## NIH CIT and OCIO: Enabling Access and Innovation With Advanced Technology Dr. Sean Mooney, Director, Center for Information Technology



Center for Information Technology & Office of the Chief Information Officer



6/9 NICHD ADVISORY COUNCIL



Computing @NIH

Success Stories

The Cloud and Data Infrastructure

Artificial Intelligence

Biomedical Cyberinfrastructure

The Future



Center for Information Technology & Office of the Chief Information Officer



## **Technology at NIH From the Ground Up**



**Cloud Services and Modular Cyberinfrastructure** 

**Enterprise Infrastructure + Cybersecurity** 



### Enterprise Computing at the NIH Enables Next Gen Biomedicine





Center for Information Technology & Office of the Chief Information Officer

## The People Building the Cyberinfrastructure Ecosystem of the NIH





IC-Specific Cyber Infrastructure.

6/9 NICHD ADVISORY COUNCIL



Center for Information Technology & Office of the Chief Information Officer

# **Success Stories From Inside the NIH**

Here are some (but not all!) of the things we do...



## The NIH Has Significant IT Successes

### The NIH Network



- 45K staff and 700K visitors in 200 NIH buildings across 9 geographic locations rely on the NIH network for secure and reliable data information sharing
- Connected with **4,300 miles** of network cabling

### **NIH Collaboration**

- **2.5+ million virtual meetings** hosted annually in Zoom and Teams
- 1+ billion emails sent and received
- 1+ million GB in Sharepoint and OneDrive

### High Performance Computing

• 100,000+ core supercomputer, BioWulf, with 1+ million CPU hours/month



## **NIH in the CLOUD** | STRIDES



## **The NIH STRIDES Initiative**

- 1. Aligned with NIH's Strategic Plan for Data Science Pursues modern, integrated, intuitive, efficient, secure, and data-driven technology
- 2. Cost savings & efficiencies for the research communityMore usage begets more savings and greater discounts for all
- 3. Strong partnerships with cloud providers

Resulting in collaborative R&D engagements and more focus and support on research











## Data Is Now at the Heart of Enabling the NIH Mission

Genomic and phenotypic Metabolomic Genetic expression and variation Microbiome Cellular structure and function Neuroscience Neuronal image **Data Types** Supported via Structural variant desDES Whole genome sequence Single-cell 'omics Microscopy image Cryo-electron microscopy **Biospecimen** And more...

364+	PETABYTES OF DATA
802M+	C O M P U T E H O U R S
2,742+	A C C O U N T S O N B O A R D E D
\$126M+	C O S T S A V I N G S
5500+	P E O P L E T R A I N E D

## Major NIH & NIH-Funded Research Programs Supported







## **NIH Cloud Lab: Experiment in the Cloud**

#### How It Works:

- 1. Fill out request form
- 2. Get account and \$500 of credits
- 3. Access tailored cloud trainings
- 4. Practice and learn for 90 days



#### **NIH Use Cases**

#### **Evaluate Utility & Cost**

Provides an easy route to evaluate the cloud's utility/cost for a project without major time or financial commitments

#### **Develop New Tools**

Allows experienced teams to prototype new architectures and evaluate software and hardware combinations

#### **Share Ideas**

Connects NIH'ers from across ICs to share ideas on how to conduct biomedical research in the cloud

#### **Learn New Skills**

Simplifies access to tools and cloud environments that participants can use for training purposes

Analysis in NIH Cloud Lab



## Cloud Data Repositories – Standard Platforms to Support AI model development, Model and Data Sharing



## **NIH's AI Investment**





Center for Information Technology & Office of the Chief Information Officer

## **Cloud Data Platforms**

Types of Platforms



**Standardization** 

NIH Research Projects support data within many **Cloud Data Platforms** 

Some tools across these platforms can be standardized

We need more enterprise thinking for those standards to make data and access easier and more impactful

**Enterprise Approach** 

Gen3 Terra Cavatica Foundry BRICS



## An Example: NIH Researcher Auth Service (RAS)



A common task for computer resources is logging into them. Authentication.



Once we log into a resource, we need to be authorized to see certain digital assets (data, web pages, pictures of cats). Authorization.



RAS solves this problem by enabling researchers to not reinvent the wheel and save both time and money using a common toolkit for logging in and access resources.



RAS is another impactful collaboration between ODSS and CIT.

**RAS** is a Lego that contributes to global biomedical cyberinfrastructure.



## **Workspaces Enable Data Integration & Analysis**

- Researchers often need to combine data from multiple repositories to answer questions
- Workspaces are (often cloudbased) analytical environments where researchers can bring together data and tools for analysis
- Workspaces themselves are not data repositories, since they do not govern or distribute data



As with data repositories, **there are many workspaces**, both commercial and open-source, supported by NIH and in use across the biomedical research community

#### 6/9 NICHD ADVISORY COUNCIL



Center for Information Technology & Office of the Chief Information Officer

## A "Plug-and-Play" Future?



How might we expand and enhance NIH's data ecosystem by making connections between repositories and workspaces faster, standardized, and more secure?



## NIH Programs Using Workspaces Today

Many major NIH programs use workspaces to facilitate data analyses



Center for Information Technology & Office of the Chief Information Officer





## **AI Based Prediction and Design of Proteins**

### Ab initio prediction and design wins the 2024 Nobel Prize in Chemistry

 The Nobel Prize in Chemistry 2024

 David Baker
 Demis Hassabis
 John Jumper

 "for computational protein design"
 "for protein structure prediction"
 "for protein structure prediction"

 "for computational protein design"
 "for protein structure prediction"
 "for protein structure prediction"

 Image: David Baker. II. Niklas Elmehed © Nobel Prize
 Image: David Baker. II. Niklas Elmehed © Nobel Prize Outreach
 Image: David Baker. II. Niklas Elmehed © Nobel Prize



### Critical Assessments and Data Challenges are Key to Advancing Data Science

The NIH Funded CASP Experiment, Critical Assessment of Structure Prediction, highlights Importance of unbiased assessment of AI methods.





From CASP in 1994 to AlphaFold in 2021

## **AI Business Cases: TrialGPT**

- TrialGPT is in the clinical space.
- Researchers are developing and evaluating the use of Large Language Models (LLMs) for matching patients to clinical trials.
- TrialGPT will be assessed for practical use by healthcare professionals.
- To assist healthcare professionals in matching patients to relevant clinical trials with improved efficiency.





[1] <u>https://pmc.ncbi.nlm.nih.gov/articles/PMC10418514/</u>
[2] <u>https://www.nih.gov/news-events/news-releases/nih-developed-ai-algorithm-matches-potential-volunteers-clinical-trials</u>



## **AI in NIH Operations**

## **AI in NIH Operations**

### **AI is Operational**

- Al is being implemented as an advanced IT system without a data scientist
- Examples include Office 365
   message management/drafting
   (Co-pilot), meeting minutes,
   secure ChatGPT access
- We must ensure that implementations are effective, safe, legal, ethical, and equitable



### **Benefits to AI Business**

- Business efficiency of common processes
- **Decision making** support
- Summarization of data -> true
   'data to knowledge'
- Interactive chatbots, both text and voice/video avatars
- Benefits of Natural Language
   Processing



Center for Information Technology & Office of the Chief Information Officer

AI Risks

- Al can be **inequitable**
- Data shared with AI may share with the company that provides the AI and may be used to train models
- Al can **impact human rights**
- Al can be **unsafe**
- Al can drift in accuracy over time
- Data privacy issues (i.e., learning data leakage/exfiltration)

## Why AI Oversight is Necessary

- Al is risky for the NIH
  - Webservices, e.g., ChatGPT
  - Data governance and policy is needed for using AI
  - Secure ChatGPT is possible (CHIRP), but users must be trained and attest to not sharing confidential information
- Can impact human rights, funding decisions, etc.
  - For example, a scientist is still today often depicted as an Einstein like man in a lab coat. This is not an average scientist.
  - What if an algorithm prefers to fund grant proposals based on a similar stereotype? This could create inequity.
- Clinical Use of AI might harm patients







## NIH AI Use Cases (From Data Call)

77 of the 195 use cases are non-retired, operational AI not qualified as R&D or productivity COTS. Of the 77:

27	16	68
Relates to Grants Management	Include a Chatbot	Include natural language processing (inclusive of chatbots)
97	9	12
Of the 195 are R&D	of the 195 use cases are non-retired productivity COTS	Are retired use cases



## **AI Business Cases: DMS Classification LLM**

- Researchers are using LLMs to enhance the evaluation of NIH Data Management and Sharing (DMS) Plans, focusing on streamlining the review process.
- The primary objective of the LLM4DMS project is to create AI/LLM **digital-assistant tools**.
- The tools will analyze submitted DMS plans to detect inconsistencies, ensure completeness, and verify alignment with NIH policies and objectives.
- It will streamline reviews, enhance consistency in policy interpretation, and **partially automate tasks** that are currently time-consuming and susceptible to human error.



#### 6/9 NICHD ADVISORY COUNCIL

## **ChIRP Pilot** NIH-secure GenAI Chatbot

#### **Overview:**

- OIR launched a 6-month pilot to test ChIRP, an NIH-secure GenAI environment for exploring large language models
- ChIRP is web based and can summarize documents, generate text/images, write/debug code, and more using GPT-40, DALL-E, and Claude 2.1 models
- Enterprise, secure, and suitable for sharing NIH confidential information
- Development led by Yang Fann with many others and built upon work Nick Asendorf did with the NHLBI Chatbot

#### How to Access:

- 1. Connect to NIH Network or VPN
- 2. Go to http<u>s://chirp.od.nih.gov</u>



### SCAN TO ACCESS CHIRP



or visit <u>https://chirp.od.nih.gov</u> (NIH Network Connection or VPN Required)



## Microsoft 0365 Co-Pilot

Example business process Efficiency

NIH Collaboration and Communication User Group

#### **Proposed Microsoft 365 Copilot Rollout Strategy**



As we begin preparations for Microsoft 365 Copilot, the proposed strategy below highlights the key anticipated milestones.



Provision licenses



## Bringing it All Together: Cyberinfrastructure

6/9 NICHD ADVISORY COUNCIL

## Digital Ecosystem and Cyberinfrastructure

Cyberinfrastructure should be like toy building blocks: **reusable and interoperable.** 





## Many Technologies to Consider And Build in Interoperability

We need to support both the new **data generating tools** and the **platforms** to support their data.



Bioinformatics Workflows



**Components of Cloud Based Data Platforms** 



Specialized Computing: Quantum, GPUs, etc.



LLMs and Generativ<u>e Al</u>



## Future: Building the 'Technical Garden' at the NIH

In research, while we let a thousand flowers bloom -- we are all in one garden

### **Exclusively Within the NIH:**

- Advancing Clinical Research and Trials with Technology
- Develop NIH's Administrative Data Strategy
- Growing and Nurturing Tech Careers @NIH
- Enabling Responsible Access and Use of Operational AI at the NIH
- Enterprising NIH systems in the Cloud



## Future: Building the 'Technical Garden' at the NIH

In research, while we let a thousand flowers bloom -- we are all in one garden

### **Broader Focus on the Biomedical Research Community:**

- Convening and Leading the Biomedical Sciences Technical Community
- o Improving Advanced Tech Fluency in the Global Biomedical Workforce
- Products and Services to Support a Cloud Ecosystem
- Supporting the NIH Core Research Facility Ecosystem

## **How We Communicate**

Please watch our efforts:

• **Read our forthcoming broad newsletter** to serve the NIH community.





## Today's Information Technology, Tomorrow's Cyberinfrastructure

# **Thank You!**