



Building the Future: Leveraging the National AI Research Resource and National Secure Data Service for AI Capabilities

Katerina Antypas, National Science Foundation, Office of Advanced Cyberinfrastructure

Lisa Mirel, National Science Foundation, National Center for Science and Engineering Statistics

National Center for Science and Engineering Statistics
Social, Behavioral and Economic Sciences
National Science Foundation

Outline

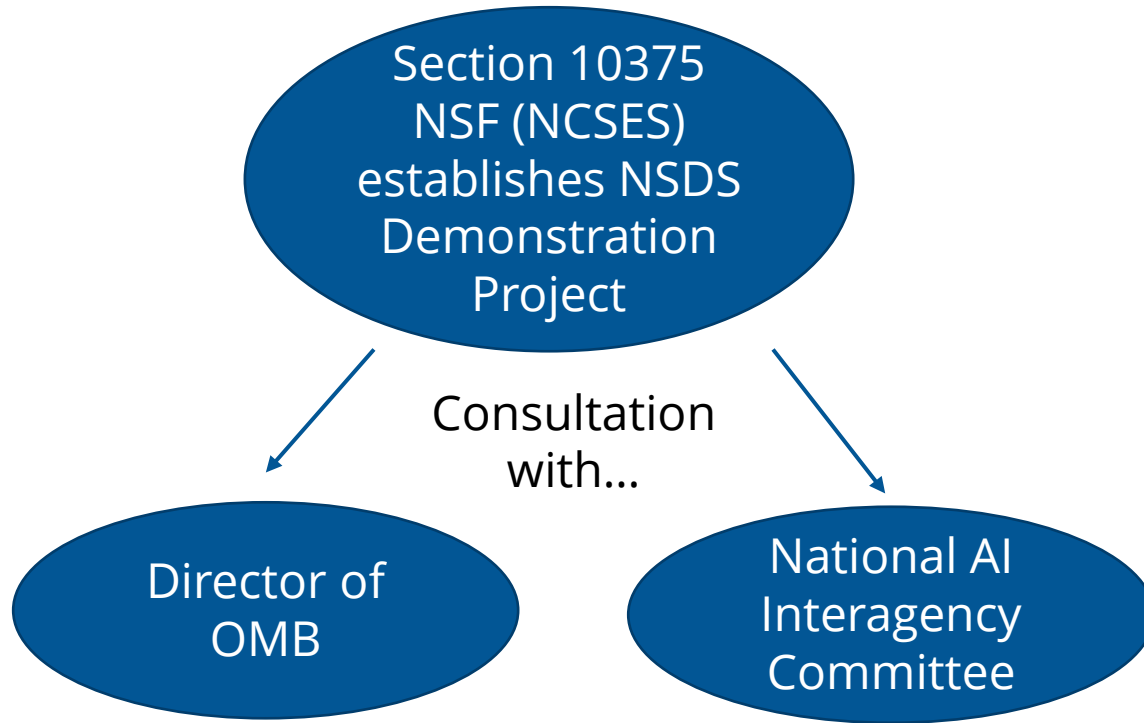
- Introduction
- National Secure Data Service (NSDS) Demonstration Project Overview
- National AI Research Resource (NAIRR) Pilot Overview
- Synergies and Opportunities with NSDS Demonstration and NAIRR Pilot
- Questions
- Open floor discussion

NSDS Demonstration Overview

CHIPS and Science Act of 2022

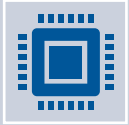
(a) In general...

Shall...



Establish a demonstration project to develop, refine, and test models to inform the full implementation for a government-wide data linkage and data access infrastructure for statistical activities conducted for statistical purposes, as defined in chapter 35 of title 44, United States Code.

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 an NSDS



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

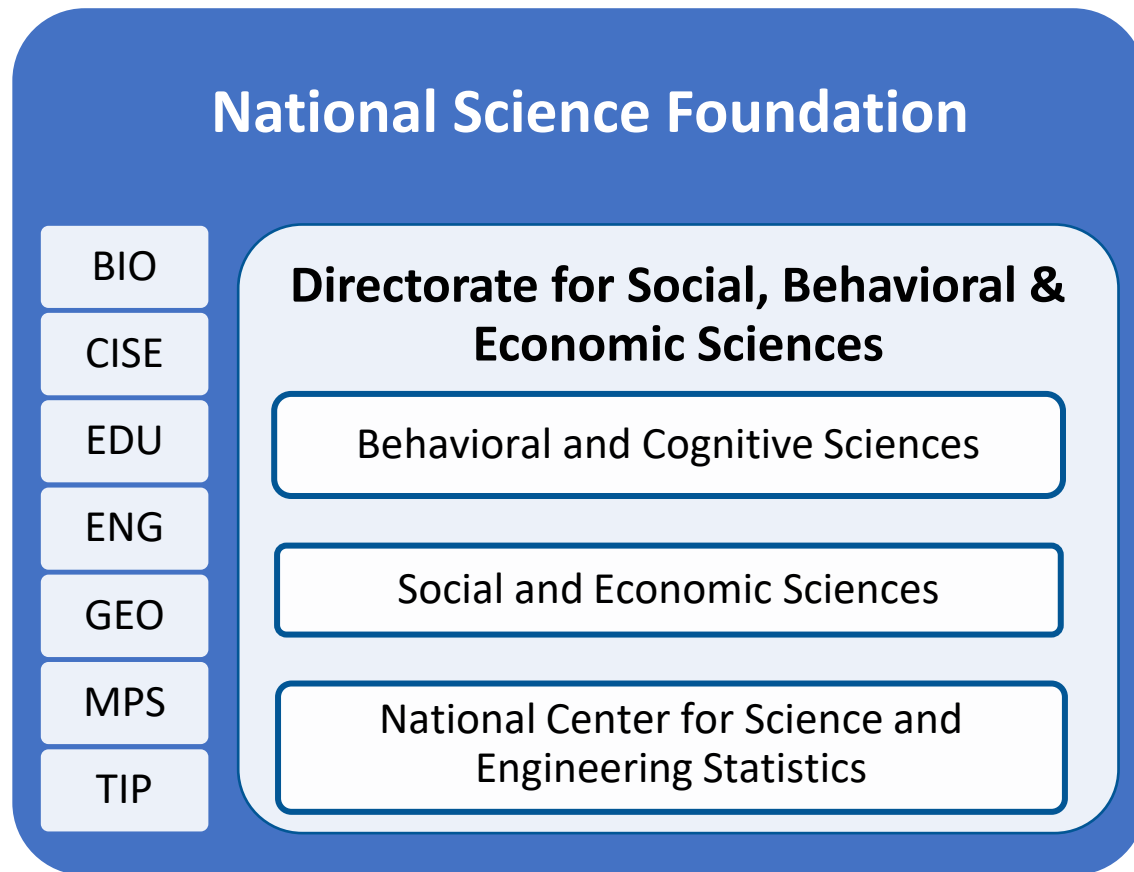


CHIPS and Science Act calls for consultation with the director of OMB, the NAIRR, and alignment with the Advisory Council on Data for Evidence Building (ACDEB) recommendations



The NSDS Demonstration will be implemented by the National Center for Science and Engineering Statistics (NCSES) within the U.S. National Science Foundation (NSF)

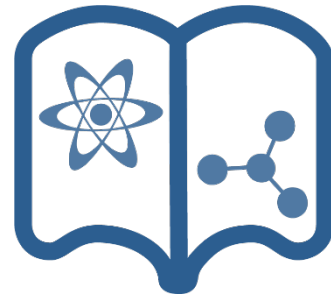
NCSES is one of 13 principal federal statistical agencies and is found within the NSF



NCSES's mission is to serve as a federal clearinghouse for objective data that provide key insights into the science and engineering enterprise



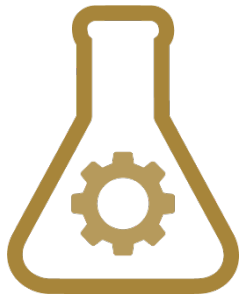
**Science & Engineering
Workforce**



STEM Education



**Innovation & Global
Competitiveness**



Research & Development

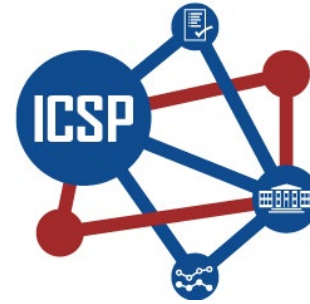


**Government Funding
for S&E**



Higher Ed R&D

NSDS Demonstration: Federal Agency Collaboration



NSDS Demonstration: Additional Stakeholder Engagement



State Agencies



Federal Statistical
Research Data Centers



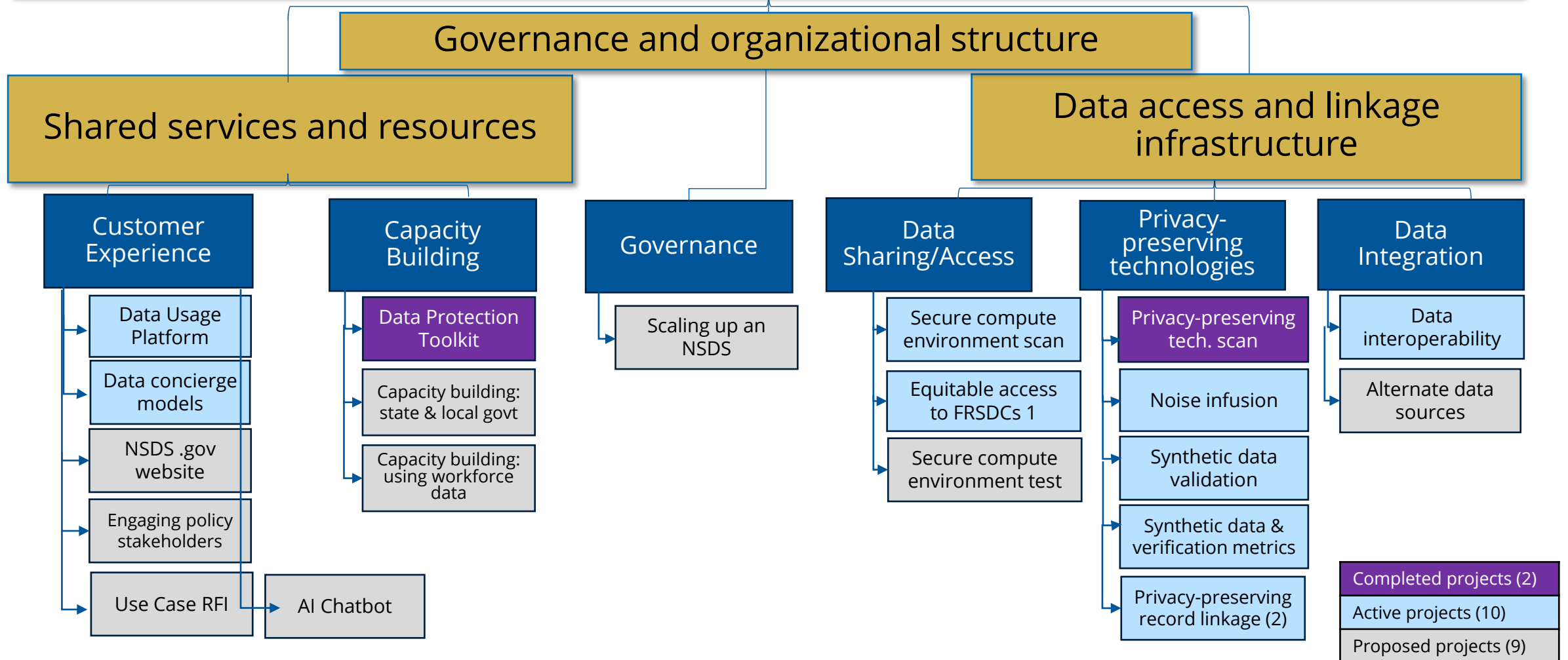
Federally Funded Research
and Development Centers



Private Sector

NSDS Demonstration: Piloting an Ecosystem

National Secure Data Service Demonstration



Vision for the NSDS



NSDS is envisioned as a government-wide set of shared services. It would serve as a **front door** and a central hub for users to discover shared services and resources and to utilize the NSDS data access and linkage infrastructure.



Shared services and resources will include a data concierge service to direct users on their evidence-building journey, toolkits for data protection and synthetic data, communities of practice where users can connect, and platforms to promote transparency concerning the use of government data.

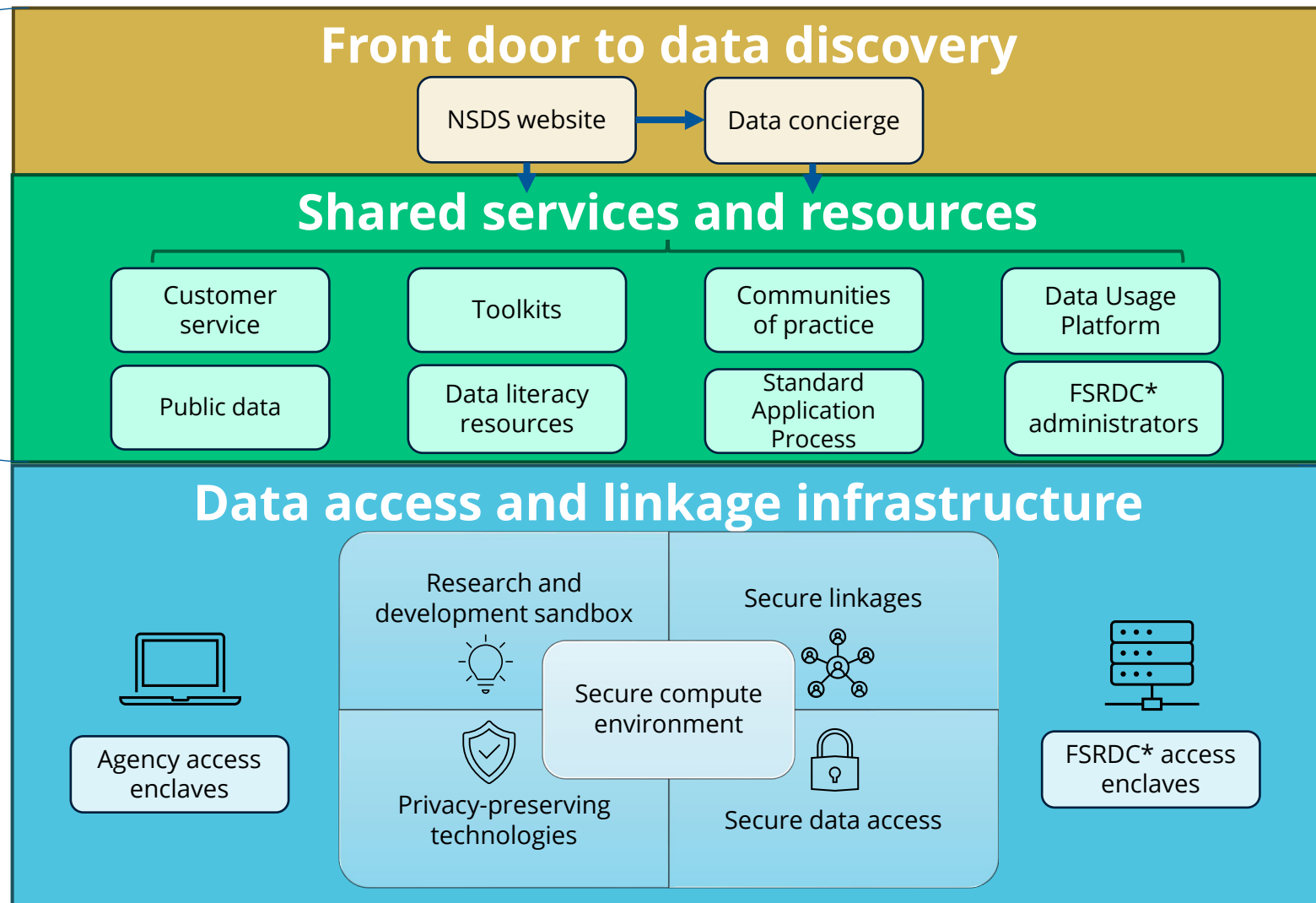


Data access and linkage infrastructure will enable users to securely access, link, and analyze powerful, high value data. The NSDS access infrastructure will complement agency enclaves and the Federal Statistical Research Data Center (FSRDC) network while addressing gaps in coverage. The NSDS will also facilitate secure data linkages in support of distinct, authorized projects.

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

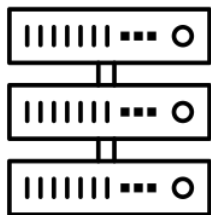
The data concierge guides users to the appropriate secure access modality, directs them to attain needed security credentials, and helps initiate linkages or leverage tools.

*FSRDC = Federal Statistical Research Data Centers

NAIRR Pilot Overview

Vision for the National AI Research Resource

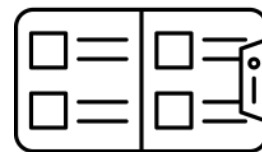
A widely-accessible, national research infrastructure that will advance the U.S. AI R&D environment, discovery, and innovation by empowering a diverse set of users through access to:



Secure, high-performance, privacy-preserving **computing**



High-quality **datasets**



Catalogs of **testbeds** and **educational materials**



Training tools and **user support** mechanisms

Objective: To strengthen and democratize the U.S. AI Innovation ecosystem in a manner that protects privacy, civil rights, and civil liberties

Why do we need a NAIRR?

- Many potential contributors lack access to requisite resources which can be costly *as well as* hard to navigate
- Researchers investigating AI to serve the public good require access to resources
- To train the next generation of researchers and AI leaders



Urgent national goals we are aiming to fulfill with NAIRR



Spur
innovation



Increase the **diversity**
of talent in AI



Improve U.S. **capacity**
for AI R&D



Advance
trustworthy AI

- Facilitate *national, coordinated* access to AI resources for the broad research and education community to accelerate discovery and innovation.
- Assure that public interest is strongly represented in AI and drives a responsible and trustworthy AI ecosystem.
- Combine forces across the S&T enterprise to increase AI resource capacity and expertise.

Full NAIRR Vision vs NAIRR Pilot Goals



NAIRR Pilot Goals

1. Demonstrate the value & impact of the NAIRR concept.
2. Support novel & transformative AI research while reaching broad communities.
3. Gain experience to advance and refine NAIRR design.

** NOTE: funding only an estimate from the Task Force. No funding has yet been appropriated for NAIRR.*

Bringing together the strengths of government, private industry and non-profit partners

Contributing Partners

Agencies

- National Science Foundation
- Defense Advanced Research Projects Agency
- Department of Agriculture
- Department of Defense
- Department of Energy
- Department of Veterans Affairs
- National Aeronautics and Space Administration
- National Institutes of Health
- National Institute of Standards and Technology
- National Oceanic and Atmospheric Administration
- US Geological Survey
- US Patent and Trademark Office (USPTO)

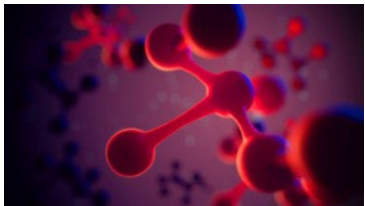
Non-governmental orgs

- AI2: Allen Institute for AI
- AMD
- Amazon Web Services
- Anthropic
- Cerebras
- Databricks
- Datavant
- EleutherAI
- Google
- Groq
- Hewlett Packard Enterprise
- Hugging Face
- IBM
- Intel
- Meta
- Microsoft
- MLCommons
- NVIDIA
- Omidyar Networks
- OpenAI
- OpenMined
- Palantir
- Regenstrief Institute
- SambaNova Systems
- Vocareum
- Weights & Biases

NAIRR Pilot Users



AI Researchers



Domain Scientists
Applying AI



Students and
Educators

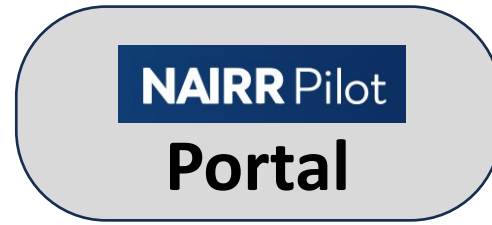
US-Based Institutions including:

- Academic institutions
- Non-profits
- Federal agencies or federally-funded R&D centers
- State, local, or tribal agencies
- Startups and small businesses with Federal grants

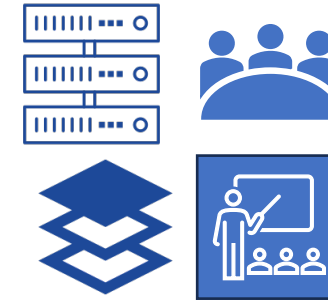
NAIRR Pilot Organization

User Journey


**US-based
Researchers,
Educators &
Students**



<https://nairrpilot.org>



**Pilot Resources
and Opportunities**

The NAIRR Pilot provides infrastructure and resources; it does not fund end-user research.

Operations

**NAIRR
Open**

Enable open AI research and access to diverse AI resources via a central portal and coordinated allocations

**NAIRR
Secure**

Enable AI research needing privacy and security-preserving resources. Assemble exemplar privacy preserving resources.

**NAIRR
Software**

Facilitate use of AI software, platforms, tools and services across platforms

**NAIRR
Classroom**

Reach new communities through education, training, user support and outreach



Governance

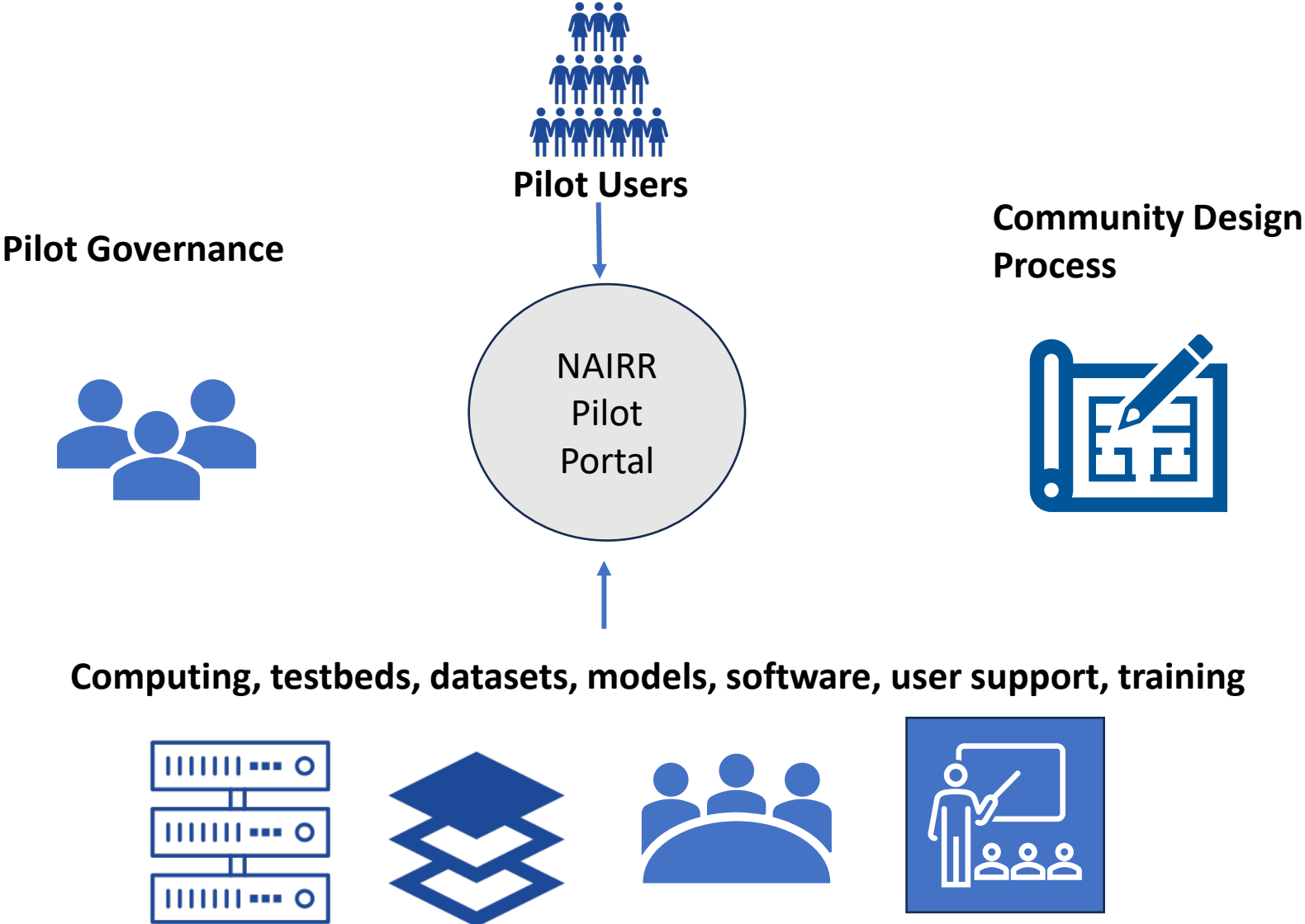


**Community Design
Process**



Initial NAIRR Pilot Components

Note, NAIRR Pilot will provide infrastructure to researchers. The NAIRR Pilot will not fund end-user research.



NSDS Demonstration Project and the NAIRR Pilot are critically connected in the goal of broadening and democratizing access

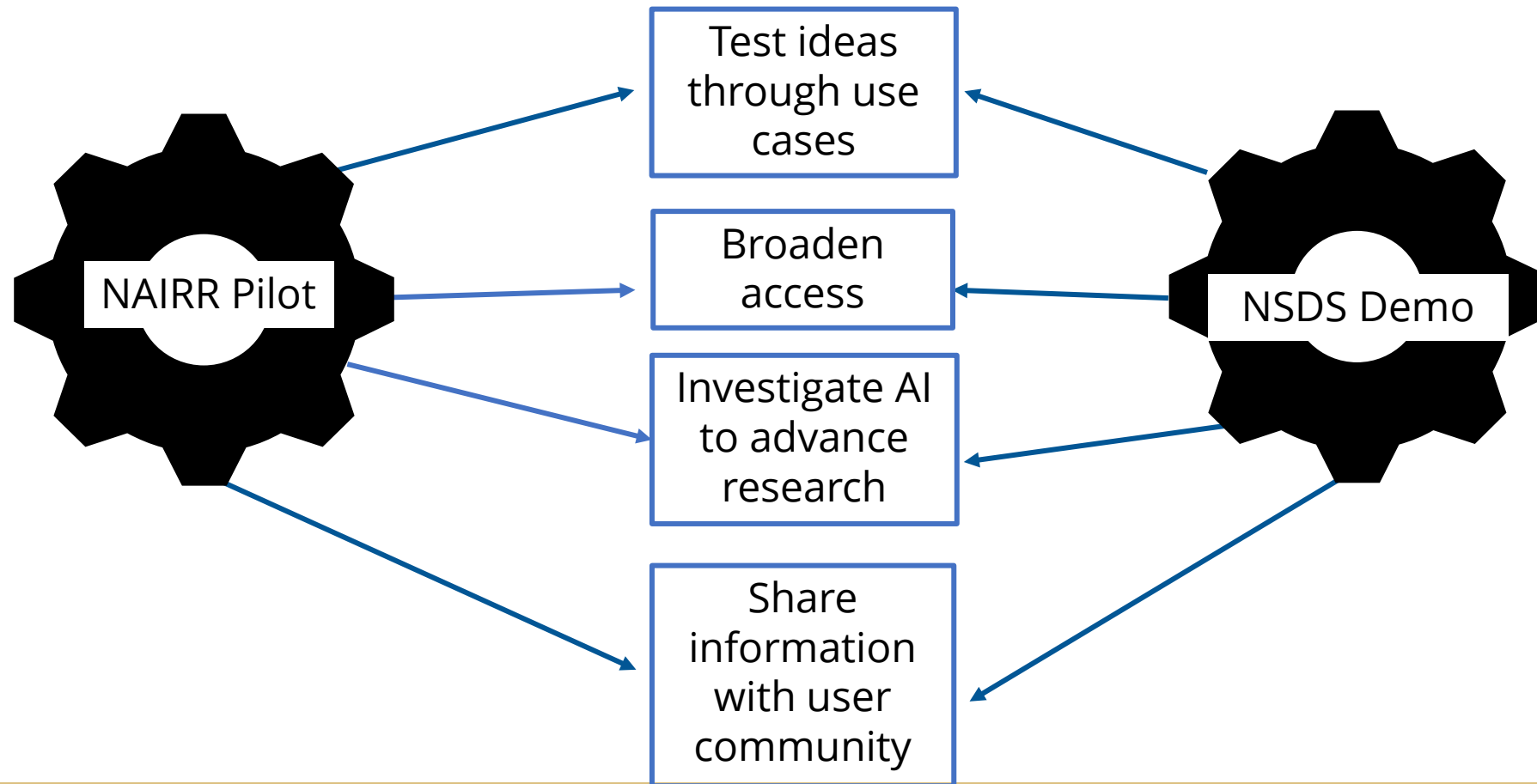
- Provides access to computing, model, software, platforms, datasets and user support
- NAIRR will not 'store' any datasets, but will facilitate discoverable datasets
- Connect users of data to the appropriate compute and other software and tools



- Serve as a central hub for users to discover shared services and resources
- Connect users to toolkits (e.g., data protection and synthetic data)
- Provide support to securely access, link, and analyze powerful, high value data for evidence building

Synergies and Opportunities with NSDS Demonstration and NAIRR Pilot

Synergies and Opportunities with NSDS Demonstration and NAIRR Pilot



Questions and Discussion