

Project Update Webinar

Privacy Preserving Technologies

Projects Covered

- Privacy Preserving Technologies Phase 1: Environmental Scan
- Utilizing Privacy Preserving Record Linkage to Link Data from Two Federal Statistical Agencies
- Utilizing Privacy Preserving Record Linkage with Parent Agency Data and Statistical Agency to Inform Programs and Policies

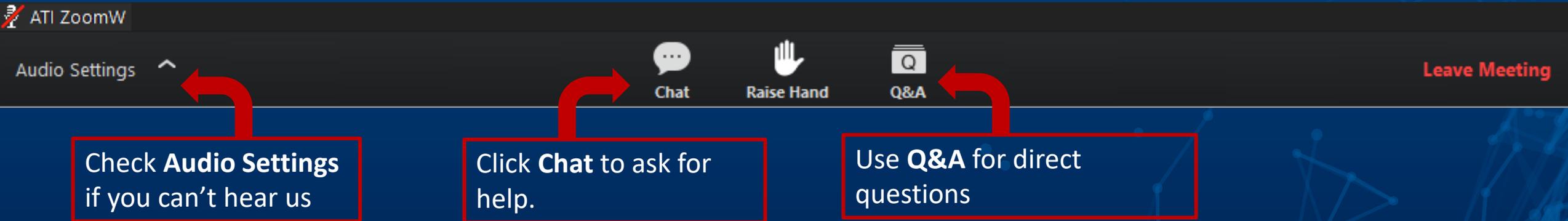


AMERICA'S DATAHUB
CONSORTIUM



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.



The image shows a screenshot of the Zoom meeting control bar. The controls include 'Audio Settings', 'Chat', 'Raise Hand', 'Q&A', and 'Leave Meeting'. Three red arrows point from text boxes below to the 'Audio Settings', 'Chat', and 'Q&A' icons. The text boxes contain the following instructions:

- Check **Audio Settings** if you can't hear us
- Click **Chat** to ask for help.
- Use **Q&A** for direct questions

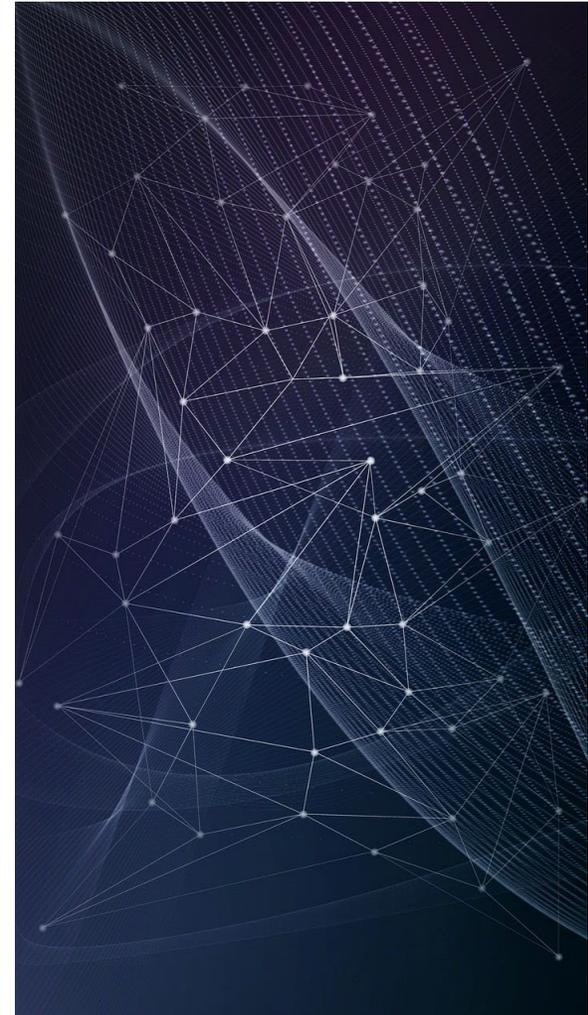


AMERICA'S DATAHUB
CONSORTIUM



Agenda

- Overview of the National Secure Data Service (NSDS), Heather Madray
- Privacy Preserving Technologies (PPTs) and the NSDS Demonstration Project
- Privacy Preserving Technologies: Environmental Scan
- Project Update on Utilizing Privacy Preserving Record Linkage Activities
- Current Status
- Lessons Learned and Final Thoughts
- Q&A
- Closing remarks



Background: National Secure Data Service (NSDS) Demonstration Project

Heather Madray
Supervisory Program Director
National Center for Science and Engineering Statistics (NCSES)



AMERICA'S DATAHUB
CONSORTIUM



CHIPS and Science Act Requirements (§10375)



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 a set of shared services and a government-wide data linkage and access infrastructure to support **evidence building**.



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



The NSDS Demonstration Project 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

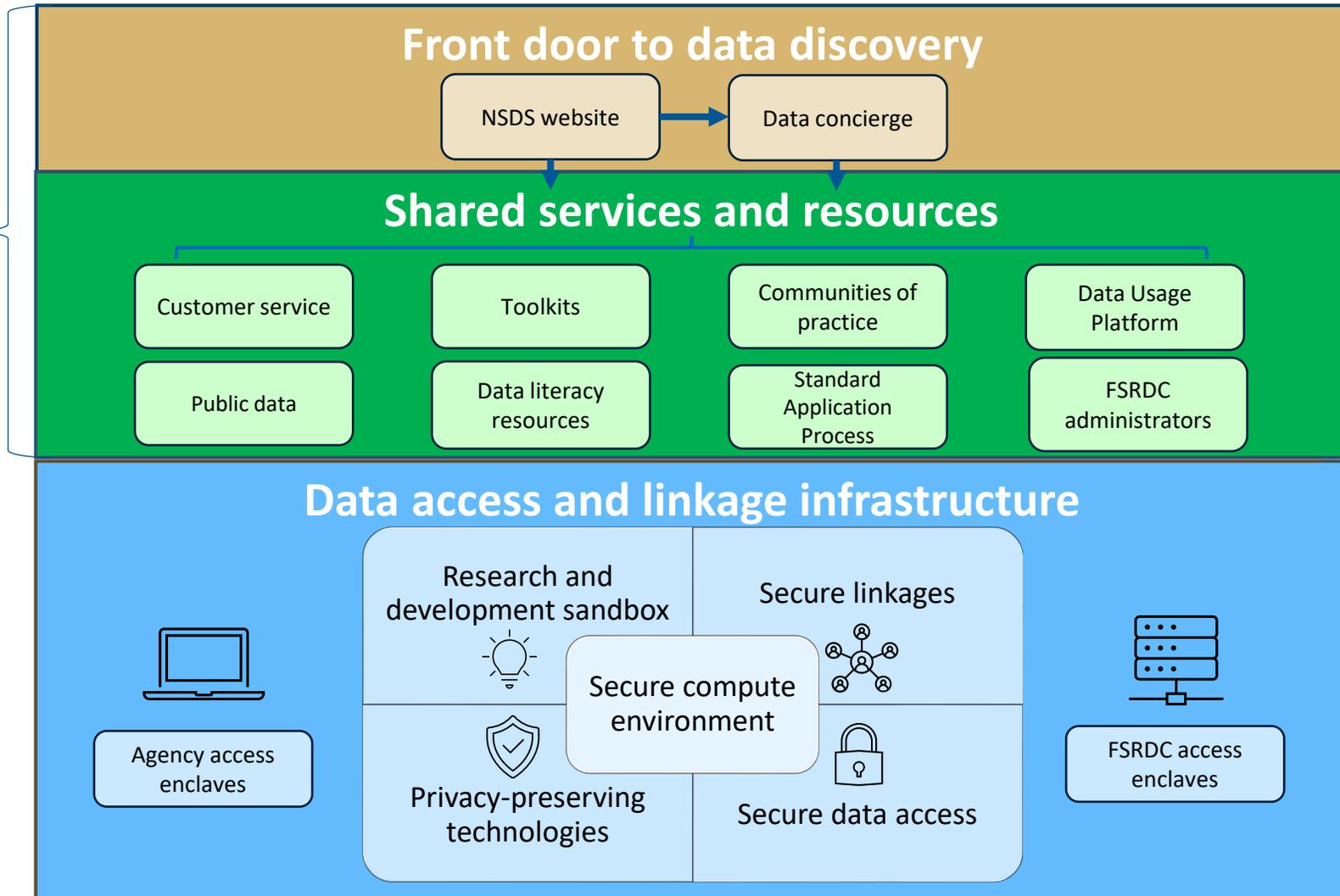
And more...



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.

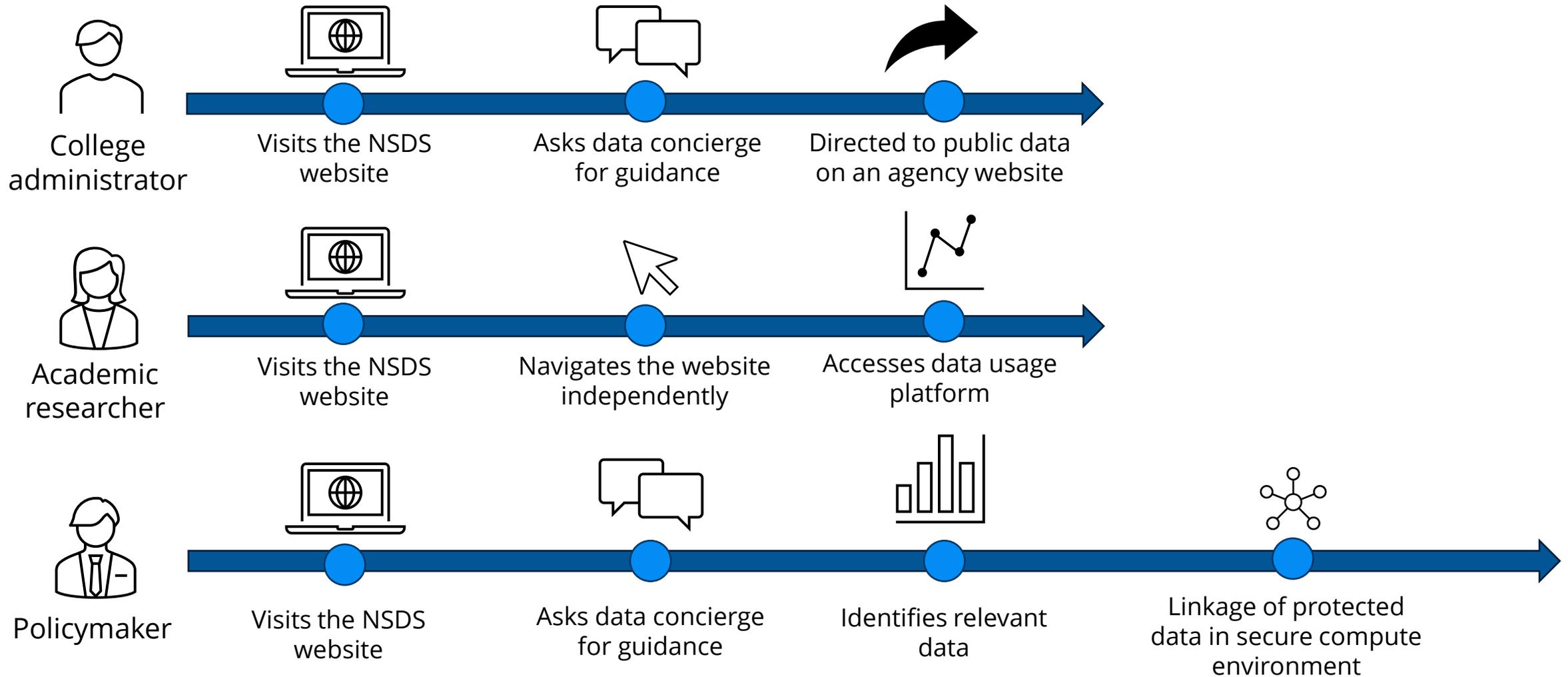


How are health outcomes influenced by access to transportation services?



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

NSDS User Journey Examples



Privacy Preserving Technologies and the NSDS Demonstration Project

Lisa Mirel

Statistical Advisor

National Center for Science and Engineering Statistics (NCSES)



AMERICA'S DATAHUB
CONSORTIUM



PPTs

- Increasing access to data to make informed policy decisions has come to the forefront
- Accessing these data in secure ways has posed challenges
- There are privacy enhancing technologies that provide tools and approaches to mitigate privacy risks with confidential or restricted data, thus enabling new opportunities for:
 - Linkage
 - Dissemination
 - Analysis

Three PPT Projects

Project Title	Project Performer/ Partner	Period of Performance
Privacy Preserving Technologies Phase 1: Environmental Scan	RTI International/ Flood Mason	July 2023- January 2024
Utilizing Privacy Preserving Record Linkage (PPRL) to Link Data from Two Federal Statistical Agencies	HealthVerity, Inc./ Mathematica, Inc.	September 2023- August 2025
Utilizing PPRL with Parent and Statistical Agency Data to Inform Programs and Policies	NORC at the University of Chicago	September 2023- September 2025

PPTs: Environmental Scan

Environmental Scan Results to Inform NSDS

Key Components for PPT Implementation

- Need for standards
- Expertise may be met by vendors/industry partners
- Training and talent growth within federal sectors are necessary to support growing need for PPT implementation

Integrating Legal, Regulatory, and Policy Requirements

- Understanding data privacy laws and regulations
- Bridging information gaps between technology and legal teams
- Leveraging technology and governance best practices to streamline data sharing agreements

Keys to Success when Utilizing PPTs

- Starting with the research/policy question → determine the right PPT solution
- Forming multidisciplinary team
- Building trust through inclusion and communication

Privacy Preserving Technologies: To Safely Process and Share Data

Input Privacy

- Allow two or more parties to submit data into a calculation without the other respective parties seeing the data
- Potential tools:
 - Secure Multi-Party Computation (sMPC)
 - Homomorphic Encryption

Output Privacy

- Modifying results of a computation such that the output cannot be used to reverse engineer the original input
- Potential tools:
 - De-identification/anonymization
 - Synthetic data
 - Noise infusion/differential privacy

https://unstats.un.org/bigdata/task-teams/privacy/guide/2023_UN%20PET%20Guide.pdf

Secure Multi-party Computation (sMPC)

- Cryptographic technique that mitigates input privacy
- Data users only learn about the output of computation and not the input
- One application of sMPC is privacy preserving record linkage (PPRL)

Utilizing PPRL Activities

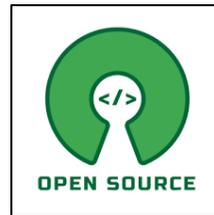
NSDS Demonstration Projects

- Two PPRL projects are underway to assess:
 - Development of data sharing and governance agreements between two separate federal statistical agencies and within a parent and federal statistical agency
 - Deployment of commercial PPRL tool (HealthVerity)



vs.

open-source PPRL tool (Anonlink)





NSDS Demonstration: PPRL Activities

Objectives:

1. Develop data sharing and governance prototype agreements
2. Utilize PPRL tools when linking person level data from disparate sources
3. Create analytic datasets that can be used to inform questions that could not be assessed with either source alone

Utilizing a commercial PPRL tool (HealthVerity) to link data from two Federal Statistical Agencies

Linking Data: Parent and Federal Statistical Agency

Data sources



National Center for Health Statistics (NCHS): National Health Interview Survey (NHIS)



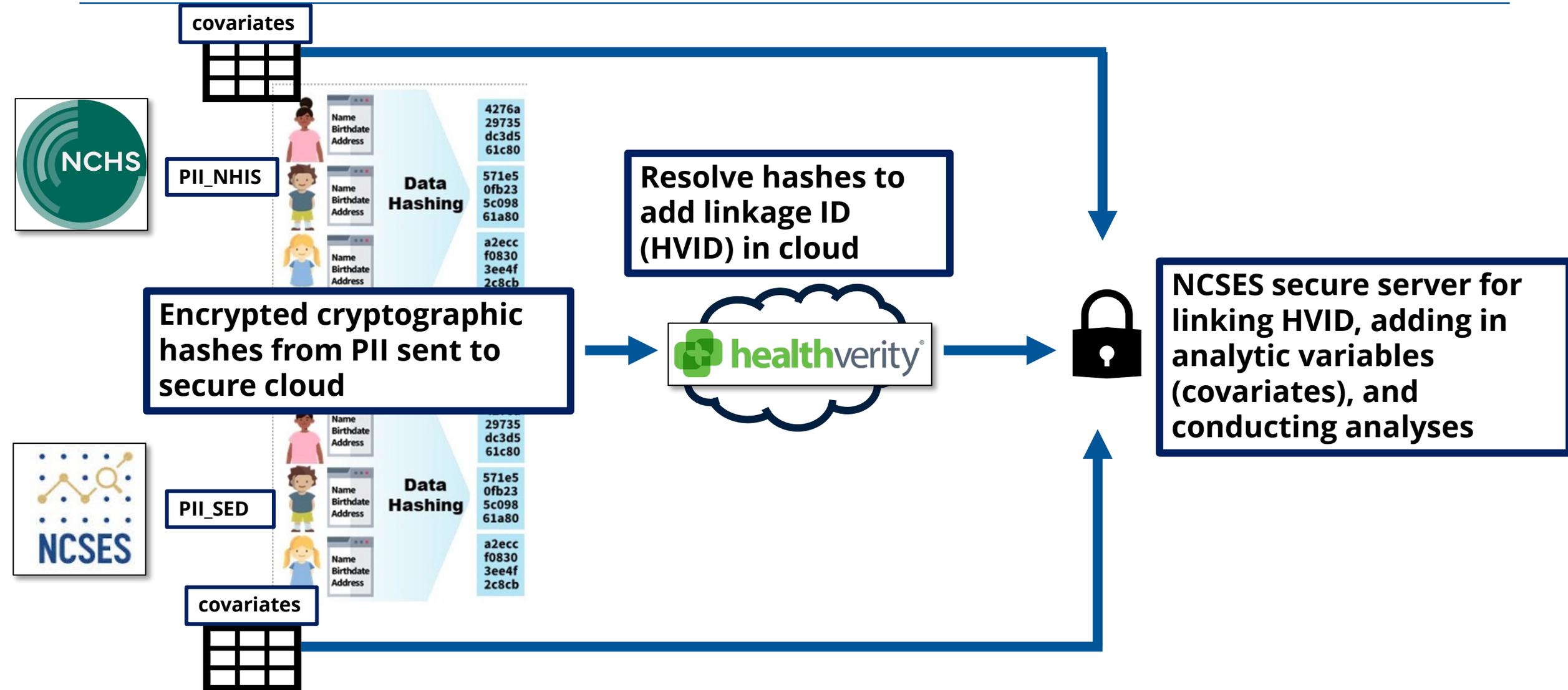
NCSES: Survey of Earned Doctorates (SED)

Goal

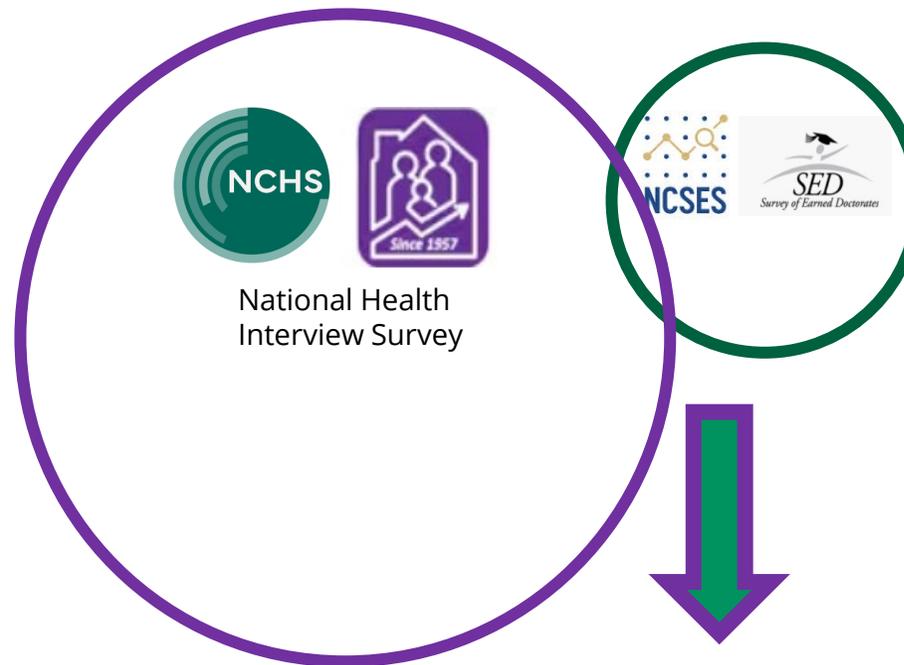
- Develop data sharing agreement to link data from two federal statistical agencies
- Establish processes for utilizing a commercial PPRL tool: HealthVerity
- Inform analytic initiatives



Data Flow: Two Federal Statistical Agencies



Data Linkage: NHIS and SED

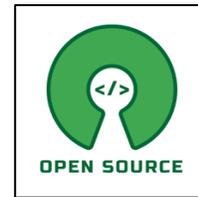


NHIS participants who earned a research doctorate from an accredited U.S. institution

Analysis: Opportunities and Limitations with the Linked NHIS and SED Data

- Compare composite estimates from each survey based on disability questions for those who linked
- Opportunities: directly compare questions and responses by the same individuals
- Limitations:
 - Timing when questions are asked
 - Small subset of population

**Utilizing an open-source PPRL
tool (Anonlink)**



**with a parent and federal
statistical agency**

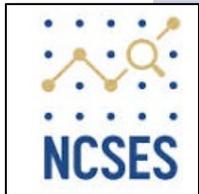


Linking Data: Parent and Federal Statistical Agency

Data sources



NSF Office of the Chief Information Officer: Principal Investigator (PI) award data



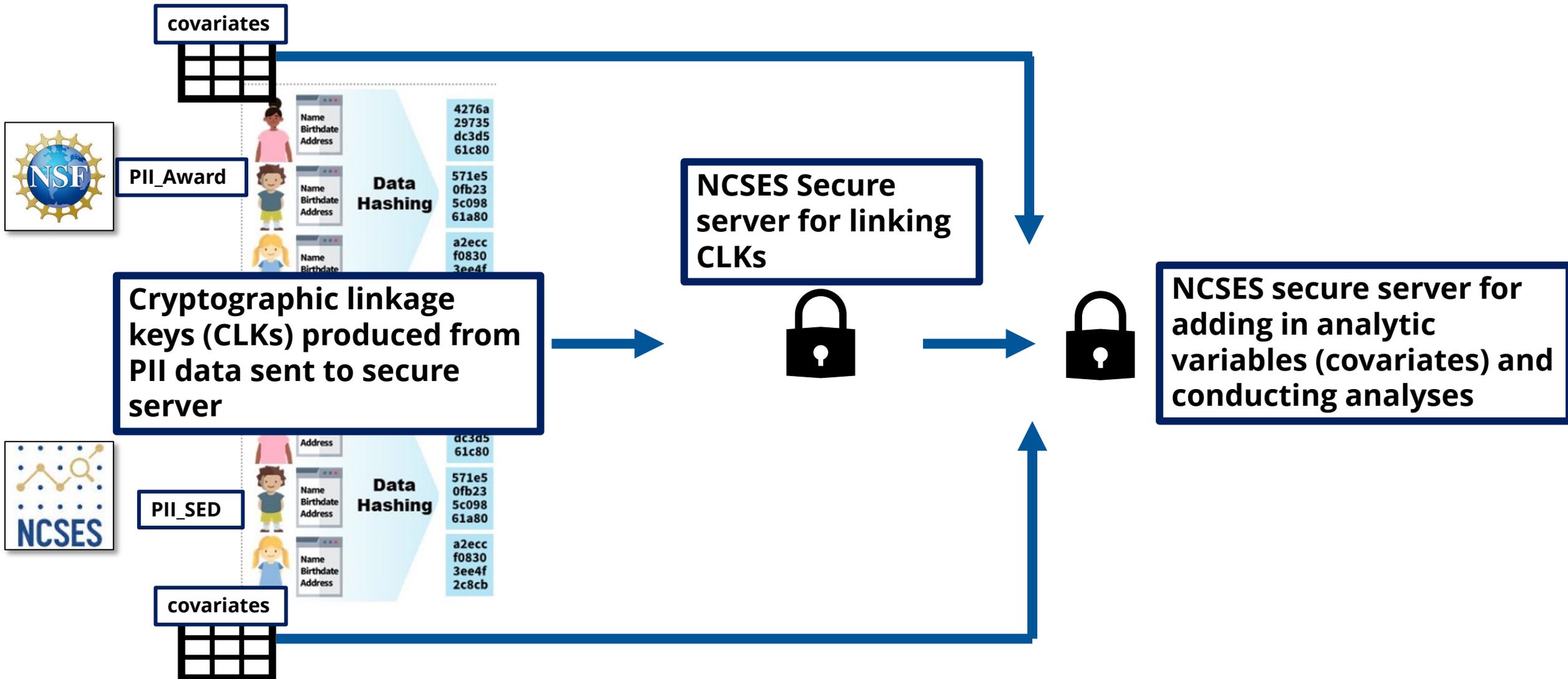
NCSES: SED

Goal

- Develop data sharing agreement to link data between parent and federal statistical agency
- Implement an open-source PPRL tool: Anonlink
- Inform programs and policies that support the NSF's strategic plan



Data Flow: Parent and Federal Statistical Agency



Data Linkage: SED and PI NSF Awards



Doctorate recipients that received an NSF Award

Analysis: Opportunities and Limitations with the Linked SED and PI NSF Awards Data

- Trajectory of U.S. doctoral recipients after completing a doctorate degree
 - Assess if there are differences by demographics and field of study
- Opportunities: inform programs and policies that support the NSF's strategic plan
- Limitations:
 - Limited variables available for linkage
 - Temporal alignment

Current Status

Completed and Remaining Steps

Objective	NHIS-SED Linkage	SED-PI Linkage
1. Develop data sharing and governance agreement	<ul style="list-style-type: none">✓ Review and sign software license agreement✓ Obtain approval to conduct linkage in a commercial cloud with FedRAMP✓ Obtain final signatures on data sharing agreement	<ul style="list-style-type: none">✓ Obtain final signatures on data sharing agreement
2. Utilize a PPRL tool	<ul style="list-style-type: none">❑ Standardize PII fields to create encrypted cryptographic hashes	<ul style="list-style-type: none">❑ Standardize PII fields to create cryptographic linkage keys
3. Create an analytic dataset	<ul style="list-style-type: none">❑ Link/analyze data in secure environment	<ul style="list-style-type: none">❑ Link/analyze data in secure environment

Lessons Learned and Final Thoughts



Lessons Learned to Date

- Establish key stakeholders and champions at both data sharing institutions early in the process
- Develop understanding of the processes and technical infrastructure needed to securely use commercial PPRL tool in cloud environment
- Allow sufficient time for finalizing data sharing agreements

Final Thoughts

- A successful, future NSDS, that supports using data for evidence building, will have transparent and standardized processes for
 - Developing data sharing and governance agreements for different types of agencies
 - Establishing a secure compute environment where unique PPRL tools (commercial and open source) are available
- Continuous need for additional use cases to further the success

Questions?



AMERICA'S DATAHUB
CONSORTIUM



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 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

