Supporting Science and Enhancing Security:

NIH Researcher Auth Service

Internet2 Meeting, May 9, 2023

*Product Owner & Director, NICHD Office of Data Science and Sharing – **Rebecca F Rosen, PhD***

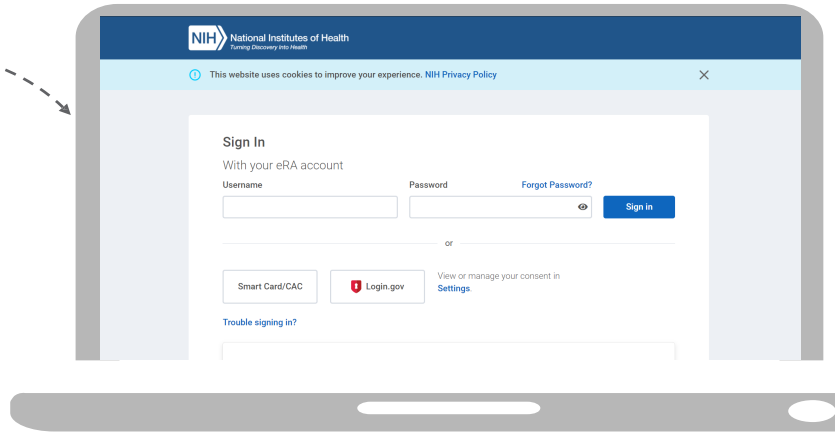
*Hosted by the Center for Information Technology (CIT) Identity, Credential & Access Management (IAM) – **Jeffrey Erickson and Ivor D'Souza***

*Funded by the NIH Office of Data Science Strategy (ODSS) – **Susan Gregurick, PhD***

NIH National Institutes of Health

Researchers use RAS to Access NIH Data & Tools

NIH RAS Login



Internal NIH and External researchers sign in with preferred credentials

Secure access to NIH-funded datasets and tools

A u d i t i n g & L o g g i n g o f E v e n t s

Identity Providers

NIH eRA InCommon LOGIN.GOV

COMING SOON eduGAIN ORCID Google

Key Features

More efficient - Authorizations from NIH dbGaP Data Access Committee (DAC) decisions are centralized and provisioned only upon login

Simple – Log in or link accounts from multiple identity providers for a single sign on experience across systems

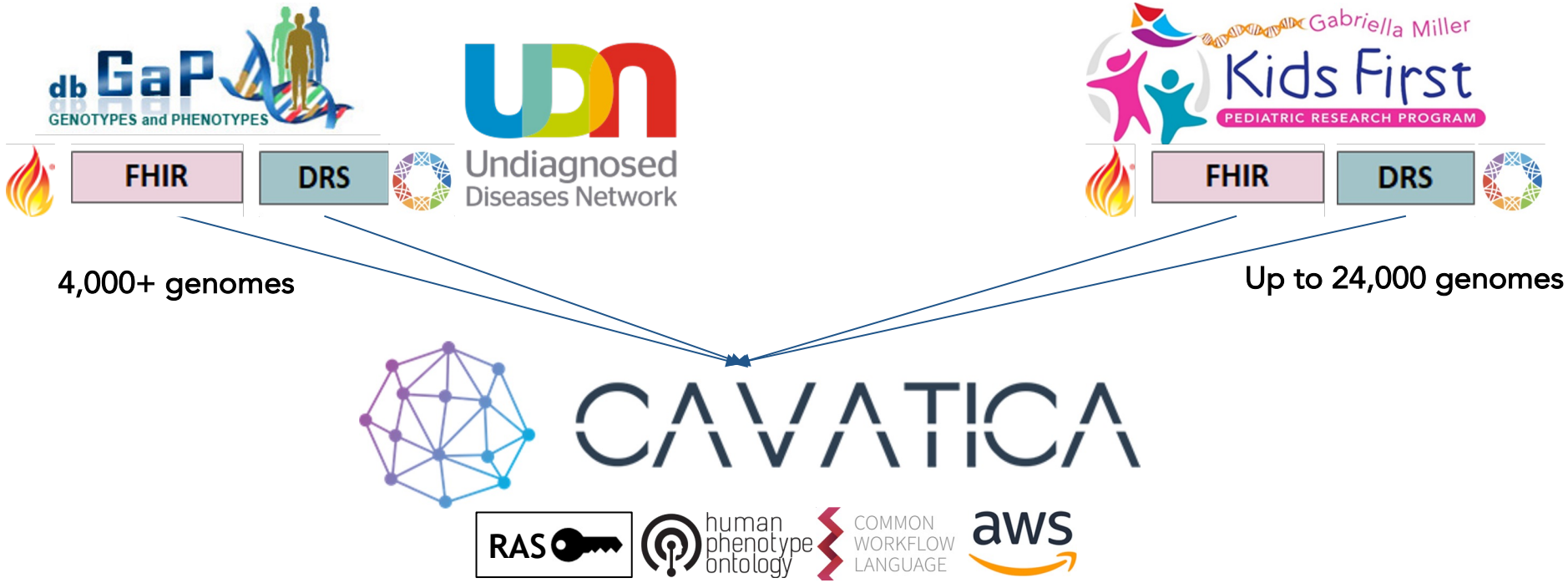
Secure – Use multi-factor authentication (MFA) for data repositories that require a higher level of access security

NIH National Institutes of Health

RAS v1.1 Release: December 2020

Supporting Science:

Researchers can co-analyze genomics and clinical datasets from Kids First & dbGaP, in the CAVATICA cloud platform using NIH RAS for single sign on and authorization



Supporting Science:

Researchers will use NIH RAS to access an NIH-funded CloudLab sandbox to set up tools and data management workflows in the cloud

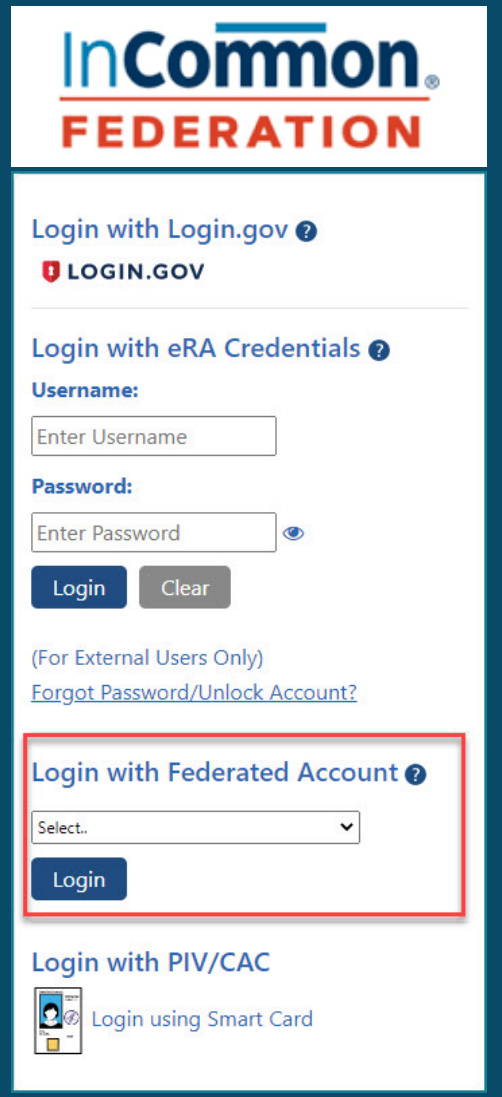
NIH Cloud Lab offers a highly supportive environment to try out cloud capabilities – with customized training designed to support NIH's research mission – and is widely available across the NIH ecosystem of researchers.

Full Access to the Cloud Console

- Deploy a full range of resources
- CPU or GPU VMs
- Managed Jupyter notebooks
- Advanced AI/ML capabilities
- Bioinformatic workflow managers
- Access to compute clusters
- High-speed networking
- Support from Cloud Team & CSPS
- On-demand training

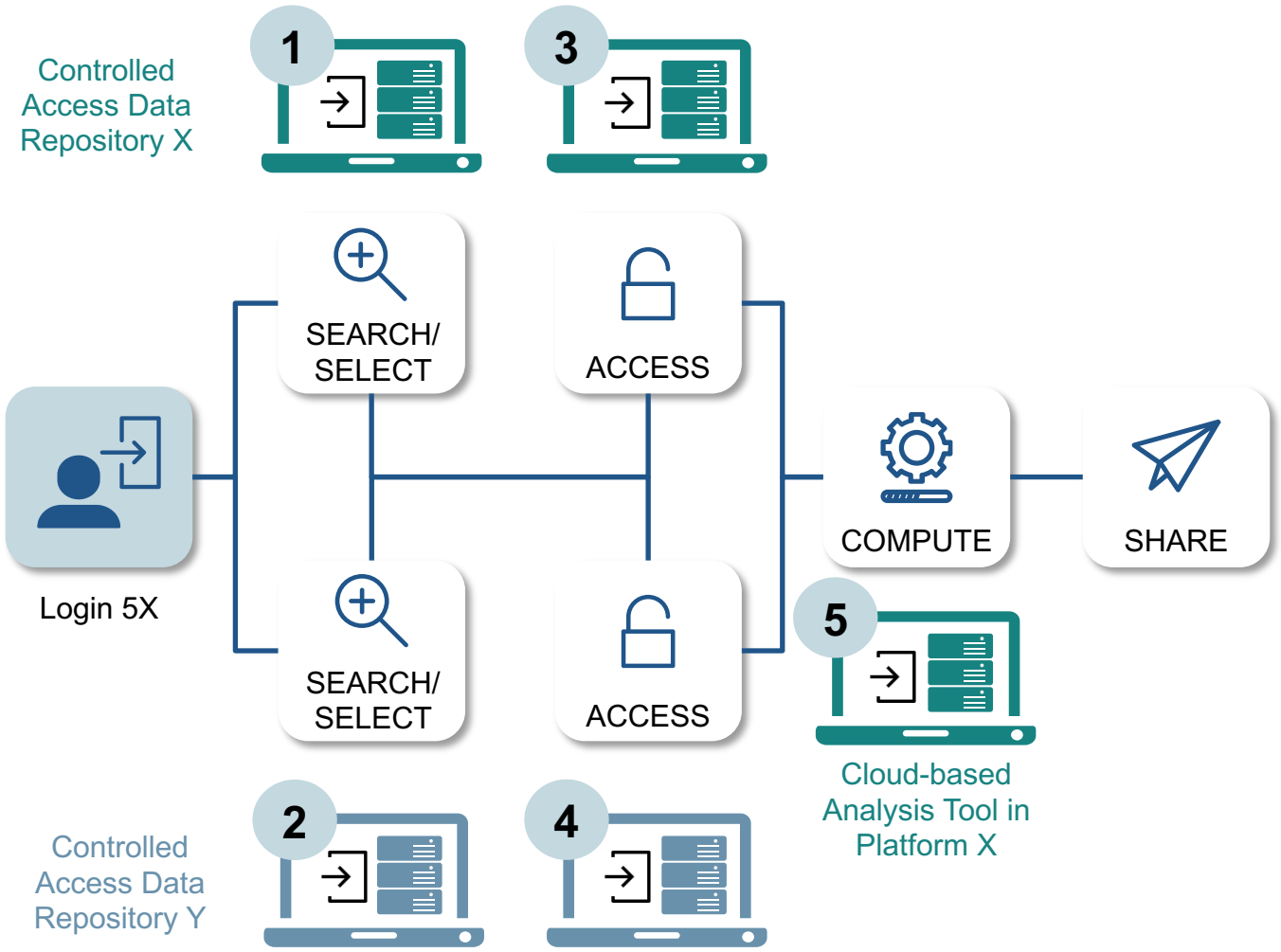
Bioinformatic Tutorials to Speed Uptake

- Variant Calling
- GWAS
- Medical Imaging
- RNA seq
- Single Cell RNA seq
- Proteomics
- Utilizing command line and Jupyter notebook environments
- Using HPC environments in the cloud



cloud.NIH.gov/resources/cloudlab

Before NIH Researcher Auth Service



Key Issues

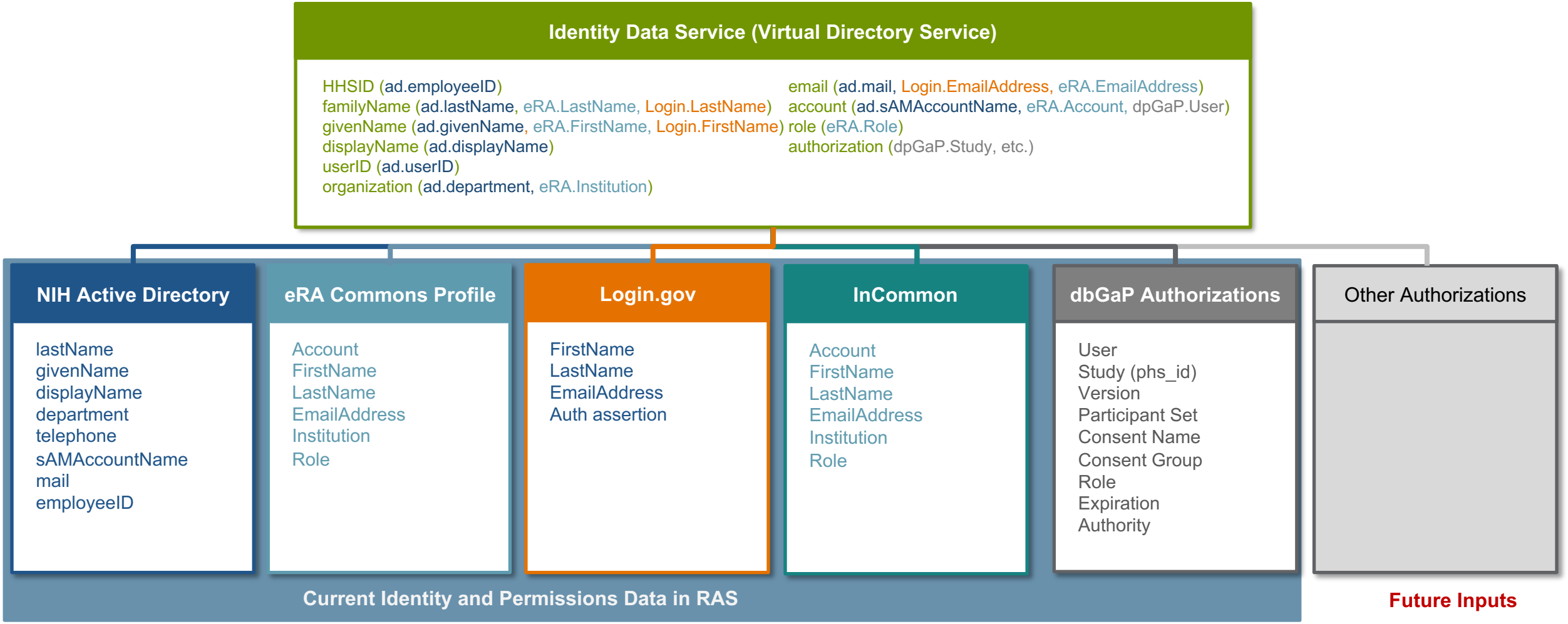
Inefficient – Authorizations from NIH dbGaP Data Access Committee (DAC) decisions replicated in multiple disconnected data repositories

Burdensome – Researchers must maintain separate accounts across multiple resources and are required to sign in over and over

Less secure – Access to controlled-access data via username and password represents risk

Disjointed – Auditing and logging is not standardized across resources

Since 2020, RAS is a Central NIH Identity and Authorization Broker



National Institutes of Health

Supporting Secure Informatics Research Across NIH Data Systems

Jane

Jane User:

- Approved access to datasets in Repositories X and Y
- Wants to analyze both datasets in a tool hosted in Repository X

NIH RAS Login

RAS Tokens State:

- Jane is Jane User
- Jane used multifactor authentication
- Jane is authorized to use datasets in Repositories X and Y

Controlled Access Data Repository X

RAS Token sent to request controlled data on Jane's behalf

Repository Y can trust Repository X:

- Both systems have an Interconnection Security Agreement (ISA) with RAS

NIH National Institutes of Health

NIH Data Systems Integrating with RAS

NIH Data Systems Integrating with RAS

1. Initial Conversations

- NCATS National COVID Cohort Collaborative (N3C)-UNA
- NCI NIH Integrated Data Analysis Platform
- NHLBI LungMap
- NHLBI BDC: RECOVER Mobile Health Data Repository
- NIA Data Enclave
- NIEHS Human Health Exposure Analysis Resource
- NIMH Cell Reprogramming Database Portal
- NIMH Repository and Genomics Resource

2. Planning (Integration request form)

- NCI Cancer Research Data Commons: Milestone 3
- NHLBI Biodata Catalyst: Milestone 3
- NHLBI Biodata Catalyst RECOVER Data DRC
- NHGRI Genomic Data Science Analysis, Visualization, and Informatics Lab-space: Milestone 3
- NHGRI Genomic Data Science Analysis, Visualization, and Informatics Lab-space DUOS
- NIAID Data Ecosystem
- NIAID ImmPort
- NIAID TB Portal
- NIDA Adolescent Brain Cognitive Development – Loris-Instance
- NIMH Data Archive: Phase 2
- NIMH/NIDA NeMo
- OD All of Us Phase 2
- OD KidsFirst Data Resource Center : Milestone 3

NIH Data Systems Integrating with RAS

3. Design (Technical design documentation)

- NICHD Data And Specimen Hub: Phase 3
- NHLBI Bio Data Catalyst RECOVER Data Gateway
- NHLBI Cardiovascular Development Data Resource Center
- NHLBI INCLUDE Data Hub
- NHLBI PCGC-HeartSmart
- OD Common Fund 4D Nucleome
- OD Common Fund The Human BioMolecular Atlas Program
- OD Rapid Acceleration of Diagnostics Data Hub 2.0
- OD RADx-Digital Health-RAPIDS

4. Development

- NCI Cancer Research Data Commons: M1
- NEI Data Commons/Biomedical Research Informatics Computing System

NIH Data Systems Integrating with RAS

5. Testing

- NIA Genetics of Alzheimer's Disease Data Storage Site ADDAPT Cloud Commons

6. Live in Production

- NCBI, Database of Genotypes and Phenotypes Power User Portal
- NCI Cancer Research Data Commons: Milestone 1
- NHGRI Analysis Visualization and Informatics Lab-space : Milestone 1
- NHLBI Biodata Catalyst: Milestone 1
- NICHD Data And Specimen Hub: Phase 2
- NIMH Data Archive: Phase 1
- OD All of Us: Phase 1
- OD Common Fund Data Ecosystem Portal/Globus
- OD KidsFirst Data Resource Center: Milestone 1
- OD KidsFirst/NCBI Database of Genotypes and Phenotypes: Interoperability
- OD Rapid Acceleration of Diagnostics Data Hub 1.0

Ongoing NIH RAS Priorities

SCIENCE

SECURITY

Support the increasingly complex science and security requirements of NIH Data Ecosystems

- For example, one new emerging data ecosystem requires secure single sign on across 7 existing NIH data repositories + 3 new data repositories
- RAS will expand to convey non-genomics controlled access data permissions

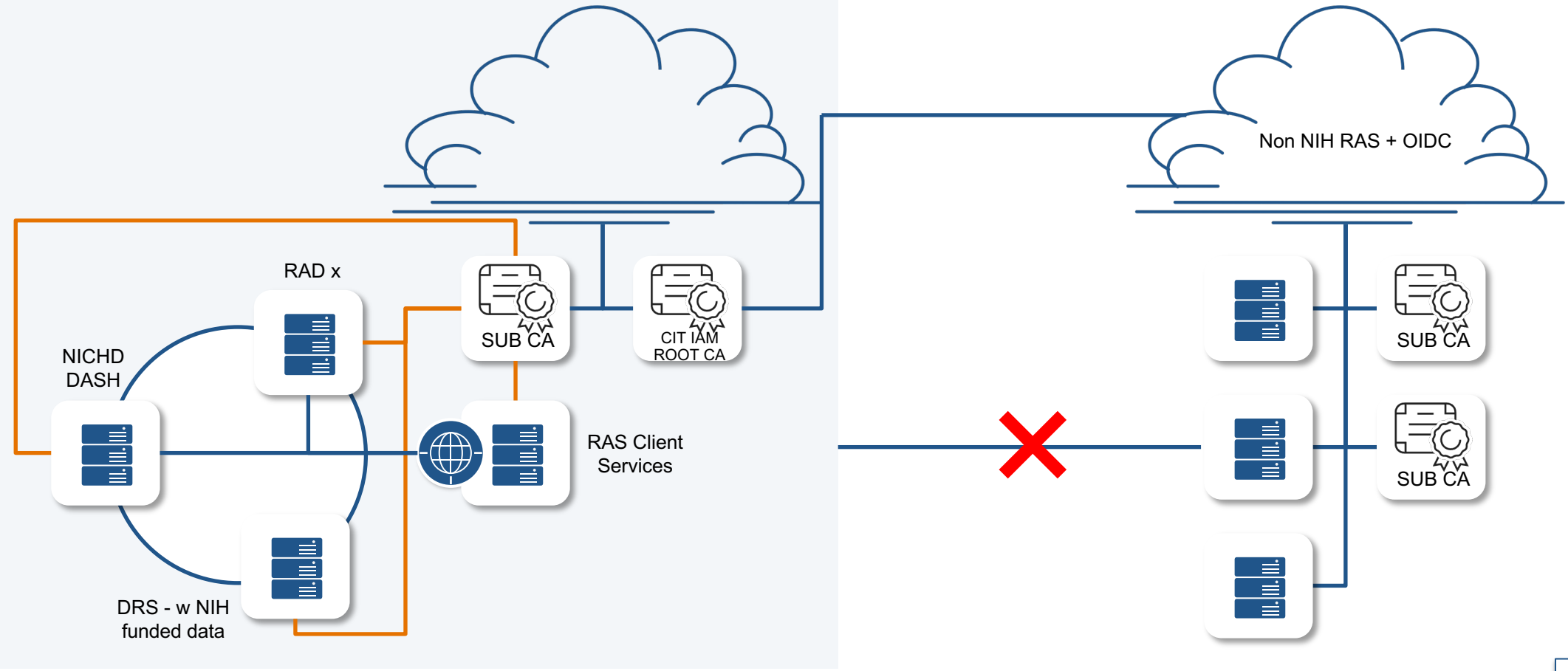
Support new science and security requirements for current and new partner systems

- RAS will support new login account credentials, new dataset permissions, new researcher workflows & enhanced performance needs
- New ODSS team is working with IC funders and partner systems to streamline onboarding and integration processes

Adapt to evolving cybersecurity trends and best practices (Zero Trust)

- Phased implementation of Security Advisory Group's recommendations, which include security design upgrades for both RAS and IC data repositories using RAS

Enhancing Security:
Toward a Federated RAS Ecosystem

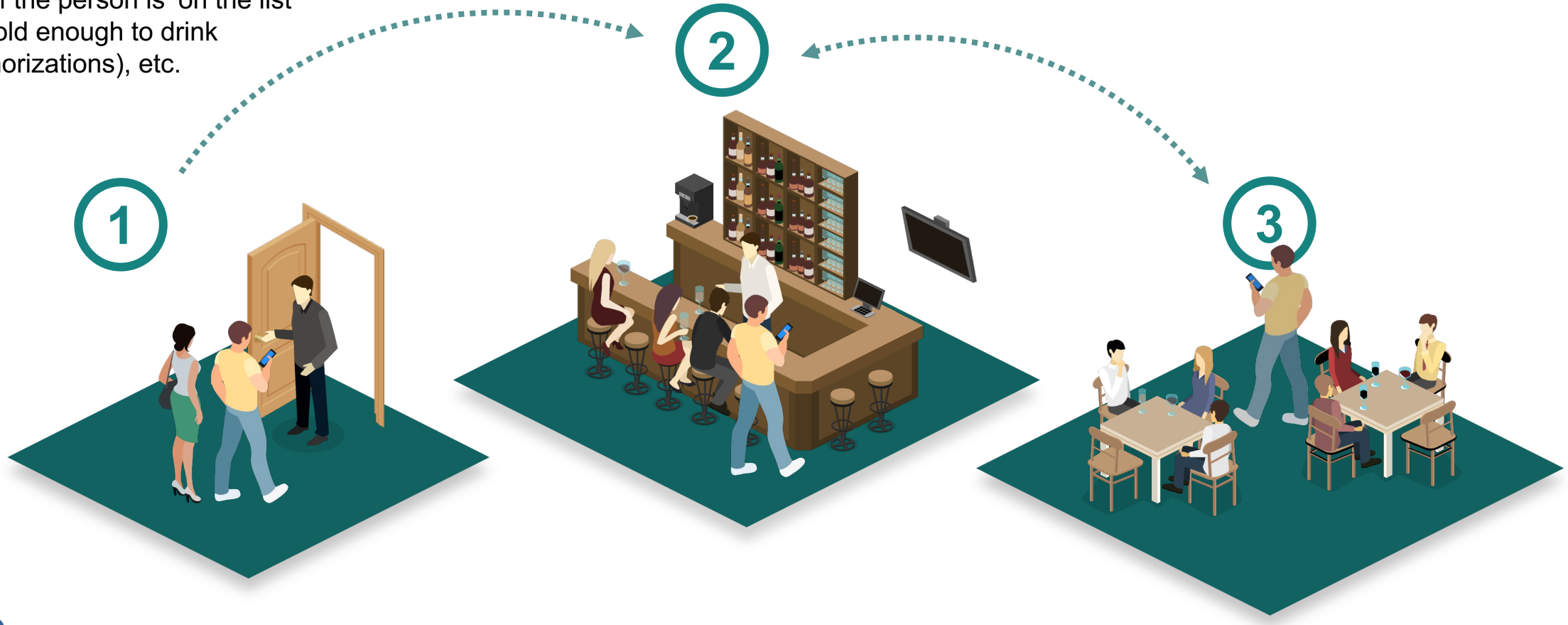


Use of a Private Certificate Authority (CA) allows NIH RAS to establish a closed loop eco system by compartmentalizing security boundaries. Certificates used for client authentication will be issued by NIH RAS CAs to partner systems, DRSs, and NIH RAS token services.

Enhancing Security:
Toward a New NIH Data Ecosystem Security Model (Zero Trust)

Bouncer checks the customer's driver's license (authentication) and if the person is 'on the list and old enough to drink (authorizations), etc.

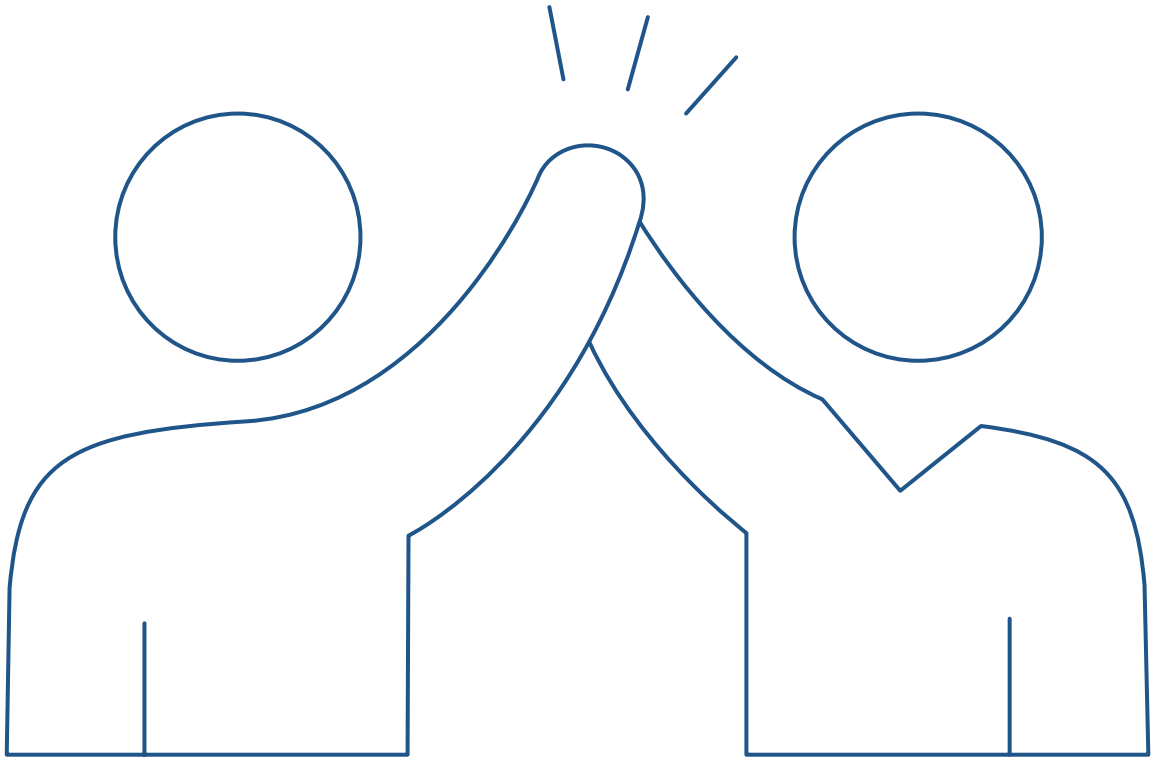
Bartender still requests to check your ID and checks your "VIP" bracelet each time you order a drink at the bar



NIH National Institutes of Health

Acknowledgements and Thanks

- Susan Gregurick & Jeff Erickson
- OCIO/CIT Leadership
- ODSS Technical Working Group
- RAS Development Team
- NIH IC System Owners and PIs/PMs/Development Teams
- NCBI and dbGaP Team
- RAS Governance Team and Security Advisory Group



Questions and Feedback

https://datascience.nih.gov/data-infrastructure/researcher-auth-service
https://auth.nih.gov/docs/RAS/serviceofferings.html

Centralizing Auth Enhances NIH Data System Security

Before RAS:

- Search Portals, data repositories, and platforms each manage their own identity and access management (“auth”) software
  - eRA Commons or other credentials for single factor login
  - Internally managed role-based access to data
  - For controlled genomics data, link eRA account and NIH permissions information (data repository makes decision)

With RAS:

- Repositories **delegate to NIH** important security controls
  - Only NIH can check user identity before a system or data access event and multi-factor authentication is standard
  - Only NIH RAS tokens can be used to access controlled data (NIH makes decision)
- Repositories adhere to security controls in the RAS Interconnection Security Agreement