From Infrastructure to Impact: Developing Actionable Solutions and Community Challenges in Biomedical Cyberinfrastructure*

Sean D. Mooney†
Center for Information Technology, National Institutes of Health, Bethesda, Maryland, USA
Email: sean.mooney@nih.gov

Nick Weber
Center for Information Technology, National Institutes of Health, Bethesda, Maryland, USA
Email: nick.weber@nih.gov

Timothy J. Hohman
Vanderbilt University, Nashville, Tennessee, USA
Email: timothy.j.hohman@vumc.org

Sarah Biber
Washington University in St. Louis, St. Louis, Missouri, USA
Email: biber@wustl.edu

Biomedical research increasingly depends on distributed data resources, cloud services, high-performance computing, and artificial intelligence (AI). Translating these capabilities into scientific impact requires cyberinfrastructure that is open and interoperable. The 2026 Pacific Symposium on Biocomputing (PSB) workshop examined the evolving cyberinfrastructure at the National Institutes of Health and established a foundation for discussing reusable infrastructure needs and building blocks and barriers to their adoption. The proposed PSB 2027 workshop, "From Infrastructure to Impact," builds on that foundation by moving from awareness to structured prioritization, solution development, and community pilot framing. The workshop will address three interconnected themes: AI-enabled biomedical infrastructure, secure and interoperable computing environments, and government-academic partnerships and ecosystem alignment. Short invited talks, collaborative breakout activities, and plenary synthesis will guide participants in developing three to four well-scoped community pilots with concrete solution pathways and measurable milestones over a two- to five-year horizon. Expected outputs include prioritized problem statements and a structured community summary or white paper. By linking infrastructure needs to actionable collaborative efforts, the workshop aims to inform sustained cross-sector engagement beyond PSB.

Keywords: Biomedical cyberinfrastructure; artificial intelligence; interoperability; secure computing; community pilots; government-academic partnerships.

* This work is supported by the Center for Information Technology (CIT), National Institutes of Health (NIH), with special thanks to Jennie Larkin, National Institute on Aging (NIA), and Susan K. Gregurick, NIH Office of the Director, Office of Data Science Strategy (OD/ODSS).

† Corresponding author.

© 2026 The Authors. Open Access chapter published by World Scientific Publishing Company and distributed under the terms of the Creative Commons Attribution Non-Commercial (CC BY-NC) 4.0 License.

1. Introduction

Biomedical discovery increasingly depends on integrating complex biomedical data (e.g., genomic, clinical, imaging, digital, and real-world) across institutional and computational boundaries. Artificial intelligence (AI) and large-scale data analytics expand research opportunities but also increase demands on supporting infrastructure. Open, standards based approaches to scalable biomedical research computing remain central challenges.[1] Addressing them requires coordinated effort rather than isolated tools.

The PSB 2026 workshop, "The Evolving Cyberinfrastructure at the National Institutes of Health to Support Data and AI in Biomedical Research," introduced reusable infrastructure needs and building blocks for cloud access, researcher authentication and authorization, and secure analytics workspaces.[1] Its accompanying program included perspectives on data platforms and harmonized resources for neurodegenerative disease research.[2] The PSB 2027 workshop, From Infrastructure to Impact, extends this foundation by moving from infrastructure needs awareness to recommended solutions and and proposed community pilots. This article describes the workshop's scope, planned activities, and anticipated outputs.

2. From Infrastructure Awareness to Coordinated Action and Impact

Existing programs provide starting points for coordinated action. NIH's Science and Technology Research Infrastructure for Discovery, Experimentation, and Sustainability (STRIDES) initiative is the NIH's enterprise solution to the cloud and reduces barriers to commercial cloud services[3] The Researcher Auth Service (RAS) supports authentication, authorization, and audit logging.[4] Finally, secure workspaces bring analysis capabilities closer to data[1]. These platforms use Global Alliance for Genomics and Health (GA4GH) standards[5] and the Findable, Accessible, Interoperable, and Reusable (FAIR) principles[6] as a "North Star" for responsible sharing and reuse.

The workshop will convene experts from the meeting to synthesize lessons across data platforms, cloud services, AI, and high-performance computing (HPC). The workshop will be largely interactive with interdisciplinary breakout teams developing tractable solutions to shared infrastructure needs. The proposed solutions and associated proposed community pilots will be summarized in a structured community report or white paper.

3. Workshop Scope and Themes

The workshop focuses on cross-cutting cyberinfrastructure technology challenges rather than domain-specific scientific problems. It is organized around three themes, each framed by a core question that anchors the invited talks and breakout discussions. Within each theme, breakout groups will define a specific problem or opportunity, map a solution pathway, and develop a community pilot or engagement framework that could build on existing initiatives.

3.1. *AI-enabled biomedical infrastructure*

This theme addresses AI as durable infrastructure within national and institutional digital ecosystems with a focus on embedding AI using standards. Topics include operationalizing AI at scale; open and responsible model sharing; benchmarking and evaluation; and workflow acceleration and automation. The central question is how to embed AI using standards. Proposed approaches may address reusable interfaces and workflow integration. The National Institute of Standards and Technology (NIST) AI Risk Management Framework provides a voluntary reference for discussing trustworthiness and responsible implementation.[7]

3.2. *Secure and interoperable computing environments*

This theme examines technical mechanisms for secure, advanced computation across sensitive and heterogeneous environments. Topics include federated learning and distributed AI; portable and containerized workflows; encryption and privacy-enhancing technologies; researcher identity, authentication, and authorization; and secure cross-platform interoperability. The central question is how to enable advanced computation across distributed data environments while maintaining strong identity, access control, and security. Existing standards provide building blocks: RAS supports authentication and authorization[4]; GA4GH's Data Repository Service (DRS) standardizes data retrieval[8]; Health Level Seven (HL7) Fast Healthcare Interoperability Resources (FHIR) supports data exchange[9]; and Common Workflow Language (CWL) describes portable tools and workflows.[10] Solutions should remain open, standards-based, and community driven.

3.3. *Government-academic-industry partnerships and ecosystem alignment*

This theme addresses the partnership, policy, and incentive structures needed to build and sustain shared infrastructure across institutions and sectors. Topics include coordination models for shared infrastructure, alignment with federal standards, governance of controlled-access and clinical real-world data, partnership models for co-developing tools and platforms, and incentives to build, sustain, and share infrastructure. The central question is how government, academic, and industry partners can jointly build, govern, and sustain shared cyberinfrastructure that serves the broader biomedical research community. The NIST Cybersecurity Framework 2.0 supports assessment, prioritization, and communication of security outcomes.[11] Breakout groups will consider technical plans alongside implementation and adoption, including key requirements and dependencies.

4. Workshop Organization and Community Pilots

The half-day workshop combines short invited talks, structured breakout activities and report-outs, and plenary synthesis. Speakers and participants will represent academia, federal, and institutional infrastructure programs, with attention to diversity across institutions and career stages.

4.1. *Invited talks and shared problem framing*

Invited talks of 10-15 minutes will present infrastructure limitations, opportunities, and successes; identify technical, operational, or partnership gaps; and articulate community pilots aligned with the three themes. These speakers will share perspectives to form a basis for problem prioritization:

- Travis Wheeler, PhD; University of Arizona
- Kenneth D. Mandl, MD, MPH; Boston Children's Hospital
- David Haussler, PhD; University of California Santa Cruz
- Roxana Daneshjou, MD, PhD, FAAD; Stanford University

4.2. *Structured problem prioritization*

Within themes, participants will converge on a limited set of high-impact, tractable problems. Proposed considerations include relevance across research settings and opportunities for coordinated action. These considerations support discussion rather than prescribing a solution.

4.3. *Breakout groups: solution development and pilot design*

Each breakout will develop a solution pathway and a community pilot, specifying feasibility and measurable milestones over a two- to five-year horizon. Groups will then share back their solutions and pilots with the full group. Table 1 proposes reporting fields to support comparison and synthesis.

Table 1. Proposed breakout reporting fields for solution and pilot development.

Field	Content to be developed by participants
Problem and scope	Shared problem, intended beneficiaries, and boundaries.
Solution pathway	Technical mechanisms, governance, and existing capabilities.
Feasibility and milestones	Barriers, dependencies, resources, and measurable two- to five-year milestones.
Stakeholders and partnerships	Collaborators, alignment opportunities, and sustained support.
Community pilot and success criteria	Bounded pilot or framework, links to existing initiatives, anticipated impact, and success criteria.
Advancement and next steps	Research, infrastructure, policy, or partnership pathways beyond PSB.

4.4. *Plenary share-back and synthesis*

Groups will present their feedback. A moderated summary from the leaders will seek convergence on three to four priority challenges. Related proposals may be aligned around shared capabilities or stakeholders, with next steps outlined for engagement beyond PSB.

5. Expected Outcomes and Community Impact

Expected outputs are three to four prioritized problems with solution pathways, community pilots, and success criteria tied to measurable two- to five-year milestones. A structured community summary or white paper will document next steps for researchers and infrastructure providers. These are planned outputs, not completed deployments or validated solutions; implementation will depend on resources, governance, and partner commitments. By connecting computational innovation with shared infrastructure needs, the workshop aims to spark new cross-sector partnerships and a community that keeps working together after PSB.

6. Conclusion

Building on PSB 2026, this "From Infrastructure to Impact" workshop will move from awareness to coordinated action across AI-enabled infrastructure, secure and interoperable computing, and government, academic, and industry partnerships. Its intended contribution is a focused set of actionable community pilots connecting computational advances to the infrastructure, governance, and collaborations needed for sustained biomedical impact.

References

1. Ramwala, O. A., Weber, N. and Mooney, S. D. The Evolving Cyberinfrastructure at the National Institutes of Health to Support Data and AI in Biomedical Research. Pac. Symp. Biocomput. 31, 859-864 (2026). doi:10.1142/9789819824755_0064.
2. NIH Workshop at PSB 2026. Workshop overview and program. https://weber-nick.github.io/psb-2026/ (accessed October 5, 2026).
3. National Institutes of Health. NIH STRIDES Initiative. https://cloud.nih.gov (accessed October 5, 2026).
4. National Institutes of Health. Researcher Auth Service Initiative. https://datascience.nih.gov/researcher-auth-service-initiative (accessed October 5, 2026).
5. Rehm, H. L. et al. GA4GH: International policies and standards for data sharing across genomic research and healthcare. Cell Genom. 1, 100029 (2021). doi:10.1016/j.xgen.2021.100029.
6. Wilkinson, M. D. et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci. Data 3, 160018 (2016). doi:10.1038/sdata.2016.18.
7. Tabassi, E. Artificial Intelligence Risk Management Framework (AI RMF 1.0). NIST AI 100-1 (2023). doi:10.6028/NIST.AI.100-1.
8. Global Alliance for Genomics and Health. Data Repository Service (DRS). https://www.ga4gh.org/product/data-repository-service-drs/ (accessed October 5, 2026).
9. Health Level Seven International. FHIR Release 5, version 5.0.0. https://hl7.org/fhir/R5/index.html (accessed October 5, 2026).
10. Common Workflow Language Project. Common Workflow Language (CWL). https://www.commonwl.org/ (accessed October 5, 2026).
11. Pascoe, C., Quinn, S. and Scarfone, K. The NIST Cybersecurity Framework (CSF) 2.0. NIST CSWP 29 (2024). doi:10.6028/NIST.CSWP.29.