Project Update Webinar
Capacity Building

Projects Covered

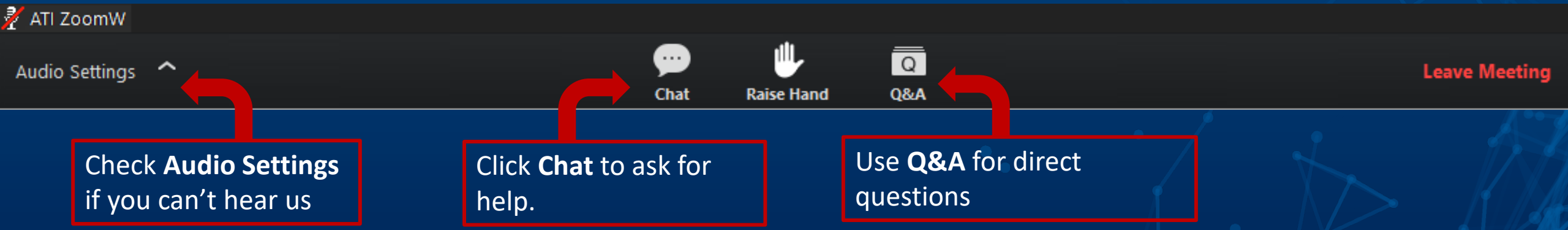
- **Data Protection Toolkit Use Case Analysis**
- **Expanding Equitable Access to Restricted-Use Data through Federal Statistical Research Data Centers**

AMERICA'S DATAHUB CONSORTIUM

NSF | NCSES

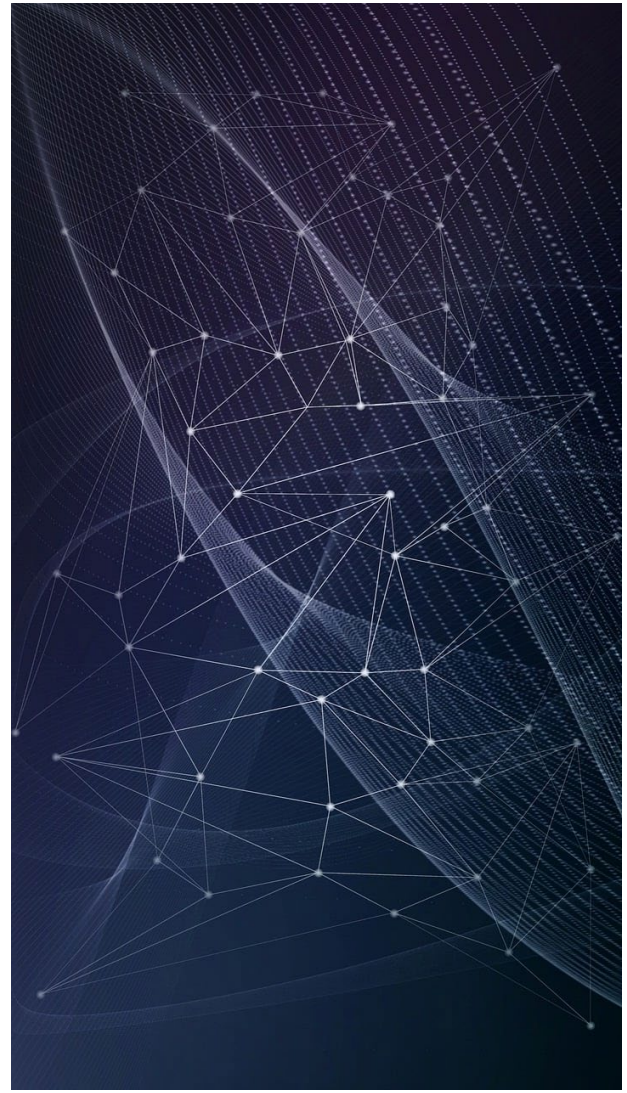
Housekeeping Items:

- 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.



Agenda

- Overview of the National Secure Data Service, Dr. Heather Madray
- Overview of Data Protection Toolkit Use Case Analysis, Dr. Heather Madray
- Project Update on Data Protection Toolkit Use Case Analysis, NORC
- Q&A
- Overview of Expanding Equitable Access to Restricted-Use Data through Federal Statistical Research Data Centers, Dr. Heather Madray
- Project Update on Expanding Equitable Access to Restricted-Use Data through Federal Statistical Research Data Centers, ICPSR, Regents of the University of Michigan
- Q&A
- Closing remarks



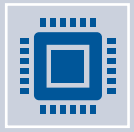
Overview of the National Secure Data Service (NSDS)

Dr. Heather Madray
NCSES, Program Director for Data Access, Confidentiality, and Quality Assessment (DACQA)

AMERICA'S DATAHUB CONSORTIUM

NSF | NCSES

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

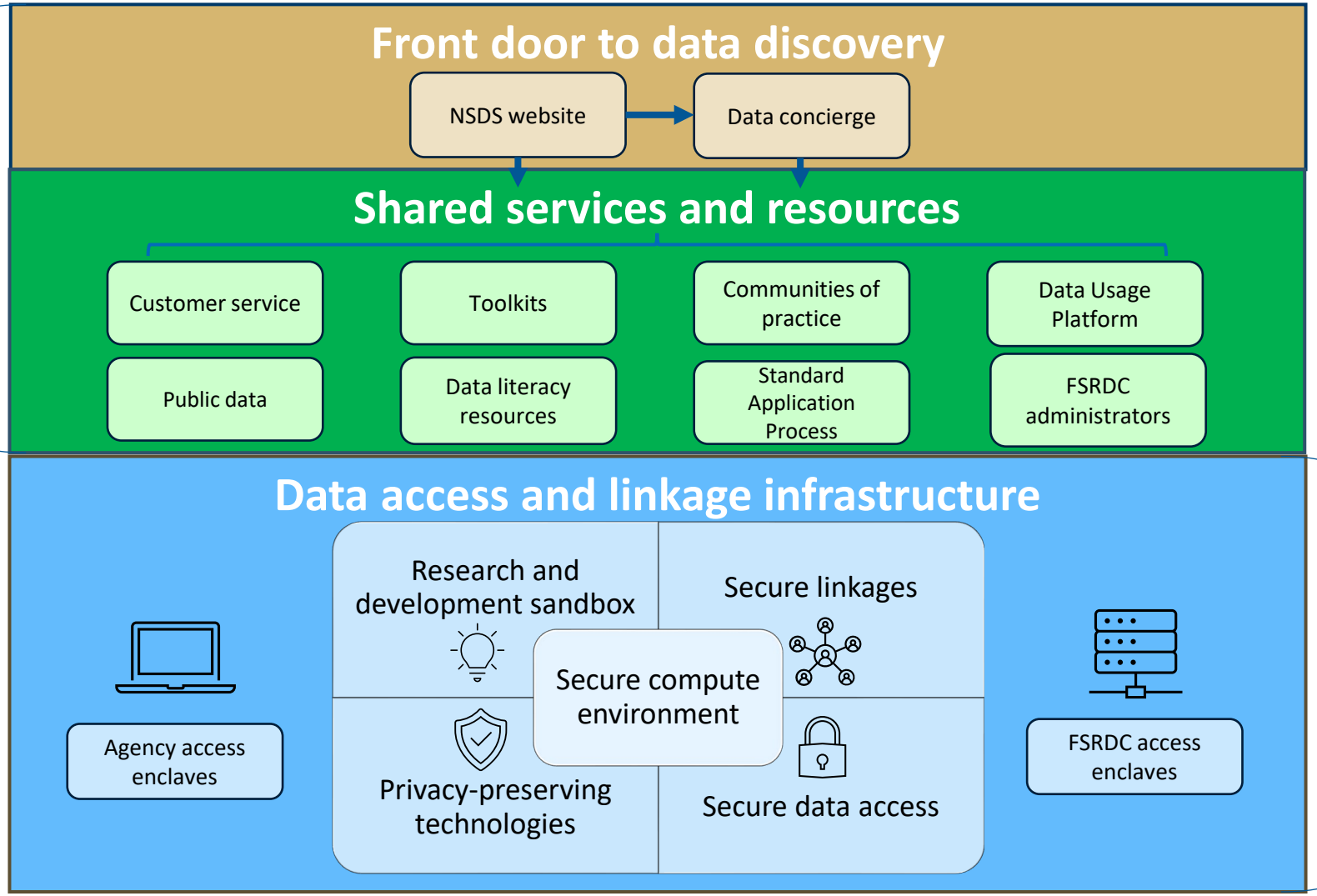
Climate data

Economic data

And more…

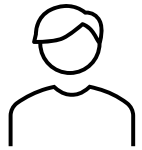
What could an NSDS look like?

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



College Administrator

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

An NSDS data concierge can guide users to public data.

Academic Researcher

How has government data on workforce education been used?

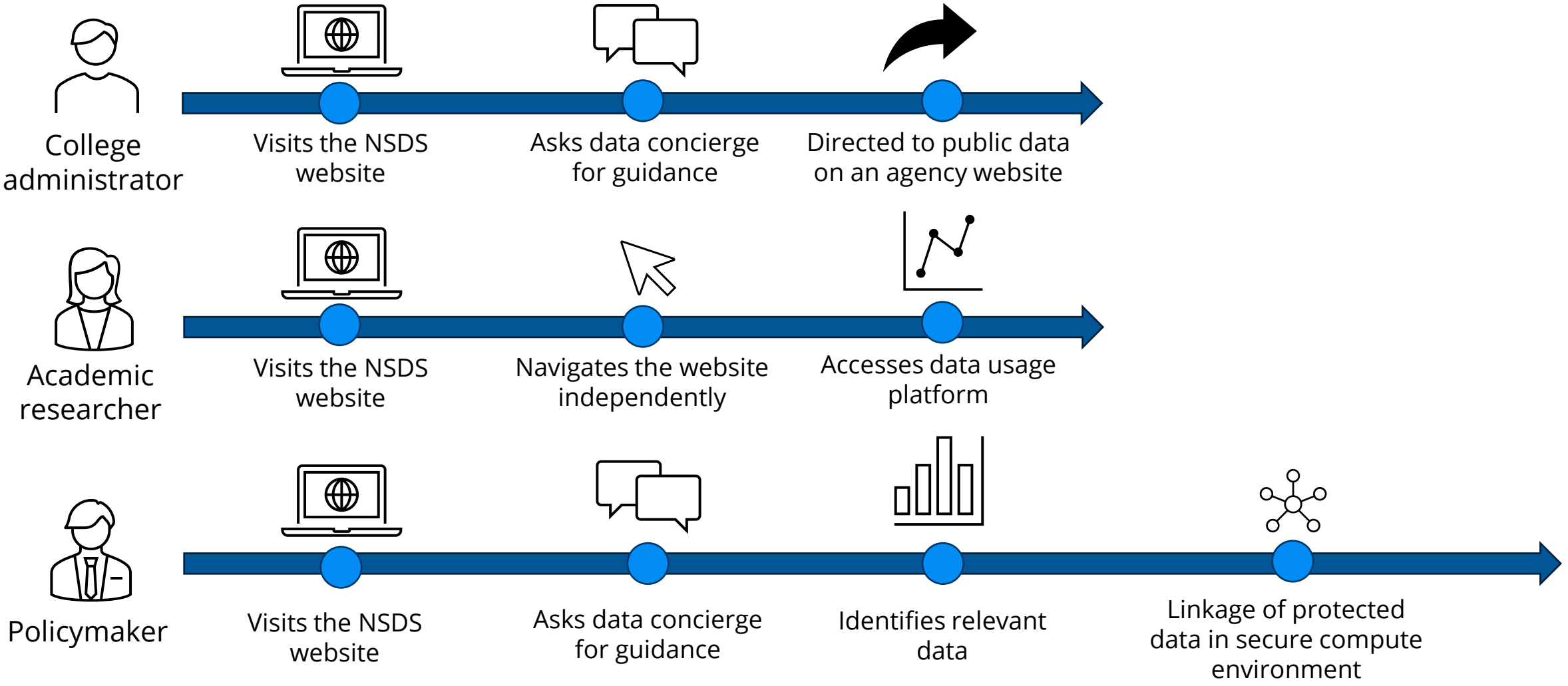
An NSDS data usage platform identifies published products using government data assets.

Policymaker

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



Use Case Analysis of the Data Protection Toolkit

September 26, 2024

Ben Reist

Team: Martha Stapleton, Seth Brohinsky, Peter Meyer, Julie Kubelka

NORC at the University of Chicago

Agenda

01 Introduction

02 Overview of Interviews

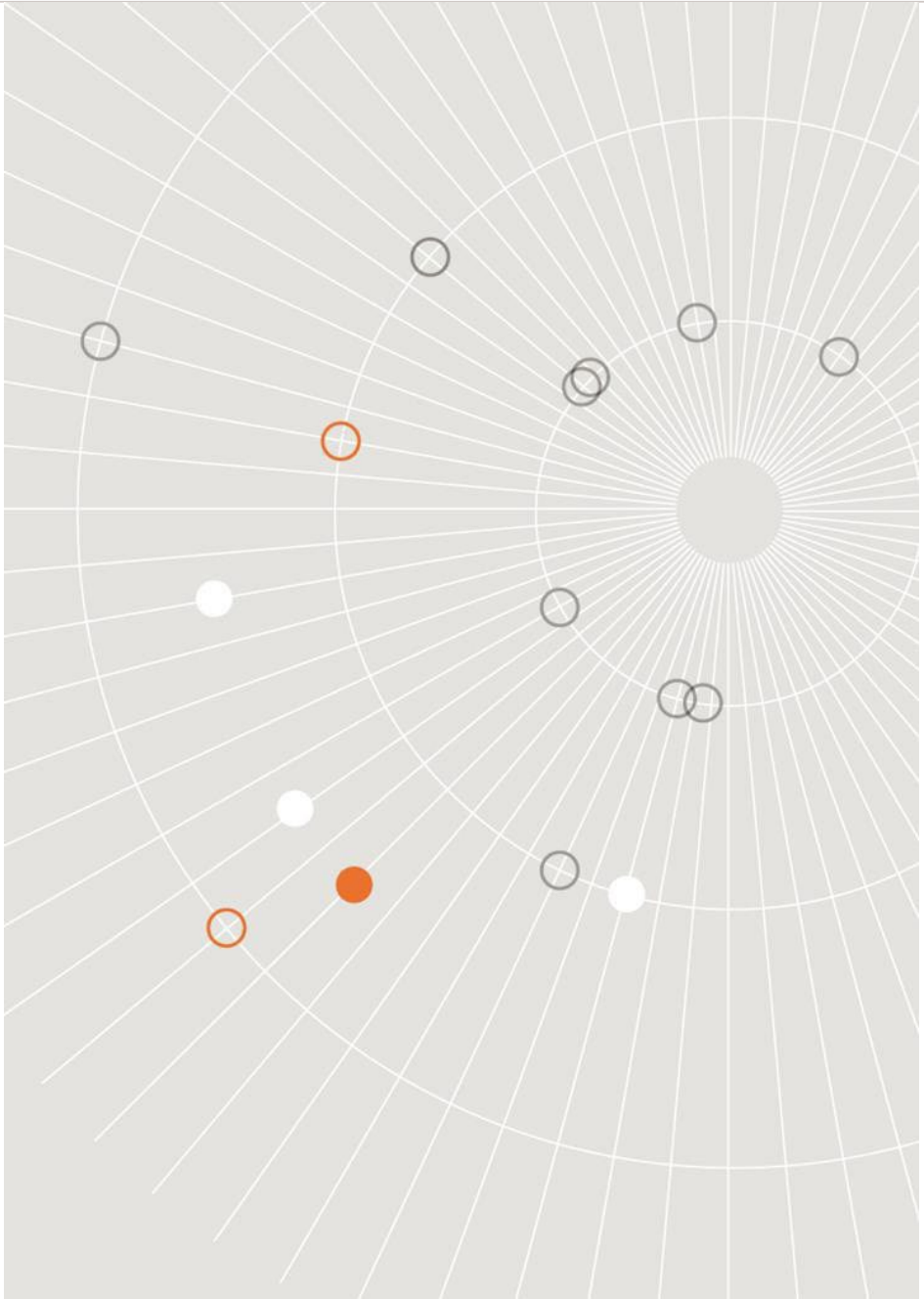
03 Findings

04 Next Steps

Introduction: Understanding the Context

Study Purpose

- Conduct a Use Case Analysis of the Data Protection Toolkit
- Identify successful uses and potential enhancements to the Toolkit
- Identify current data protection practices reported by federal and non-federal staff
- Receive feedback and suggested improvements to the Toolkit



About the Data Protection Toolkit (DPT)

Background

- Developed by the Federal Committee on Statistical Methodology (FCSM)
- DPT project is part of the Federal Data Strategy
- Currently hosted by NCES

- Provides information for Statistical Officials, Chief Data Officers, agency staff, and practitioners on increasing access to data assets while protecting confidentiality
- Includes—
  - Inventory of methods and models
  - Best Practices
  - Case Studies
  - Tools, Templates, and Automated Resources

NORC conducted 15 qualitative interviews to obtain the data reported in the final report.

Interview Structure

- Conducted between September 22 and November 9, 2023
- 60-minute virtual interviews via Zoom
- NORC moderated the interviews

Participant Sample

- NORC and NCSES collaborated to create the list of potential respondents to ensure representation across government, academia, and the private sector
- 9 respondents worked for federal agencies and 6 worked for non- federal organizations (state governments, universities, and research organizations)

Findings

We begin with findings about the current data protection environment from both federal and non-federal respondents, and then move to findings about the toolkit

The Current Data Protection Environment – Federal

Federal agencies employ multiple techniques to avoid disclosure of private and confidential information when disclosing or releasing data.

Federal agencies interviewed use a tiered access model for handling external researcher data access requests.

Two respondents noted that their data were never linked due to the siloed nature of their agency's program offices.

All federal agencies interviewed have a disclosure review process for publicly released data.

Some agencies interviewed have a formal Disclosure Review Board (DRB) to support the review process.

The federal agencies interviewed are balancing the desire to make more data available to the public, while protecting confidential information.

The Current Data Protection Environment – Non-Federal

- Non-federal organizations employ similar techniques to avoid disclosure of private and confidential information.
- They follow formal disclosure risk review guidelines and described the resources used to guide their disclosure review process.
- They described their processes to control access to sensitive data and ensure that output is void of confidential information.
- Respondents from state governments noted the importance of meeting both state legislative requirements around data release and data confidentiality and federal collection requirements for the state-based federal surveys.
- New differential privacy practices were concerning to non-federal respondents as they were perceived as unnecessary and would ultimately reduce data usability.

Feedback on the Toolkit – Generally Positive

Federal agency staff described the toolkit as helpful, saying it provided a source for standard information to share across agencies.

Respondents envisioned using the toolkit to learn and understand new data concepts such as differential privacy.

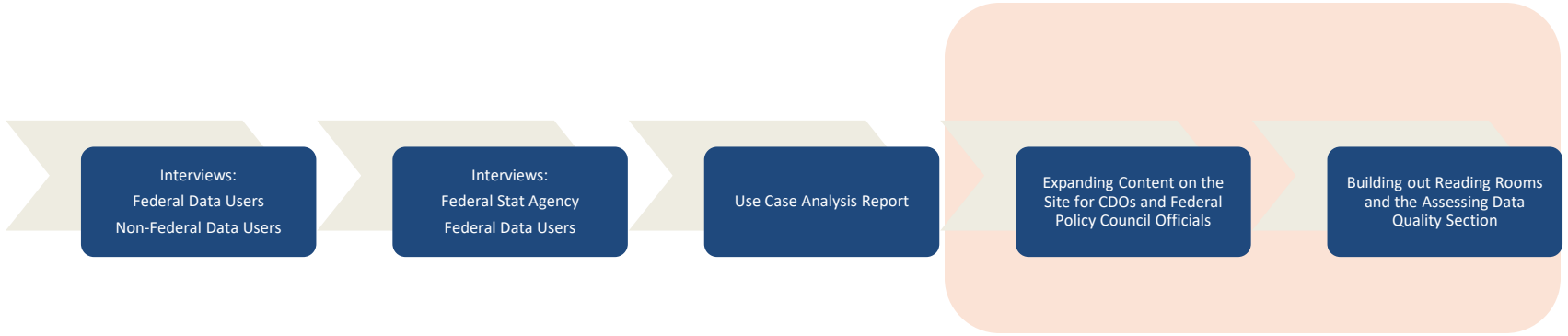
One respondent was using resources in the toolkit to support their work in establishing a DRB at their agency.

Feedback on the DPT - Opportunities for Improvement

- Some federal agency respondents were less sure if they would use the Toolkit as a reference because they knew the field and would turn to their colleagues with any questions.
- Respondents wondered if there would be additional information in the Toolkit on different data types.
- Non-federal staff noted that some of their organizations had DRBs but that the information on DRBs in the Toolkit was specific to federal DRBs.
- Respondents thought that the resource list in the Toolkit was extensive and difficult to navigate.
- Respondents noted that although there were different sections of the Toolkit, the intended audience and knowledge level were not explicit.
- Non-federal respondents suggested specific methods to introduce the Toolkit to new audiences.

Respondents found the DPT to be a great resource, but provided ideas to make the Toolkit more accessible and user-friendly

Timeline of DPT Project and Follow-up Activities



Fall - Winter 2023

Delivered: January 2024

2024

Late 2024

Questions?

Expanding Equitable Access to Restricted-Use Data through Federal Statistical Research Data Centers

Mike Castro – U.S. Census Bureau

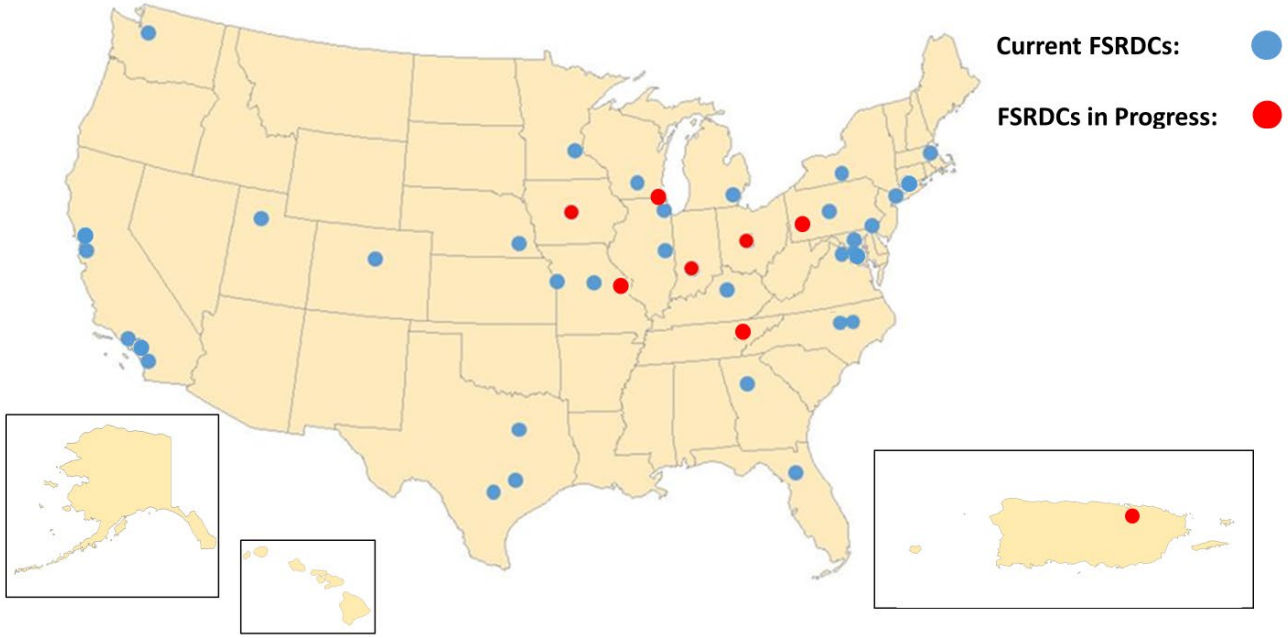
Mary E. Campbell – Texas A&M University

What are FSRDCs?

- Partnerships between federal agencies, research institutions, researchers
  - Provide access to restricted microdata for research
  - Findings benefit agencies, researchers and organizations, public

- Census Bureau-managed secure labs at partner institutions
  - Universities and Federal Reserve Banks
  - Virtual access option for affiliated researchers on qualifying projects

FSRDC Program Reach



- 33 secure, networked locations
- Collaborations between local institutional partners
- 125+ academic and non-profit research organizations members

Example - Atlanta FSRDC Partner Institutions:

Auburn University	Georgia Institute of Technology	University of Georgia
Clemson University	Georgia State University	University of South Carolina
Emory University	Tulane University	University of Tennessee
Atlanta Federal Reserve Bank	University of Alabama-Birmingham	Vanderbilt University

Challenge: Expand Equitable Access

- Expand user-base beyond Research-1 institutions
- Be a resource for
  - Minority Serving Institutions
  - Non-R1 universities
  - State and local governments
  - Non-profit institutions
- Level the playing field such that qualified users are equally capable of applying for and utilizing FSRDC data
  - Remove individual and institutional barriers

Current FSRDC access is dominated by high-resource institutions

Institution Type	n	%
R1: Doctoral Universities – Very High Research Activity	78	62%
R2: Doctoral Universities – High Research Activity	4	3%
M1/M2: Master's Colleges & Universities	5	4%
Special Focus Four-Year: Research Institution	3	2%
Academic Affiliations (Non-Carnegie) (e.g., research centers)	10	8%
Non-Academic Organizations	25	20%
Total:	*125*	*100%*

Opportunity: Research the Issue

- Share equity goals with NSDS
- How to resolve current inequities in FSRDC program and not perpetuate into NSDS?
- Goal – Conduct environmental scan of prospective users to understand:
  - Who are our underserved users?
  - What are their needs?
  - What barriers to entry do they face?
  - How can we remove those barriers?
- NCSES/NSF – ADC Award: Research Team
  - Texas A&M
  - University of Michigan
  - National Latino Research Center @ Cal State San Marcos

Research Overview and Recruitment

- Surveys and focus groups
  - Emails to Minority Serving Institutions (MSIs), nonprofit organizations, and state/local government research groups
  - Additional recruiting through existing FSRDC networks
- Survey N=360
  - 201 academic (130 MSIs and 38 PWIs), 38 nonprofit organizations, 73 state/local government, 67 other or not specified
- Focus groups: 8 groups with total N=23
  - 11 from MSIs, 1 from PWI, 4 from nonprofit organizations, 7 from local/state government

Preliminary results

	MSI (N=130)	PWI (N=38)	Academic (N=201)	Non-Academic (N=116)
Yes, can easily find information about FSRDCs	29%	55%	37%	29%

Preliminary results

Usefulness	MSI (N=130)		PWI (N=38)		Academic (N= 201)		Non-Academic (N=116)	
	M	Confidence Level	M	Confidence Level	M	Confidence Level	M	Confidence Level
Usefulness of in-person access	3.1	Neutral	3.5	Neutral	3.3	Neutral	2.7	Neutral
Usefulness of online access	4.8	Very Useful	4.9	Very Useful	4.9	Very Useful	4.6	Very Useful
Usefulness of Hybrid Access (online and in-person)	4.0	Useful	4.3	Useful	4.1	Useful	3.7	Useful

Note: Not Useful (NU) = 1 – 1.50; Somewhat Useful (SU) = 1.51 – 2.50; Neutral (N) = 2.51 – 3.50; Useful (U) = 3.51 – 4.50; Very Useful (VU) = 4.51 – 5.00.

Preliminary results

Confidence…	MSI (N=130)		PWI (N=38)		Academic (N=201)		Non-Academic (N=116)	
	M	Confidence Level	M	Confidence Level	M	Confidence Level	M	Confidence Level
…your organization will support a long-term research project	2.7	Confident	2.8	Confident	2.7	Confident	2.3	Somewhat Confident
…you can gain travel support	2.5	Confident	2.7	Confident	2.6	Confident	2.2	Somewhat Confident
…you can travel internationally for no more than 6 months at a time	2.9	Confident	3.3	Confident	3.0	Confident	2.6	Confident
…designing research methods	3.1	Confident	3.2	Confident	3.1	Confident	2.6	Confident
…finding information in codebooks	2.9	Confident	3.1	Confident	3.1	Confident	2.9	Confident
…using technical memos	2.4	Somewhat Confident	2.7	Confident	2.6	Confident	2.9	Confident

Note: Not Confident (NC) = 1 – 1.50; Somewhat Confident (SC) = 1.51 – 2.50; Confident (C) = 2.51 – 3.50; Very Confident (VC) = 3.51 – 4.00

Lessons learned so far

- Recruitment challenges mirroring usage patterns
- Feedback from the survey
- Impressions from the focus groups

Questions?

Stay Engaged

- Solicitation and Contract Related Questions: ADC-Contracts@ati.org
- General/Membership Questions: adc@ati.org
- Join the ADC Mailing List: www.americasdatahub.org/adc-mailing-request-form/
- Read Lessons Learned: 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

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	View Final Report
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	View Final Report
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 tche 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	

- Project Quarterly Lessons Learned
- Final Reports
- Project Status

Thank you!

AMERICA'S DATAHUB
CONSORTIUM