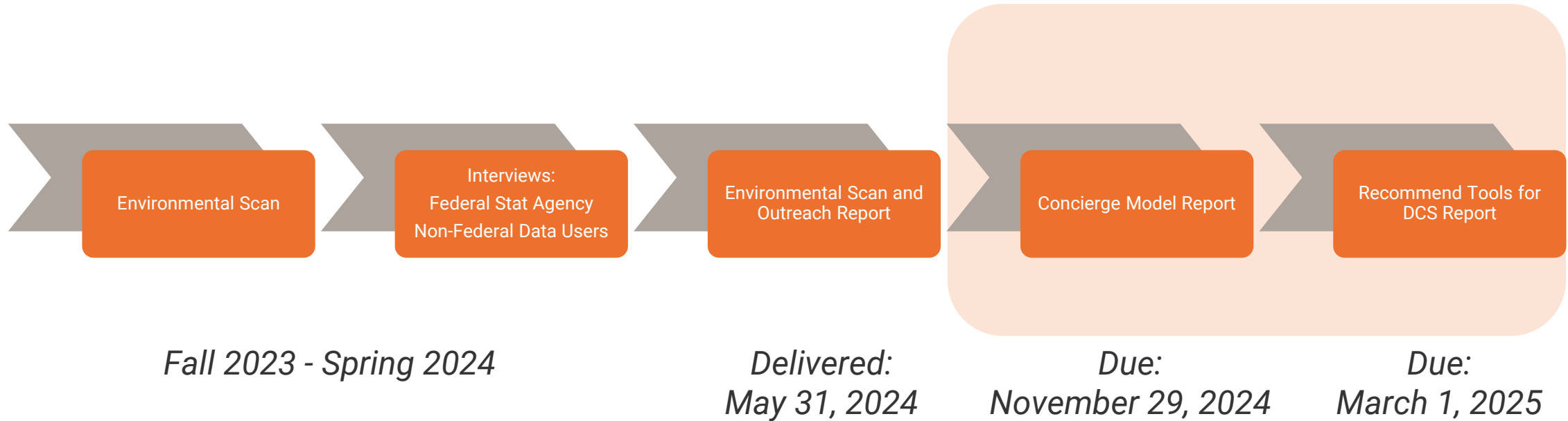
## Objective

Develop models for data concierge services that will provide users with technical assistance for data discovery in support of evidence-based research

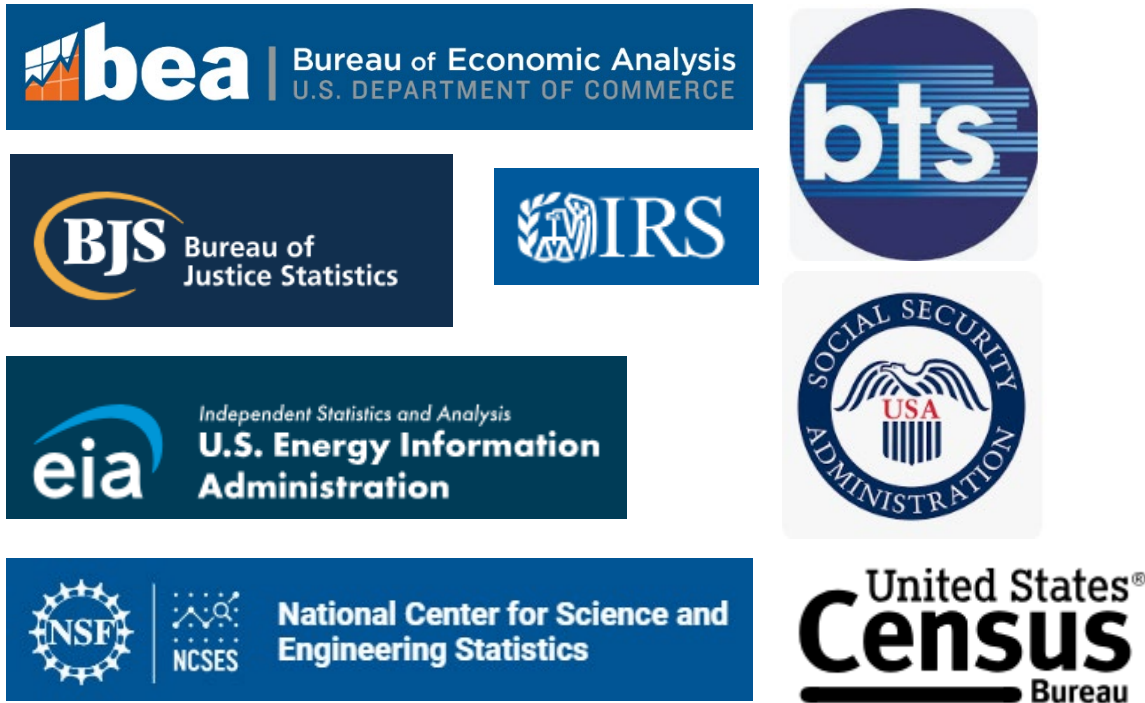
## Activities

- Environmental scan of existing data infrastructure providers
- Interviews with federal agencies and data users
- Model and tool recommendations

# Timeline



26 completed interviews (9 federal statistical agencies and 17 data users)



*Examples of Federal Statistical Agencies interviewed*

<i>Data Users</i>	<b>Interviews</b>
Research/Non-Profit Organizations	7
Minority Serving Community Advocacy Organizations	2
State/Local Government	6
Economic Development Organizations	2

*Summary of data users interviewed by type*



---

# Interim Findings

# Existing examples of data concierge services from environmental scan



Access **federal data** to address state policy maker and federal employee needs



**Compile** and make data available

Support data **applications**

Conduct **trainings**



Assist state agencies with population estimates using **public data**



Provide opportunities for **researchers** to analyze restricted-use statistical data



## Topics covered in interviews with data users and data providers

### Data discovery

Determining fitness for use

Availability of metadata

Online inventories and portals



### Data access

Restricted data

Gaining access

Securing legal/policy authority



### Data usage

Authority to publish

Ability to link data and use link keys

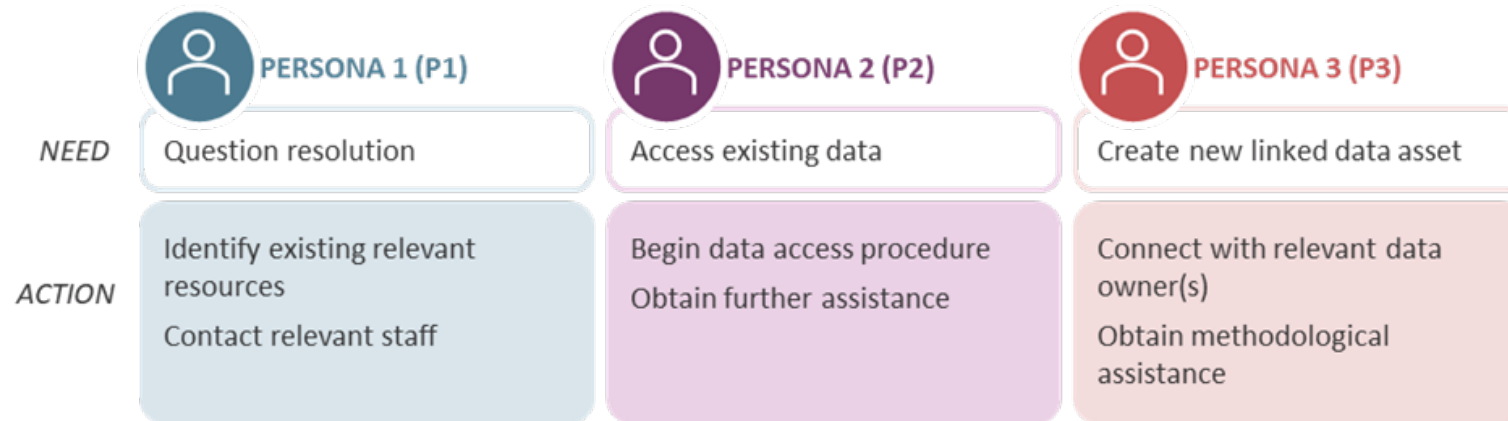
Determining quality of linkage



# Challenges faced by data users and data providers

**Search:** *“I know what variables I’m interested in, but I… feel like my searches are incomplete, or that I’m looking in the wrong place for the wrong thing”*

**Fitness:** *“I don’t understand what data contain or how they can be used without documentation… I usually need to download and explore data if possible”*

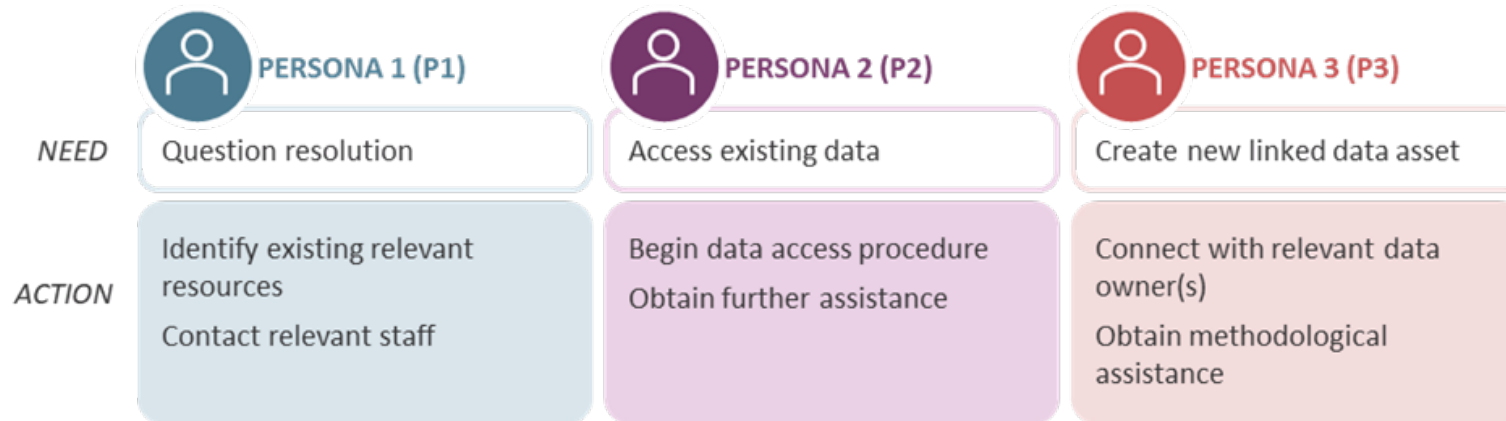


Note: Generalized personas derived from ACDEB Year 2 report. Quotes paraphrased from interviews.

# Challenges faced by data users and data providers

**Navigating restricted access:** “Without an agency partner, it can be difficult to qualify for access to restricted data”

**Meeting training and credentialing requirements:** “The terms of CIPSEA do not fully authorize data access”

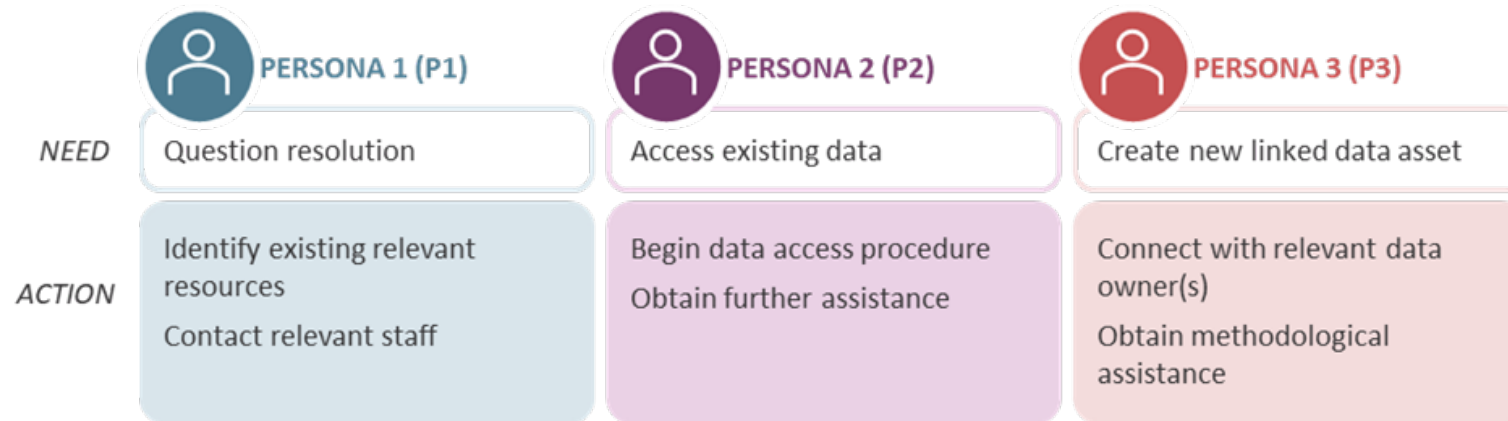


Note: Generalized personas derived from ACDEB Year 2 report. Quotes paraphrased from interviews.

# Challenges faced by data users and data providers

**Validating statistical analyses:** “I find myself wondering if others have done a similar analysis with this data and whether I’m on the right track”

**Linking data:** “I understand the potential, but I’m not sure which data are eligible, how to set up a project, or how to evaluate the quality of links for my needs”



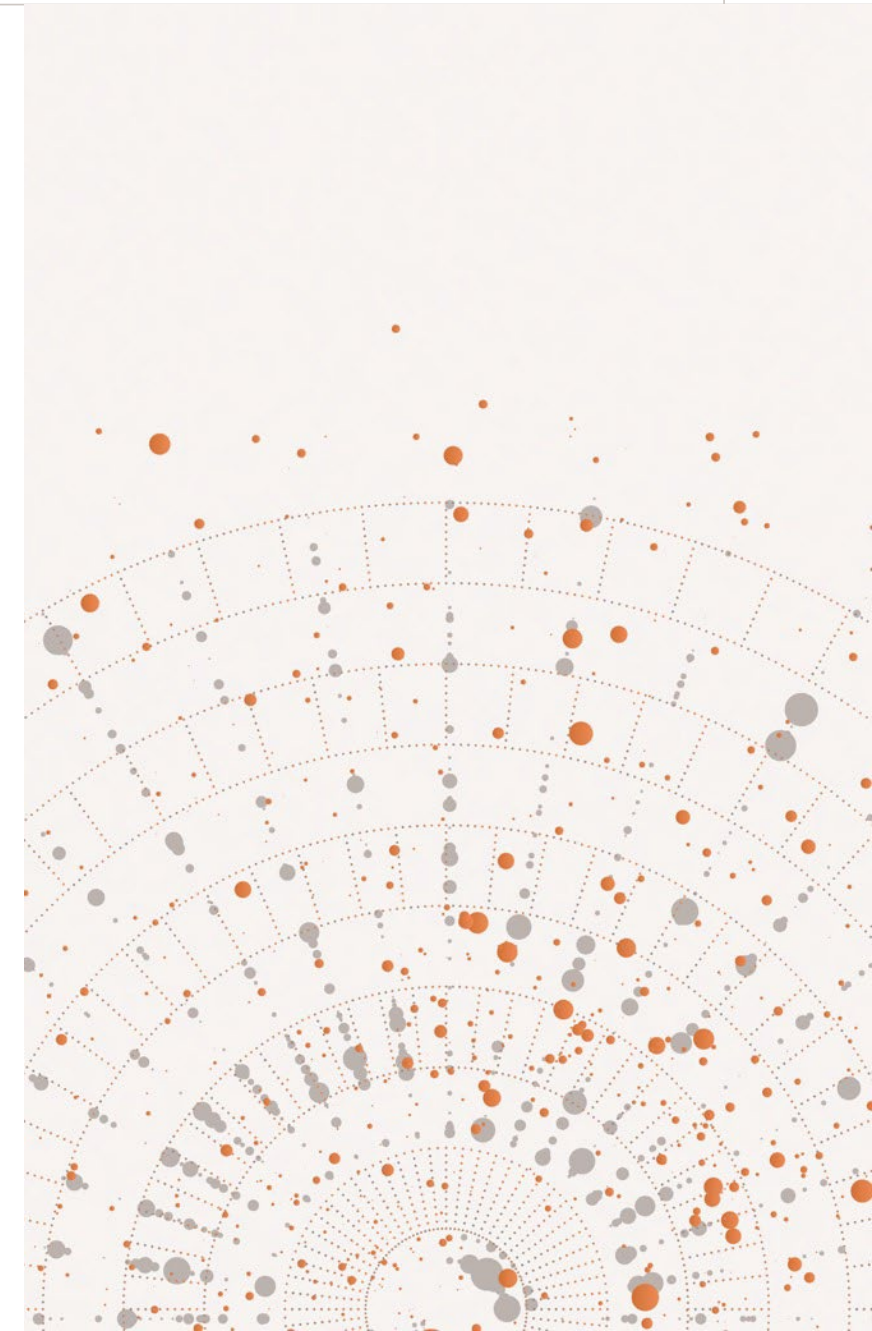
Note: Generalized personas derived from ACDEB Year 2 report. Quotes paraphrased from interviews.

---

# Lessons Learned

## High-level findings

- **Data services** provided by statistical agencies vary in scale and scope, and are not standardized
- **Data accessibility** is limited by legal and policy restrictions
- **State and tribal data users** are often not well-resourced to discover and access data
- **Experienced data users**, such as academic researchers, rely on “insider” agency contacts to discover and access data
- **Less experienced data users**, such as students, have more difficulty identifying and accessing data
- **Enhanced data discovery and access tools** would benefit all users, from new users to “power users”





## Potential solutions – Core functionality of a DCS

### Data Discovery Services

- Centralized assistance for data access
- Chatbot for general inquiries

### Data Access Services

- Centralized assistance for navigating legal requirements
- Anonymized queries on restricted data

### Data Use Services

- Statistical expert consultations
- Library of data use best practices



## Potential solutions – Additional services

### Guide cross-agency requests

- Enhance documentation to support complex searches
- Provide wayfinding guidance across agency resources

### Build on success of the SAP

- Offer SAP adopters expanded capabilities
- Enable application tracking and improvements to metadata

### Establish standardized tools

- Expand analytic capabilities through shared services (PPRL)
- Broaden data accessibility by enabling tiered access



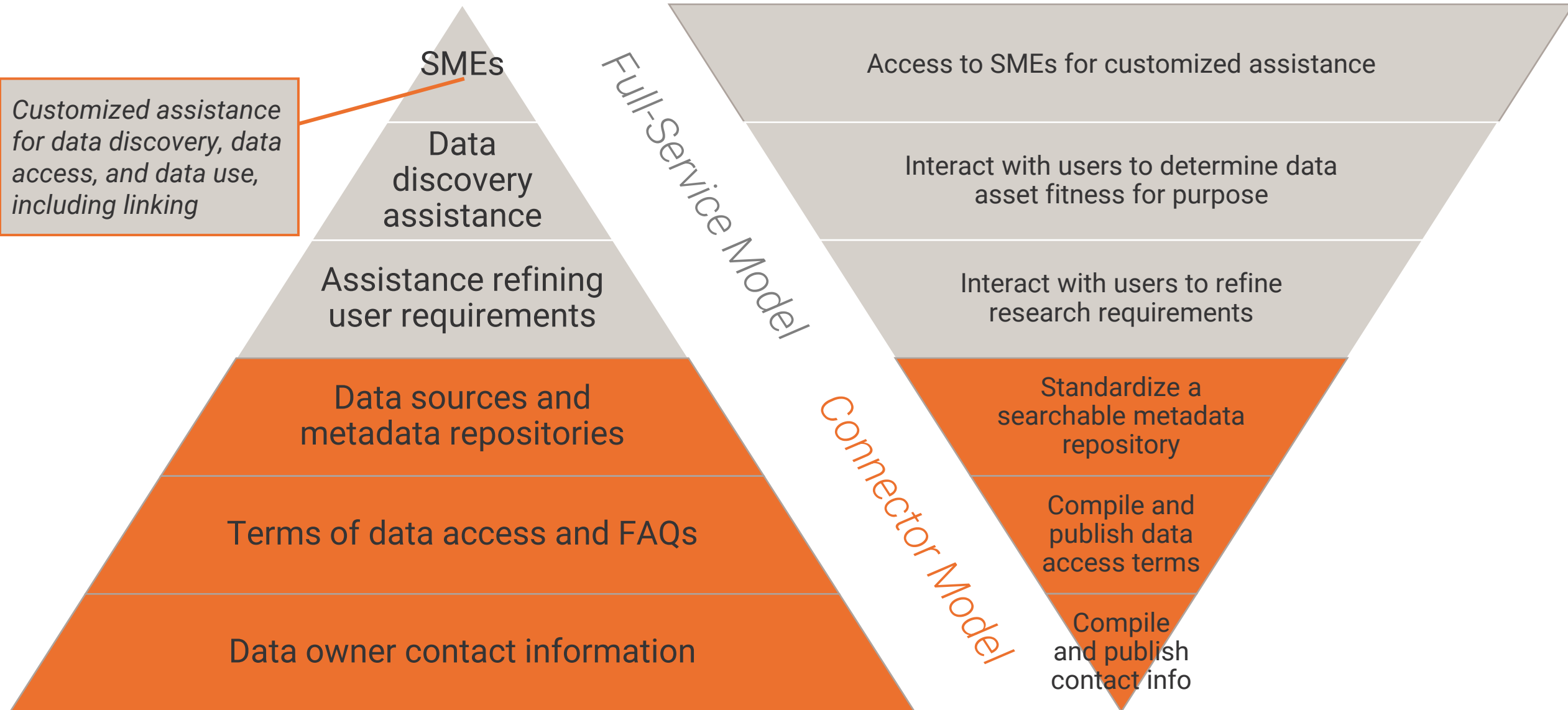
---

# Next Steps

## Concierge model components

## Resource requirements

*Customized assistance for data discovery, data access, and data use, including linking*



# Thank you.

---

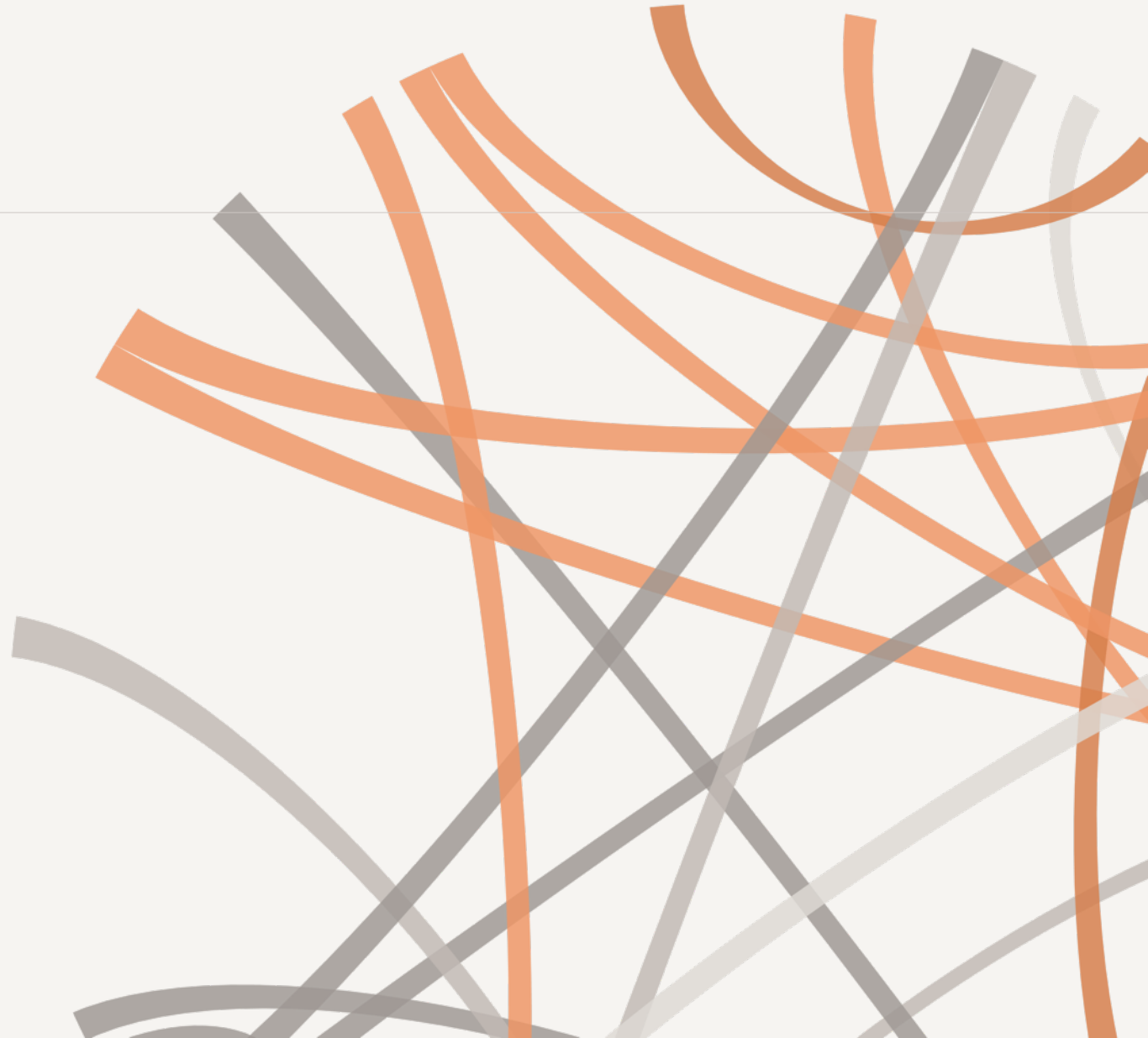
 Research You Can Trust™

---

 **NORC** at the  
University of  
Chicago

---

Questions?



# Project Update: Federated Data Usage Platform

Dr. May Aydin  
NCSES, Science Advisor



AMERICA'S DATAHUB  
CONSORTIUM



# Understanding Federal Data Usage as a shared service within the NSDS and the greater federal data ecosystem...

- Enables a variety of data user communities to connect and partner on research, analysis and application of the data
- Builds a data profile of the agency and system for decision-making
- Supports transparency objectives for improved data access and use
- Enhanced visibility of data application builds public trust in official statistics





# Considerations for exploring the DUP as a shared service

- Requires feasibility and sustainability as a federated resource
- Federal data agencies (not just statistical agencies) can leverage to build awareness of user communities and non-user communities.
- Emphasis on standardization for agencies and end-users
- Serve a wide variety of end-user needs, including researchers, policymakers, and the media
- Build on successes and lessons learned of the Democratizing Data Initiative



# Federated Data Usage Platform (DUP)

August 27, 2024





# Agenda

- **Mathematica's Approach to Federated Data Usage Platform (DUP) Development**
- **DUP: Phase 1- User Research**
  - User Research Participants
  - Key Findings on usage of DUP
- **Phase 2: Prototype Development**
  - Meeting DUP User Needs
  - Prototype Development - Methodology
  - Milestones / Timelines
- **Q & A**

# Overview of Federated DUP



# Mathematica's Approach to Federated DUP Development

## Phase 1

Conduct User Research, perform gap analysis of the existing platform and recommendations

## Phase 2

Develop a prototype to understand the value of the federal data to the public audience

## Phase 3

Develop communication strategies for the broader federal data ecosystem, with specific relevance to the NSDS



**Outcomes:** Insights from this project inform the design of future services to help federal policymakers access and use data through a National Secure Data Service to drive decision-making

# DUP: Phase 1 – User Research



# DUP: User Research Participants

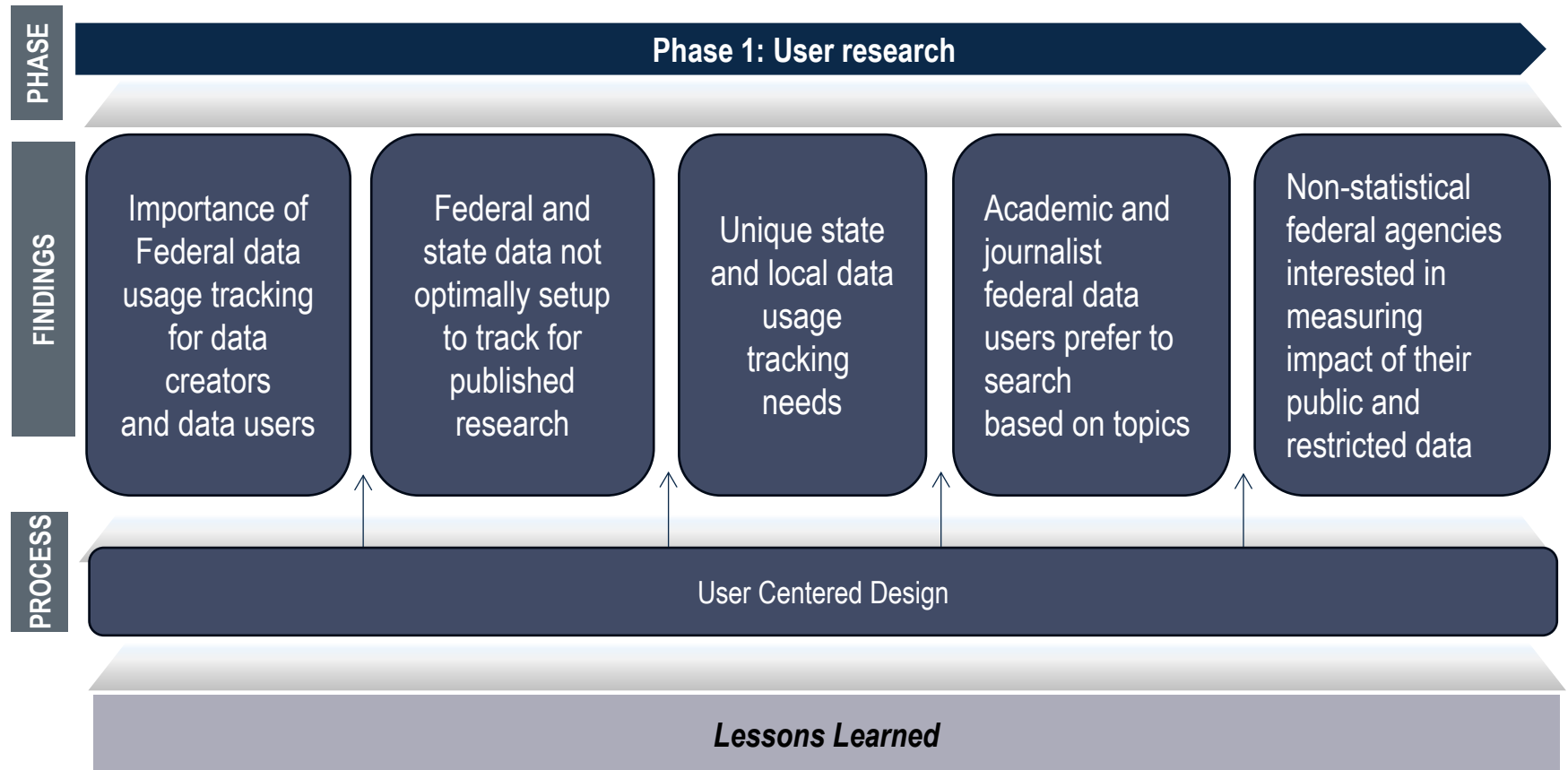
## / Participant Groups

- ❖ Federal and State Agency CDOs, Analytics and Leadership staff
- ❖ IT & Technology
- ❖ State and local agency staff
- ❖ State-run Federal Program Managers
- ❖ Journalists & State policymakers
- ❖ Academics & Researchers





# Key Findings on usage of DUP





# DUP: Phase 2 – Prototype Development



# Meeting DUP User Needs At a Glance

## / Federated DUP

- ❖ Display data usage statistics by topic, agency
- ❖ Measure data usage by summary or details
- ❖ Integrate "impact" measurement
- ❖ Allow data creators for edit/update assets
- ❖ Track usage data by location, where possible
- ❖ Track usage data trends, where possible

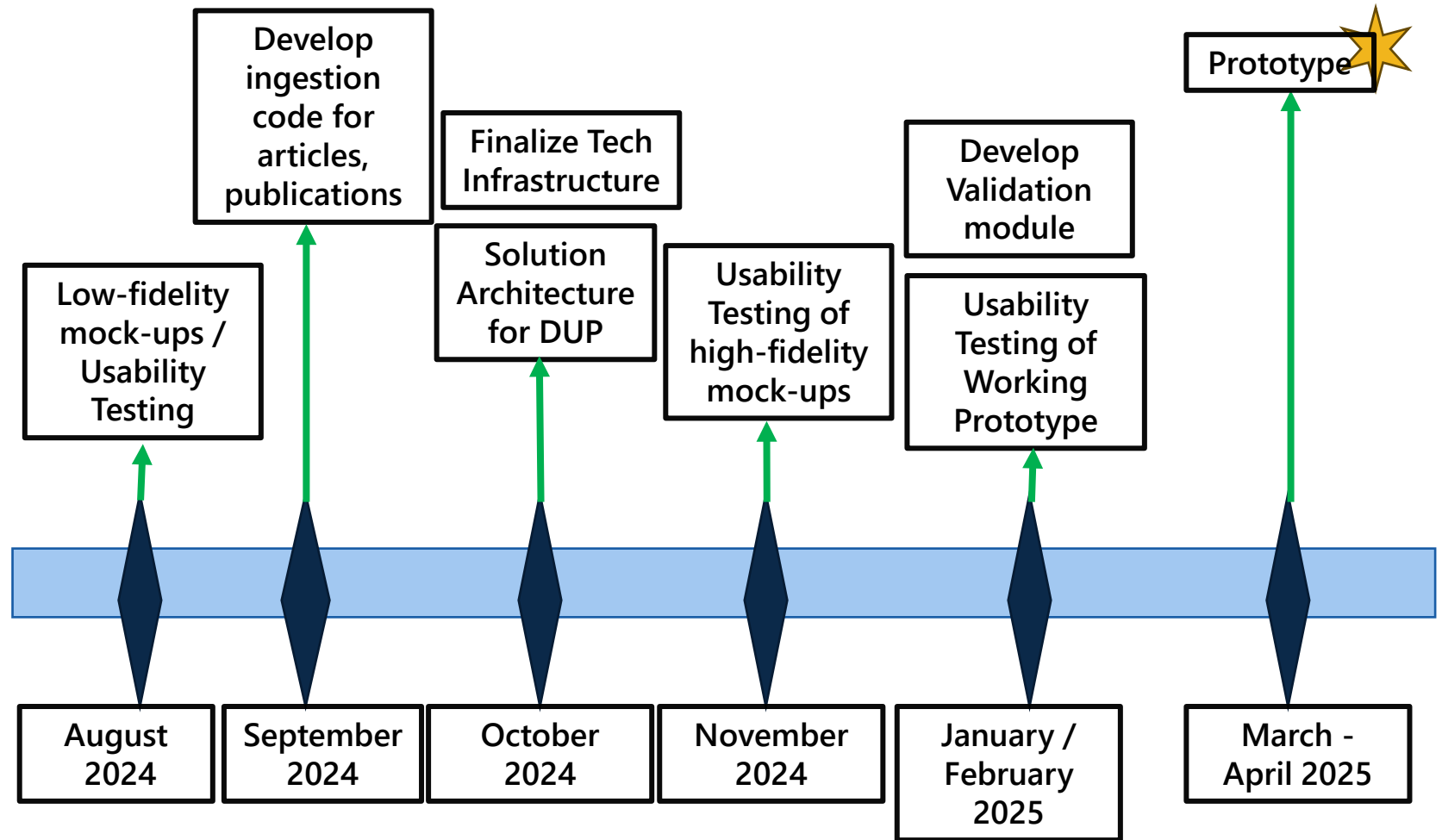


# Prototype Development: Methodology

Tasks	Activities Description
<b>Workstream Planning</b>	<ul style="list-style-type: none"><li>• Data ingestion pipeline</li><li>• Data visualization, design, and testing (and DUP infrastructure)</li></ul>
<b>Development Process</b>	<ul style="list-style-type: none"><li>• Iterative design and development</li><li>• Usability testing as part of user-centered design</li></ul>
<b>Tech Infrastructure Set-up</b>	<p><i>Cloud based architecture and open-source solutions</i></p> <ul style="list-style-type: none"><li>• Data collection: API ingestion, web scraping, bulk data downloads<ul style="list-style-type: none"><li>○ Data validation checks for missing &amp; malformed data during ingestion</li></ul></li><li>• Front end architecture: Dashboard (Drupal, and visualization)</li><li>• Back-end architecture: Structured database</li></ul>
<b>Final Product</b>	<ul style="list-style-type: none"><li>• Working prototype and code/documentation</li></ul>



# Prototype: Milestones / Timelines





# Q & A





Thank You

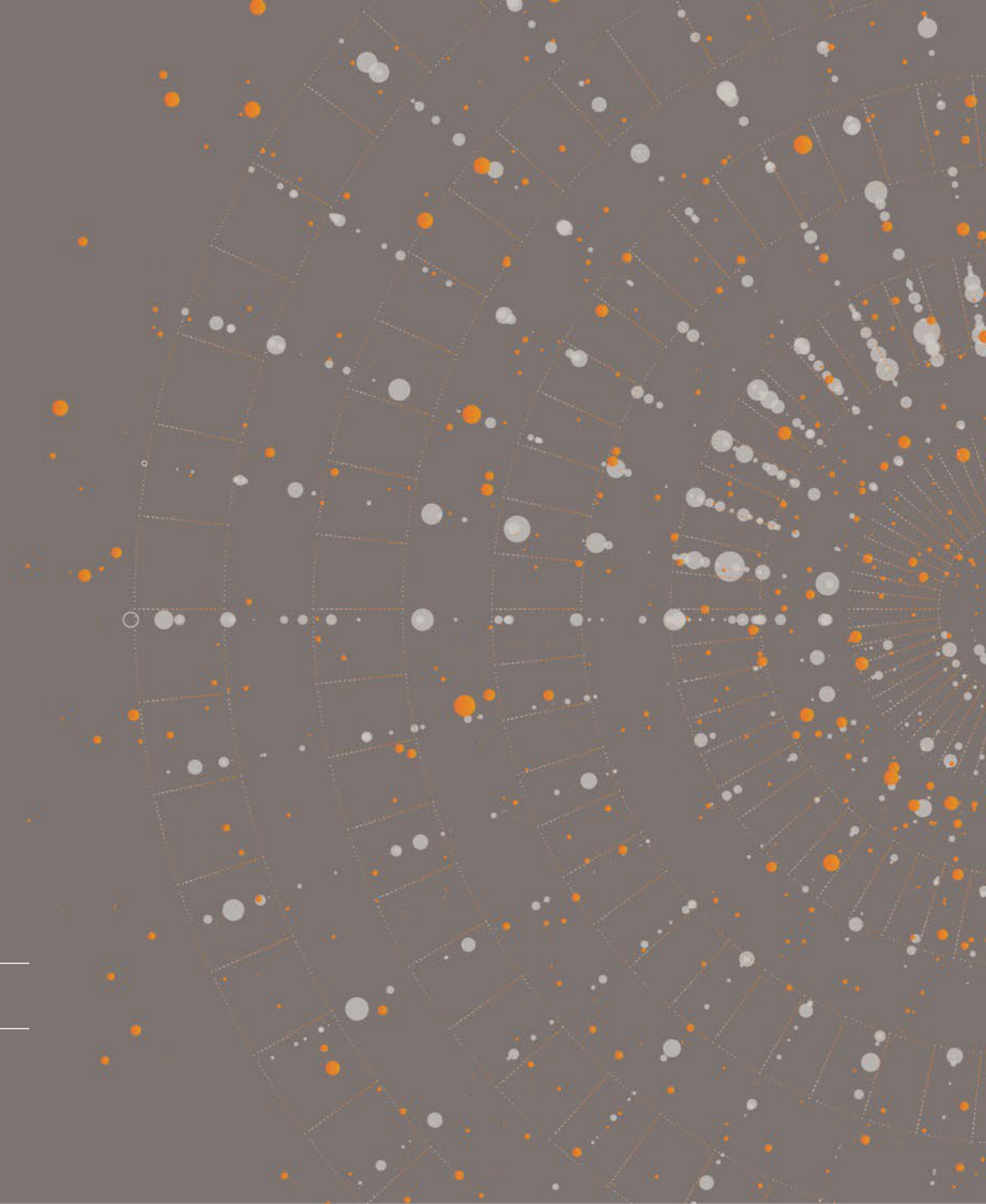
# Project Updates: Federated Data Usage Platform

---

8.27.2024

---

Martha Stapleton



# Agenda

01 Project Timeline

---

02 Interim Project Findings

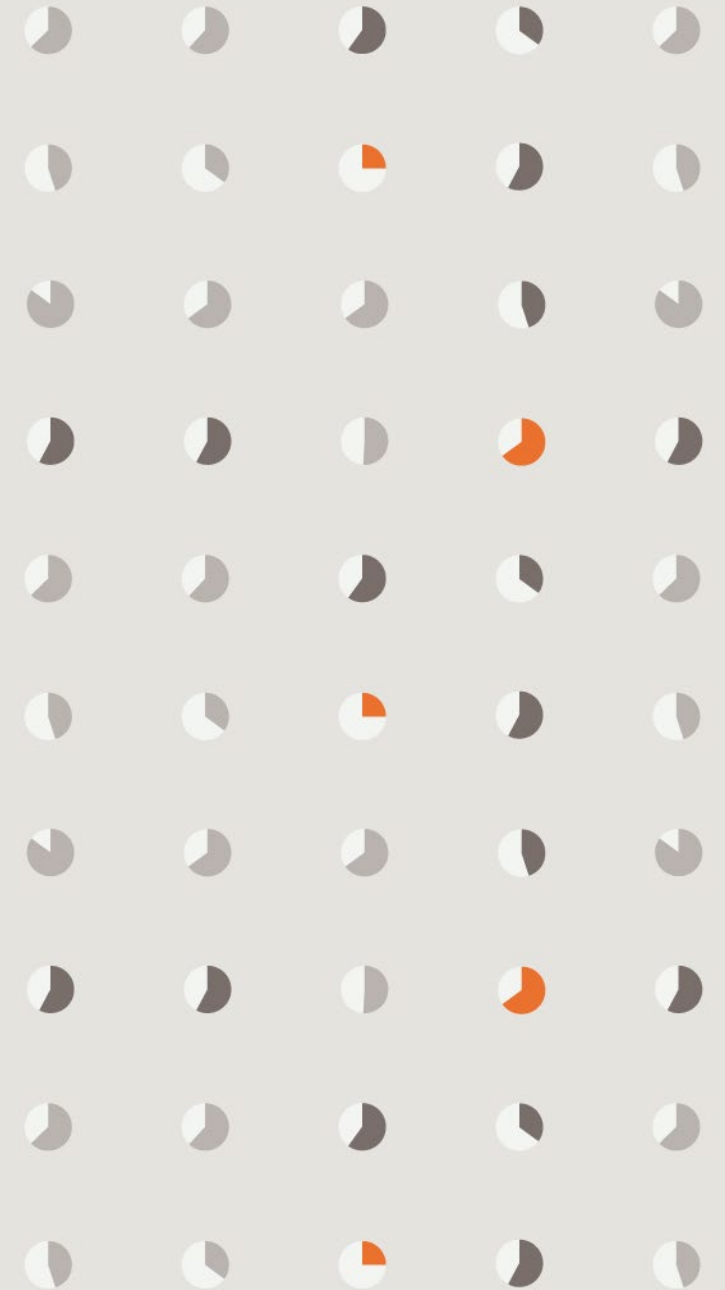
---

03 Lessons Learned

---

04 Next Steps

---

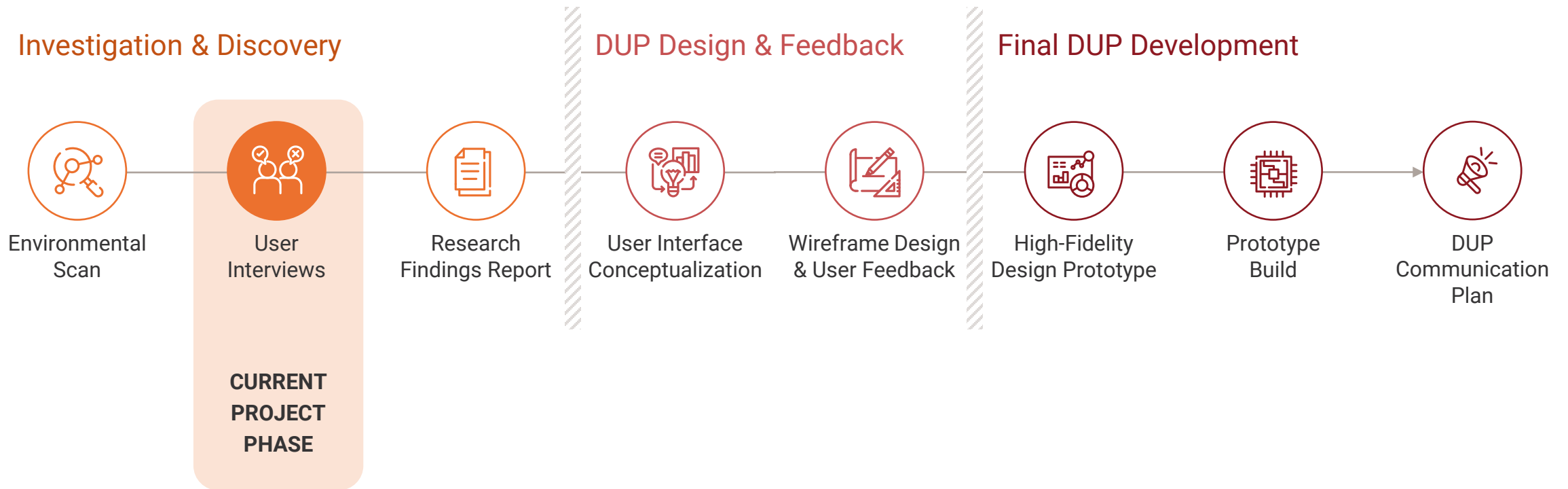




---

# Project Timeline

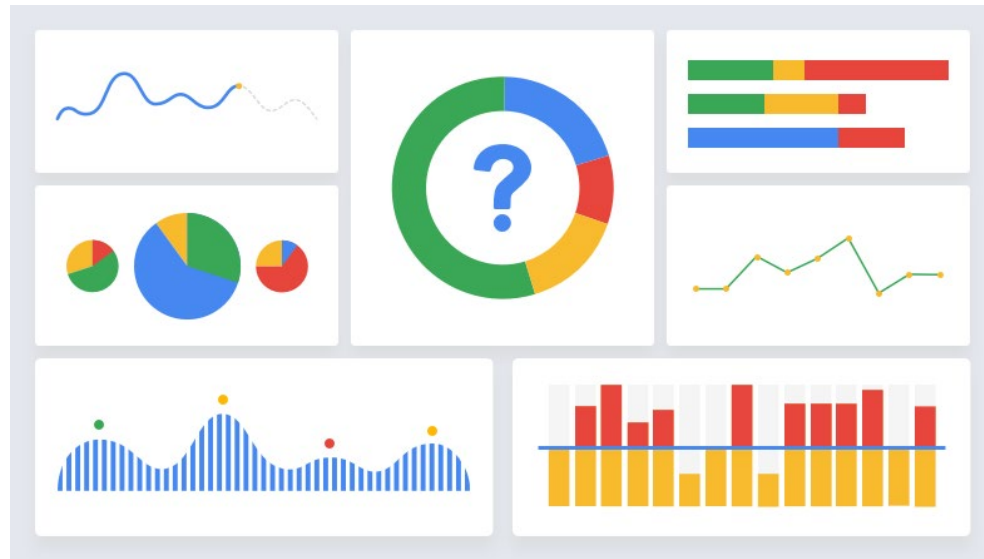
# Federated DUP Project Phases



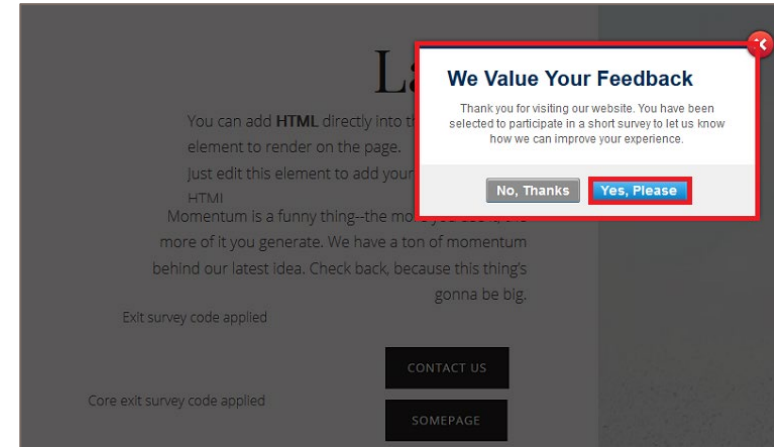
---

# Interim Project Findings

# Methods for tracking public data usage



Google Analytics: views, downloads



Push web surveys: features



Focus groups: usability

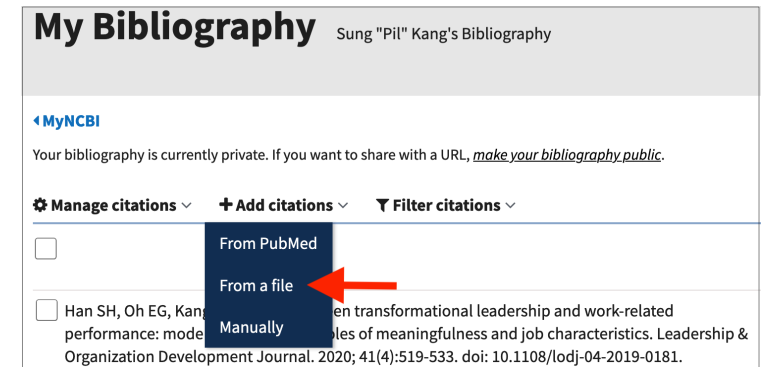
# Methods for tracking restricted data usage



SAP: track applications and engagement

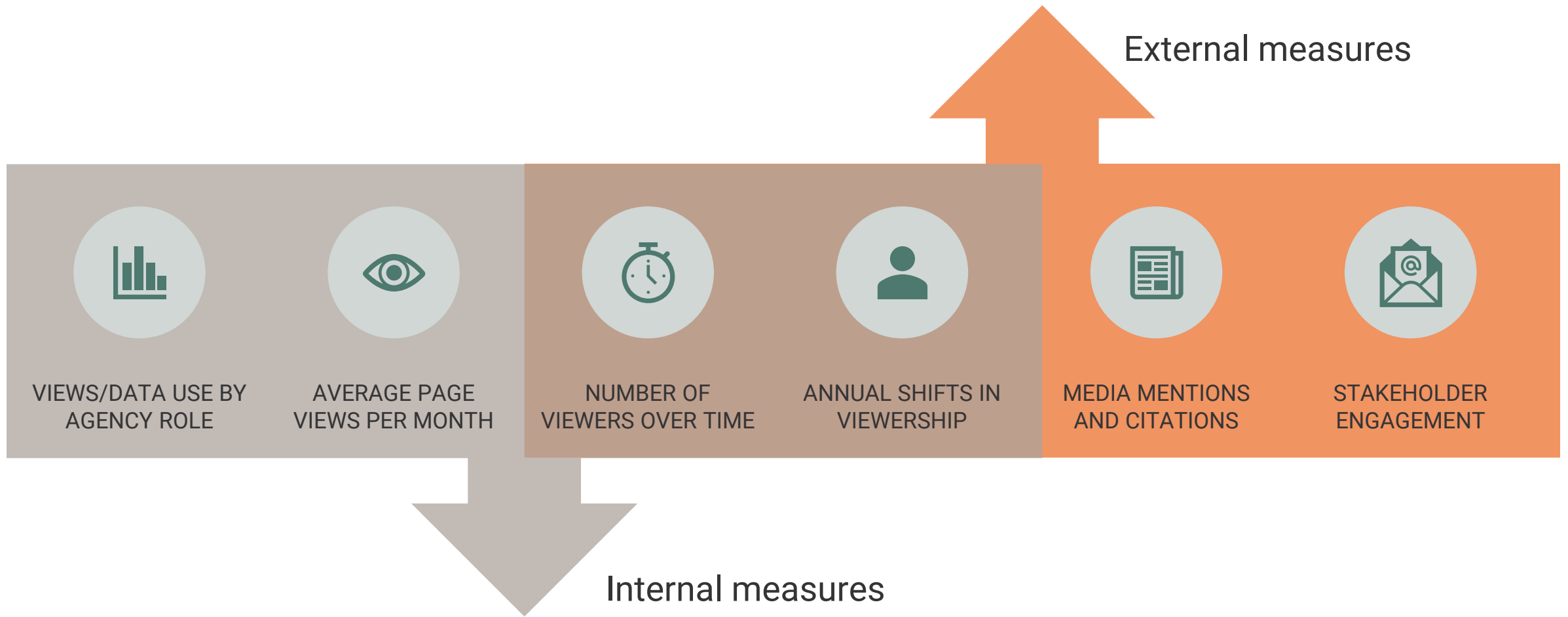


Secure computing: monitor usage



Publication lists: review for disclosure, curate by topic

# Data usage metrics inform agencies' decision-making activities



## Federal agency **perspectives** on tracking data usage

“We don’t want to fall into the trap of valuing things that are easy to measure.”

“We have good metrics on restricted data but not on public use data.”

“Just because a resource isn’t popular doesn’t mean that it’s not important.”



## What agencies want to understand about data use

1.

### **State, local, and tribal use**

*Who uses data?*

2.

### **Supported methods**

*What methods  
do the data  
support?*

3.

### **Tiered access\***

*How does data  
use vary by  
access level?*

4.

### **Decision- making**

*Which data  
are used as  
evidence in  
decisions?*

5.

### **Cross-agency use**


*How are  
data used  
collaboratively?*

\* Tiered access designates modes between restricted and open levels to simultaneously expand data access while protecting data. By tiered access, we mean application of the principle of least privilege (granting access to data assets needed to meet a user's needs) ranging from restricted data only accessible to approved agents to open data with no access controls (ACDEB Year 2 Report: Expanding Secure Access to CIPSEA Data Assets, p. 29)



What might potential shared services look like within the context of a National Secure Data Service?

### Potential Shared services within a future NSDS

- **Data Usage Platform (DUP)**
  - Standard Application Process (SAP) portal
  - Data concierge (DCS), or certain components
  - Noise infusion (SDRN)
  - Privacy preserving record linkage (PPRL)
- 
- Identifying functional needs
  - Establishing proofs of concept for solutions

### From vision to design

- Design considerations for current projects to maximize future utility
- Facilitate the integration of shared services into common NSDS offerings

## Implications for Developing the DUP as a shared service

The shared services model for a potential DUP could inform decision-making at various points in the design process, including:

- Roles and options for dashboard customization for agencies
- Overall look and feel of the platform, including branding
- Long-term support of the platform

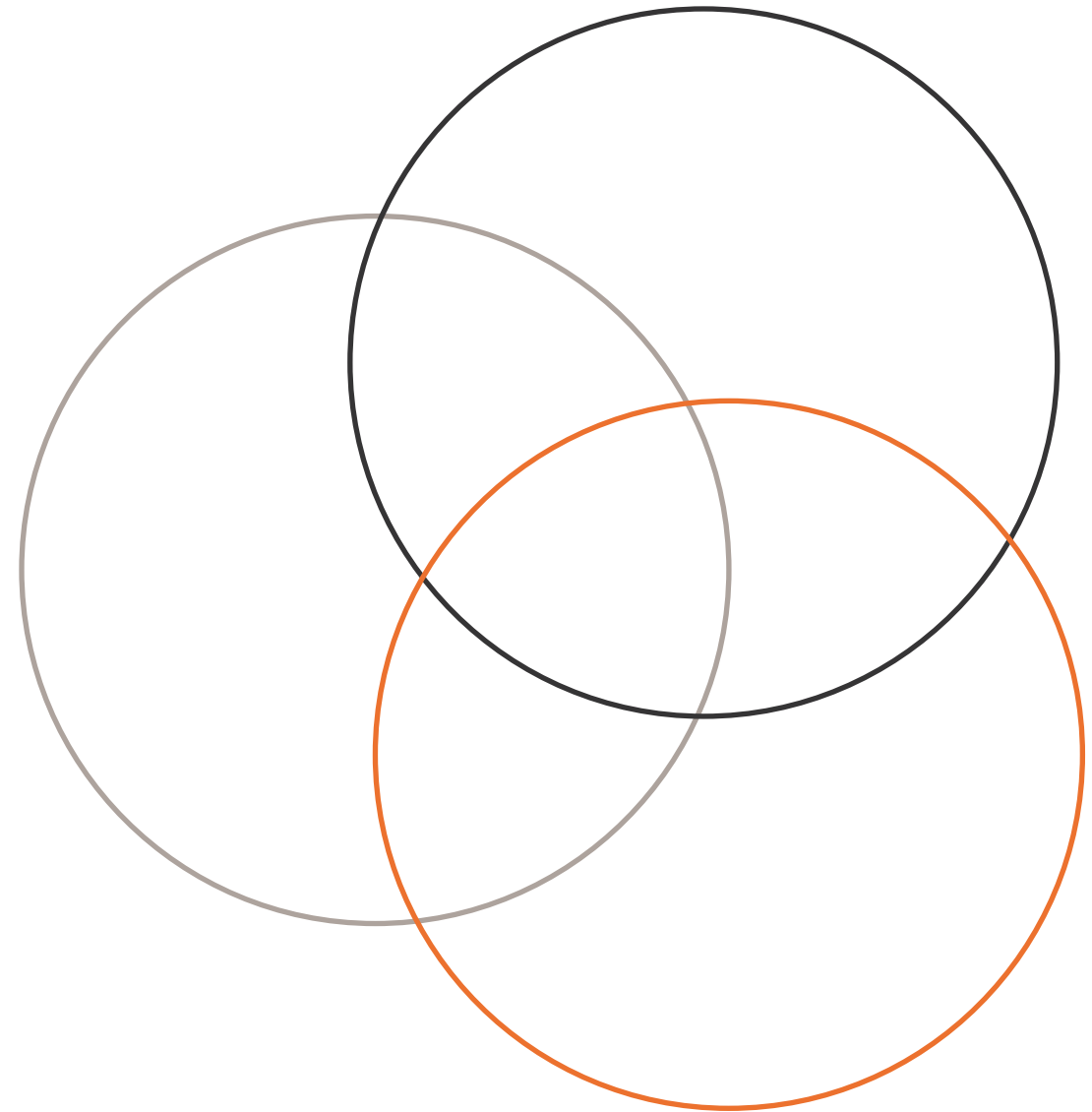
A flexible infrastructure is needed.

### **Centralize where possible**

- Reduce cost
- Increase efficiency

### **Accommodation where needed**

- Budgeting/appropriations considerations
- Agency-specific privacy and data security policies



How might a future, potential DUP be translated into a shared service?

## Defining "shared service"



Who owns and maintains each component?



Where do the DUP components live?



Can agencies control their metrics' visibility?

---

# Lessons Learned

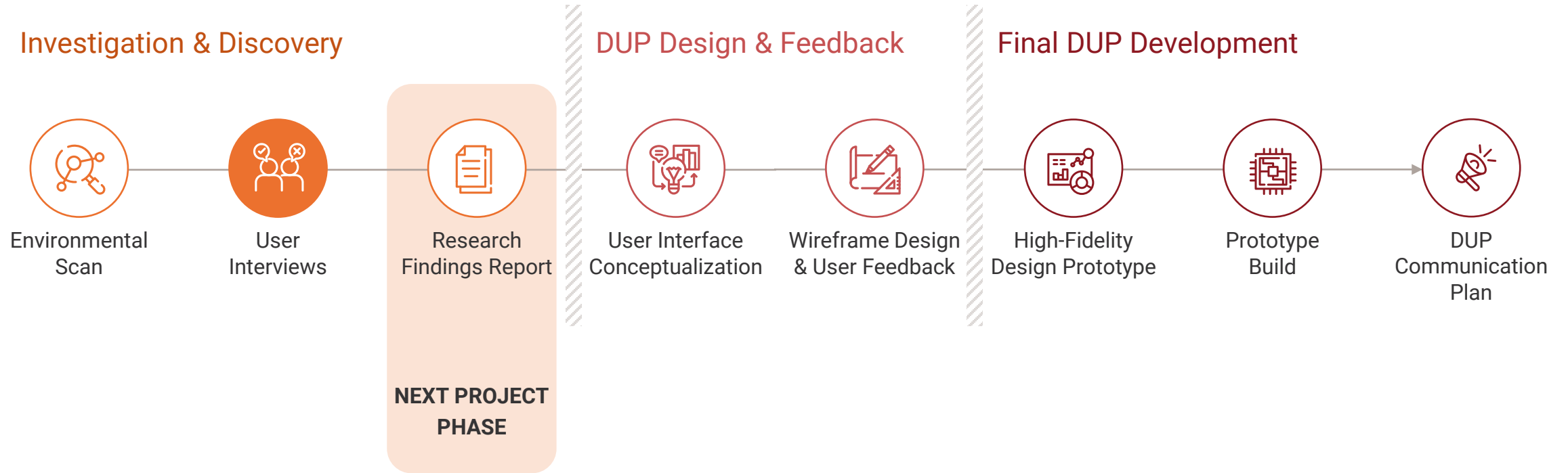
# Lessons Learned

- Interim findings from federal user interviews suggest that many potential platform users are interested in learning more about individual data users, their needs, and how they use agency-level data.
- Many agency staff have expressed a desire for cross-agency collaboration to expand research initiatives and potentially combine data products.
- Agency staff articulated a desire for a DUP to account for diverse needs which vary by agency type as well as a diverse audience of users.

---

# Next Steps

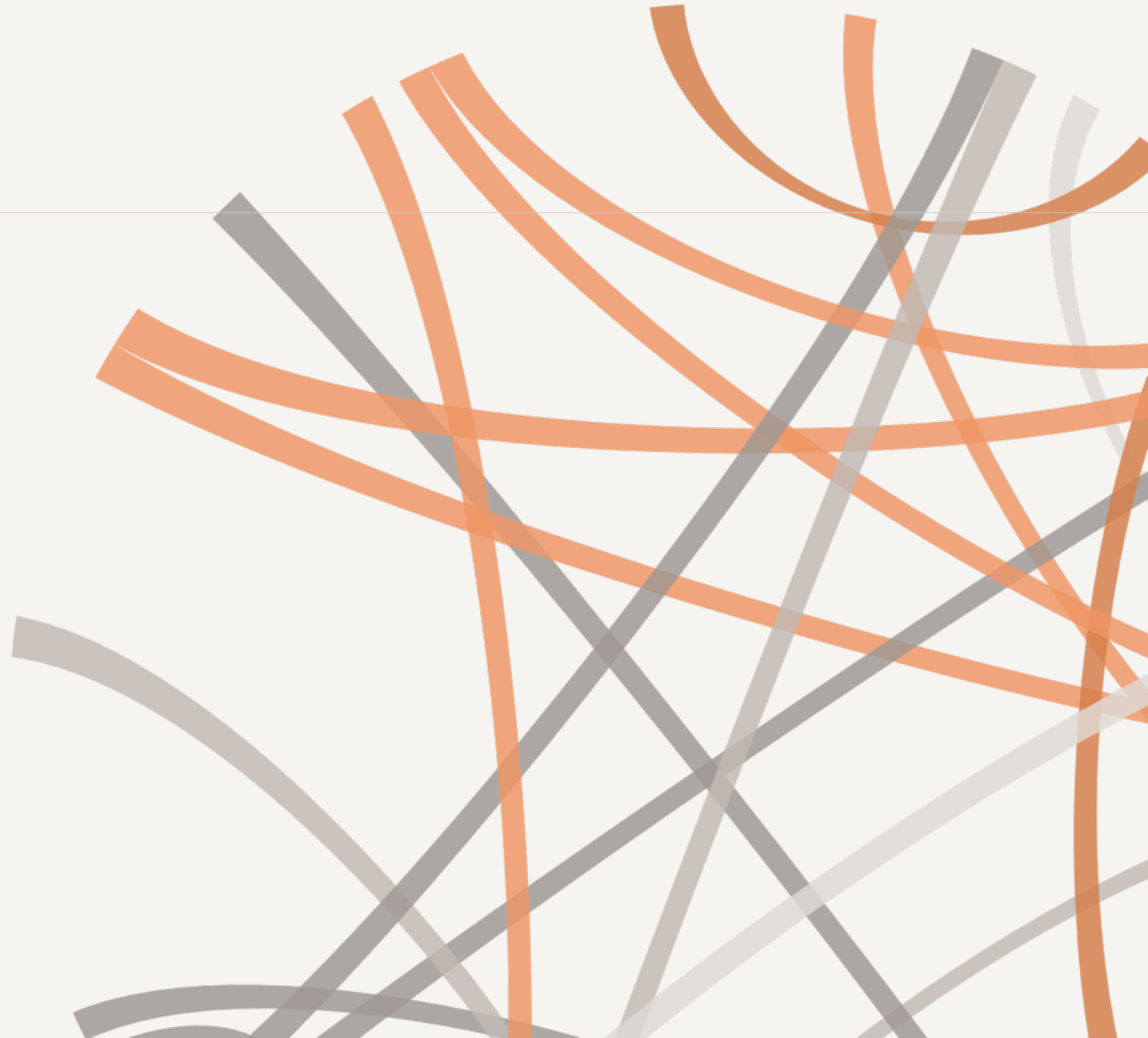
# Next Steps





---

Questions?



# Stay Engaged

✉ Solicitation and Contract Related Questions: [ADC-Contracts@ati.org](mailto:ADC-Contracts@ati.org)

❓ General/Membership Questions: [adc@ati.org](mailto:adc@ati.org)

✉ Join the ADC Mailing List: [www.americasdatahub.org/adc-mailing-request-form/](http://www.americasdatahub.org/adc-mailing-request-form/)

📖 Read Lessons Learned: [www.americasdatahub.org/adc-lessons-learned/](http://www.americasdatahub.org/adc-lessons-learned/)

## Visit the ADC Website

Stay in the loop with ADC events, solicitations, project awards, news, members, and more!

[www.americasdatahub.org](http://www.americasdatahub.org)





## Lessons Learned

Project Name	Project Status	Final Report (if applicable)
ABSSyn-23-N06: Creating and Validating Synthetic Data (NCSES/Census, Annual Business Survey) - Knexus Research Corp	Active	
DCS-23-N03: Models for a Data Concierge Service for a National Secure Data Service - NORC at the University of Chicago	Active	
DPT-23-N001: Data Protection Toolkit Use Case Analysis - NORC at the University of Chicago	Complete	<a href="#">View Final Report</a>
DUP-23-N02: Federated Data Usage Platform - Mathematica, Inc.	Active	
DUP-23-N04: Federated Data Usage Platform - NORC at the University of Chicago	Active	
FBSE-22-05: Foreign Born Scientists and Engineers in the Workforce - Coleridge Initiative	Complete	<a href="#">View Final Report</a>
FBSE-22-08: Foreign Born Scientists and Engineers in the Workforce- NORC at the University of Chicago	Active	
FBSE-22-09: Foreign Born Scientists and Engineers in the Workforce - RTI International	Active	
FSRDC-23-N03: Expanding Equitable Access to Restricted-Use Data through Federal Statistical Research Data Centers - Regents of the University of Michigan	Active	
NVSS-23-N02: National Vital Statistics System Modernization – New Opportunities for Interoperable Data - Clinovations Government + Health	Active	



# Thank you!



AMERICA'S DATAHUB  
CONSORTIUM

