ACD Workgroup Proposal
Dr. Sean Mooney, CIT Director
Workgroup Proposal

- Introduction to CIT and IT at NIH
- Need for Guidance on Information Technology, Cyberinfrastructure, and Cybersecurity
- Proposed Workgroup Logistics
- Discussion
Today, all parts of the biomedical research life cycle contain data, computers, software, analytics, and their connectivity.

How can we cultivate the most advanced technology and the most capable workforce to tackle such a broad mission?
Digital NIH

- Strategic plan and visionary roadmap for our shared future in technology
- Developed through a committee and intended for 2023-2028
- A great starting point for the future of CIT

Co-Chairs
- Andrea Norris (OD/CIT)
- Patricia Flatley Brennan (NLM)

Committee Members
- Jill Barnholtz Sloan (NCI)
- Andy Baxevanis (NHGRI)
- Raymond Dillon (OD)
- Miles Fabian (NIGMS)
- Inna Faenson (OER)
- Gregory Farber (NIMH)
- Greg Germino (NIDDK)
- Darla Hayes (OD)
- Dyung Le (OD)
- Janice Lee (NIDCR)
- Colleen McGowan (ORS)
- Elaine Ostrander (NHGRI)
- Kate O’Sullivan (NHLBI)
- Taunton Paine (OD)
- Kim Pruitt (NLM)
- Rebecca Rosen (NICHD)
- Jeff Shilling (NCI)
- Xavier Soosai (CIT)
- Michael Tartakovsky (NIAID)
- Alastair Thompson (NHLBI)
Digital NIH

Innovation, technology, and computation for the future of NIH

Digital NIH Framework
Modern, integrated, intuitive, efficient, secure, and data-driven technologies

- Extramural Research Mgmt.
- Intramural Basic Research
- Intramural Clinical Research
- Administration & Management
Digital NIH

Innovation, technology, and computation for the future of NIH

Digital NIH Framework
Modern, integrated, intuitive, efficient, secure, and data-driven technologies

Extramural Research Mgmt.
Intramural Basic Research
Intramural Clinical Research
Administration & Management

Common architecture with well-defined standards to enable integration
Digital NIH

Innovation, technology, and computation for the future of NIH

Digital NIH Framework
Modern, integrated, intuitive, efficient, secure, and data-driven technologies

- Extramural Research Mgmt.
- Intramural Basic Research
- Intramural Clinical Research
- Administration & Management

Common architecture with well-defined standards to enable integration

Innovative, cutting-edge storage, analytics, and computational infrastructure
Digital NIH

Innovation, technology, and computation for the future of NIH

Digital NIH Framework
Modern, integrated, intuitive, efficient, secure, and data-driven technologies

- Extramural Research Mgmt.
- Intramural Basic Research
- Intramural Clinical Research
- Administration & Management

- Common architecture with well-defined standards to enable integration
- Innovative, cutting-edge storage, analytics, and computational infrastructure
- Increased technically competent workforce
Digital NIH

Innovation, technology, and computation for the future of NIH

Digital NIH Framework
Modern, integrated, intuitive, efficient, secure, and data-driven technologies

- Extramural Research Mgmt.
- Intramural Basic Research
- Intramural Clinical Research
- Administration & Management

- Common architecture with well-defined standards to enable integration
- Innovative, cutting-edge storage, analytics, and computational infrastructure
- Increased technically competent workforce
- Technology to support anywhere, anytime workplace of the future
Digital NIH

Innovation, technology, and computation for the future of NIH

Digital NIH Framework
Modern, integrated, intuitive, efficient, secure, and data-driven technologies

- Extramural Research Mgmt.
- Intramural Basic Research
- Intramural Clinical Research
- Administration & Management

- Common architecture with well-defined standards to enable integration
- Innovative, cutting-edge storage, analytics, and computational infrastructure
- Increased technically competent workforce
- Technology to support anywhere, anytime workplace of the future
- Risk-based, embedded cybersecurity protections
The CIT View of IT @ NIH

- Cloud Technologies
- High-Performance Computing
- Applications
- Research Computing
- Identity & Access Management
- Network
- Cyberinfrastructure
- Collaboration
- Cybersecurity
We need enterprise-level thinking around FAIR data... and I need your help!

**FINDABLE**
- iD Persistent Identifiers (PIDs)
- # Rich Metadata
- Indexed Data Repositories
- π PIDs in Metadata

**ACCESSIBLE**
- Share Standard Comm. Protocol
- Open & Free Protocol
- Authentication Where Necessary
- ∞ Metadata Always Available

**INTEROPERABLE**
- Vocabulary Vocabularies are FAIR
- Linked Metadata

**REUSABLE**
- Metadata with Multiple Attributes
- Usage License
- Provenance
- Community Standards
“New directions in science are launched by **new tools** much more often than by new concepts.”

Freeman Dyson, Theoretical Physicist

We need to support both the new data generating tools and the platforms to support their data.
We Need Intelligent Investments

- Standards for platform interoperability
- AI/data science platforms on controlled access data
- A balance of innovation, harmonization, and cybersecurity
- Efficiencies
- Improved IT governance
Mission-Based Approach to Cybersecurity

Security is a balance. It needs to be strong enough to keep us safe and light enough not to interfere with our work, like a hard hat.

More effective MISSION
Open environment for research, collaboration, and clinical care

Stronger SECURITY
Confidentiality, integrity, availability, and trust

WHAT WE PROTECT
Life & Safety
Personally Identifiable and/or Personal Health Information
Research Data Integrity
Financial Stewardship
Administration and Operations
Clinical Research Information Systems and Cyberinfrastructure for the NIH Campus

There are lots of opportunities in clinical research IT and cyberinfrastructure.

- Supporting Clinical Trials
- Data Resources and Data Governance
- AI and Data Science Capabilities
- Translation of Research Back to the EHR
We have many axes to contemplate improving:

- Many REDCap instances across campus
- (To my knowledge) no Enterprise Clinical Trials Management System (CTMS) or EHR/CRO integration
- Few enterprise tools for patient eligibility identification
- Opportunities to improve enterprise data systems and data governance of intramural clinical research data
- Opportunities to integrate research into the EHRs including the Clinical Center and affiliates
Success Stories

A look at some (but not all!) of the things we do and how they benefit NIH
The NIH Network

- **4,300+** miles of network capacity
- **65,000+** devices connected
- **100 Gbps** network infrastructure
- **30,000** connections a day
- **150 TB** of Internet traffic a day

<table>
<thead>
<tr>
<th>Equipment Type</th>
<th>Quantity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Provider-Grade Routers</td>
<td>50</td>
</tr>
<tr>
<td>Switches</td>
<td>3,000</td>
</tr>
<tr>
<td>Access Points</td>
<td>9,000</td>
</tr>
<tr>
<td>Firewalls</td>
<td>200</td>
</tr>
<tr>
<td>Uninterruptible Power Supplies</td>
<td>1,100</td>
</tr>
<tr>
<td>Servers</td>
<td>50</td>
</tr>
</tbody>
</table>
Enabling NIH-ers to Work from Anywhere

In FY 2023:

- More than **2.5M** virtual meetings
- More than **1B** emails sent and received
- More than **1,000,000 GB** in OneDrive and SharePoint

<table>
<thead>
<tr>
<th>Service</th>
<th>Webex</th>
<th>Zoom</th>
<th>eFax</th>
<th>SharePoint Storage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Meetings Organized</td>
<td>176,856</td>
<td>466,096</td>
<td>3,782,184</td>
<td>265 TB</td>
</tr>
<tr>
<td>Meetings Participants</td>
<td>1,129,868</td>
<td>3,226,990</td>
<td>357,929</td>
<td></td>
</tr>
<tr>
<td>Zoom Webinar</td>
<td>1,268</td>
<td>1,808,768</td>
<td>107,699,158</td>
<td>777 TB</td>
</tr>
<tr>
<td>Meetings Organized</td>
<td>172,196</td>
<td>5,494,315</td>
<td>633,206,257</td>
<td></td>
</tr>
<tr>
<td>Meetings Participants</td>
<td>1,268</td>
<td>1,808,768</td>
<td>107,699,158</td>
<td></td>
</tr>
<tr>
<td>Total Email Sent</td>
<td>107,699,158</td>
<td>633,206,257</td>
<td>999,153,779</td>
<td></td>
</tr>
<tr>
<td>Total Email Read</td>
<td>633,206,257</td>
<td>999,153,779</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total Email Received</td>
<td>999,153,779</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Supporting Growing Demand for High-Performance Computing

Biowulf, the NIH supercomputer, is:

- **105,000** compute cores
- **828** graphics processing units (GPUs)
- **50** PB storage
- **~1,000** biomedical applications supported
- **100** Gbps connection to the NIH network
- **5,050+** scientific papers supported
- Responsible for sequencing the last **8%** of the human genetic code
STRIDES | A partnership with cloud vendors to provide low-cost access

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

- 260+ Petabytes of data
- 606M+ Compute hours
- 2,225+ Research programs
- $92M+ Cost savings
- 5500+ People trained

as of February 29, 2024
Building for the Future

• In biomedical science, we spend 100s of millions each year unnecessarily due to duplicative platforms we choose to support.

• This is expensive. Not interoperable. Not sustainable.

• Having one platform for all data is obviously (I think) unrealistic.

• However, we can provide a common toolkit to make these platforms more efficient, interoperable, sustainable, and impactful.
Digital Ecosystem and Cyberinfrastructure

Cyberinfrastructure should be like Legos: reusable, interoperable, and open.
An Example: NIH Researcher Authentication Service (RAS)

A common task for computer resources is to log in to 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 accessing resources.

RAS is another impactful collaboration between ODSS and CIT.

**RAS is a Lego that contributes to global biomedical cyberinfrastructure.**
The People Building the Cyberinfrastructure Ecosystem of the NIH

NIH Drives Biomedicine and Contributes to IC-Specific Cyberinfrastructure.
AI Platforms Strategy

We have a heavy use of foundation models (e.g., ChatGPT, others) and generative AI.

There is a community of practice for AI at the NIH and many area-specific strategic planning efforts.

There are many needs, such as secure/private AI environments, workspaces to bring disparate data together, etc.

This is an important area for discussion and advice is needed based on the work group’s experience and needs.
How You Can Help

We need a **roadmap** for Cyberinfrastructure and AI/data science platforms.

Thus, we need **your input** on our strategy and direction.

This will have a wide impact for **both intramural and extramural** and, I expect, will be welcomed.
Thank You!
• Provide critical input on NIH IT and cybersecurity governance;
• Articulate high-priority areas for NIH investment in cyberinfrastructure, cybersecurity, AI, and data science;
• Validate or provide additional insight on proposed strategies to support new data generating tools and platforms;
• Evaluate established or proposed platforms for platform interoperability;
• Identify gaps, challenges, and opportunities on issues related to national biomedical cyberinfrastructure;
• Evaluate and potentially expanding cloud-based tools to expand or enhance access to rich datasets and advanced computational infrastructure, tools, and services;
• Identify and promoting solutions to challenges in training, recruitment, and broadening the technical workforce in biomedical research; and
• Provide insight and feedback on risk-based cybersecurity protections.
The ITCC Working Group of the ACD will:

- Seek broad input from IT and scientific communities in public and private sectors;
- Include perspectives from those with experience in information technology, cyberinfrastructure, and cybersecurity in a research or clinical environment;
- Establish a workgroup of 1-2 internal members and 6-10 external members (external to NIH and the ACD);
- Hold quarterly or bi-annual 90-minute meetings to provide feedback to efforts made on strategies and direction; and
- Deliver an interim report within 12 months; a final report within 18-24 months.