

**U.S. Department of Health and Human Services (HHS)  
National Institutes of Health (NIH)  
Office of the Director (OD)  
Division of Program Coordination, Planning, and Strategic Initiatives (DPCPSI)**

**Office of Data Science Strategy (ODSS)**

**Artificial Intelligence–Readiness and Workforce Development  
Closeout Meeting**

**October 24 and 31, and November 1, 2022**

**Virtual**

**Final Executive Summary**

**Introduction and Purpose**

This series of three meetings, sponsored by the NIH Office of Data Science Strategy (ODSS), invited awardees of the Notice of Special Interest (NOSI) for Administrative Supplements for Workforce Development at the Interface of Information Sciences, Artificial Intelligence (AI) and Machine Learning (ML), and Biomedical Sciences ([NOT-OD-21-079](#)) and the NOSI for Administrative Supplements to Support Collaborations to Improve the AI/ML-Readiness of NIH-Supported Data ([NOT-OD-21-094](#)) to a virtual closeout principal investigators (PIs) meeting. The purpose was to provide awardees an opportunity to showcase their work and share their experiences with NIH and other colleagues working in AI and AI-workforce development. This cohort of NIH-funded investigators has laid the foundation to support the next wave of biomedical discovery empowered by AI. Detailed descriptions of the funded projects for [AI-Readiness](#) and [AI-Workforce](#) can be found on the ODSS website.

Three main sessions composed the joint closeout event. Meeting 1—Closeout Discussions in AI Workforce Development—on October 24, 2022, convened the AI Workforce Development cohort. Meeting 2—Closeout Discussions in AI Readiness—on October 31, 2022, convened the AI Readiness cohort. Meeting 3—Joint Closeout—on November 1, 2022, consisted of bidirectional conversations between the AI Readiness and AI Workforce Development groups. The agendas included presentations from ODSS and other experts in the field; lightning talks describing funded projects; breakouts to better understand common interests, challenges, and themes; and discussions on envisioning the future for AI in biomedicine. ODSS staff and other representatives from co-sponsoring NIH institutes and centers (ICs) attended the breakout sessions. More than 100 participants, including recipient team members and federal observers, registered and attended.

**Summary of Meeting**

On October 24 and October 31, 2022, Dr. Laura Biven, ODSS Lead, Integrated Infrastructure and Emerging Technologies, provided an update on AI at NIH and conveyed that the U.S. Department of Health and Human Services (HHS) is leveraging national initiatives (e.g., the National AI Initiative Act of 2020) and AI to solve previously unsolvable problems by continuing to lead advances in the health and well-being of the American people, responding to the use of AI across the HHS ecosystem, and scaling trustworthy AI adoption across HHS. Workshop participants were reminded that the NIH Advisory Committee to the Director Working Group on AI outlined opportunities to fuse AI/ML with exponential increases in biomedical data. They identified three focus areas: data, people, and ethics. Additionally, the

[\*NIH Strategic Plan for Data Science\*](#) outlines a vision of a biomedical data ecosystem that is modernized; integrated; and findable, accessible, interoperable, and reusable (FAIR). As such, ODSS views AI as cutting across multiple areas: data infrastructure, data ecosystem, tools and analytics, community engagement, and workforce development. Two programs—Artificial Intelligence/Machine Learning Consortium to Advance Health Equity and Researcher Diversity (AIM-AHEAD) and Bridge to Artificial Intelligence (Bridge2AI)—have been established. In 2021, NIH hosted Micro Labs, which focused on relevant stakeholders; key opportunities, challenges, and themes; and organizing and understanding opportunity. Several interdisciplinary themes were identified, including moving toward a systems approach to ethics for the AI data ecosystem and using AI as a tool for ethics research. Last, an Innovation Lab was focused on a data ecosystems approach to ethical AI for biomedical and behavioral research.

On November 1, 2022, Dr. Susan Gregurick, ODSS Director, presented on “Data Science at NIH” to provide a broad overview about data science initiatives and accomplishments at NIH. ODSS works with more than 200 NIH staff from 23 ICs and offices to catalyze data science across NIH. Dr. Gregurick affirmed that the overall goal is a modernized, integrated, FAIR biomedical data ecosystem. NIH has had several accomplishments in data science during the past 3 years. The NIH Science and Technology Research Infrastructure for Discovery, Experimentation, and Sustainability (STRIDES) Initiative, a partnership with cloud service providers, has shifted 200 petabytes of biomedical data to the cloud, including SARS-CoV-2 sequences collected from the global research community. At least nine NIH ICs are participating in the NIH Cloud Platform Interoperability effort. ODSS is aligning NIH-funded awards for databases and knowledge bases with the White House Office of Science and Technology Policy–desired characteristics for data repositories. The Data and Technology Advancement National Service Scholar Program is supporting experts in computer science, information science, and data science from academia, nonprofits, and other agencies with 1- to 2-year sabbaticals to work at NIH on challenging data science problems, including AI-readiness for pandemic preparedness and interoperability for COVID-19 data ecosystems.

Highlighting accomplishments specific to AI, Dr. Gregurick noted that NIH anticipates that Bridge2AI will generate new flagship biomedical and behavioral data sets that are ethically sourced, trustworthy, well defined, and accessible. Bridge2AI emphasizes ethical best practices and promotes diverse teams, and funded data generation projects are returning promising results. AIM-AHEAD, primarily established to enhance participation and representation of researchers and communities currently underrepresented in the development of AI/ML models to address health disparities and inequities, awarded 23 research fellowships in fiscal year 2021 (FY21), 25 leadership fellowships in fiscal year 2022 (FY22), and 22 pilot research projects in FY22.

Also on November 1, 2022, invited expert Dr. Bradley Malin, Accenture Professor of Biomedical Informatics, Biostatistics, and Computer Science and Vice Chair for Research Affairs, Department of Biomedical Informatics, Vanderbilt University, elaborated on recognizing and integrating social good into AI development. Participants understood that when creating AI, the principles of ethics must be embedded in the models from the beginning. Ethics can help concretize the goals of AI/ML and inform researchers of the pitfalls along the way. The computer science community is reframing the ethics of AI as AI for social good, thus building this technology for societal benefit. In terms of the AI life cycle (extending from people to data to models), flawed knowledge and unexplainable, unusable results can lead to a loss of context. Transparency also is necessary when reusing large data sets.

Generalizability is another key aspect that must be considered in model development. Infrastructure crosscuts AI, ethics, and the AI life cycle and creates research-ready environments. The NIH *All of Us* program is one such environment and demonstrates this infrastructure in practice. Dr. Malin expressed that studies of teams conducting AI and research-environment work should center on repeatability,

replicability, and reproducibility. Key takeaways for AI development were underscored and included the following: The problems faced are enormously complex and likely beyond researchers' current recognition; current ethics quandaries will take a long time to address; and ethics should be addressed before AI is created.

## Discussions

The October 24 and October 31, 2022, meetings with the AI-Workforce and AI-Readiness cohorts, respectively, included time for research teams to share their results in interactive lightning talks. During the lightning talks, participants were provided opportunities to learn about new AI readiness approaches, tools, and resources and to understand training related to AI workforce development. Awardees highlighted the successes of the AI Supplements in the context of their projects, some of which are summarized here.

- From the AI-Readiness discussions
  - Supplements helped keep critical talent in the awarded laboratories and teams.
  - Supplements were used to build a community around shared data and AI standards and benchmarking.
  - Supplements helped bring together researchers with different expertise.
- From the AI-Workforce discussions
  - AI-readiness is a new concept.
  - Although students may have some understanding and skill in AI/ML, it was less common to see students who could use their own data in AI/ML applications. Supplements helped address that gap.
  - Supplement projects were seen as logical extensions of existing courses offered and, in that way, synergistic with other learning.
- Topics common across the cohorts
  - Existing NIH-supported projects have been leveraged for data sharing.
  - Disease- and subject-specific projects are leveraging national and regional research networks to access data for their particular needs.
  - Limitations associated with ML approaches (e.g., bias) continue but there are some approaches to address and control these challenges.
  - The program involved diverse topics and a broad range of individuals, fostering cross-disciplinary interactions.
  - Awareness has been raised regarding how AI can be applied across a wide range of problems within multiple disciplines.
  - Data imputation and synthetic generation of data sets has improved.
  - The focus on reducing data curation barriers has led to better outcomes; the methods and quality of data labeling have improved.
  - Progress has been made related to pipelines for data curation by individual researchers, crowdsourcing, and “bot” methods. These data pipelines are being made more accessible.
  - Interactions among AI researchers have been catalyzed, and connections with researchers at AI institutes have been generated.
  - Resources are available to make the data useful for other researchers.
  - Communication between ML researchers and other scientists in different disciplines has been enabled and improved.

On each of the three days, breakout discussions provided an opportunity for researchers to exchange insights and experiences on topics of mutual interest. In the breakout discussions, awardees addressed key questions to better understand the gaps, challenges, and opportunities in AI workforce development and AI readiness. They also highlighted future training opportunities, tools and resources, and the expertise required to make data AI-ready. Examples of research gaps, challenges, and opportunities are summarized below.

### ***Gaps and Challenges***

- Pedagogical standards (e.g., curriculum development) and “messy” practice data sets are difficult to find.
- Focusing the training on AI readiness too early in the curriculum or training program can be a problem if students do not have sufficient foundational knowledge. A functional course to meet students where they are with their current skill level is needed.
- Full data deidentification is challenging. Institutional Review Board (IRB) approval also is challenging when considering data sets that might exist in perpetuity.
- Ethics are discussed but often are not implemented in research; this topic often is siloed.
- Data availability and access remain challenging with Health Insurance Portability and Accountability Act requirements.
- Computational expenses associated with cloud computing are cost prohibitive and not accessible to all.
- Data partners are not at equal levels of understanding and technology capacity to support AI/ML-ready platforms.
- AI literacy and domain knowledge remain highly variable.
- Researchers still are reluctant to share data for aggregation and model building; NIH mandates would provide increased incentives for data sharing.
- Unique operational challenges exist in data management and documentation pipelines for “living” data sets, to which data are added continually.
- Traditional biostatisticians still have reservations about AI methods.
- Staffing challenges include finding people with the relevant background and competing with market salaries.

### ***Opportunities***

- Develop training focused on specific research areas (e.g., environmental health sciences).
- Establish a knowledge base of experts.
- Investigate strategies to make data utilization skills common.
- Consider the topic of community ownership of data, as well as environmental- and justice-focused extensions of FAIR principles.
- Explore incorporating in-service training and education into AI processes.
- Investigate mechanisms to incentivize publishing data sets.
- Develop methods to assess and validate the AI tools and support more training for the users to better understand the limitations of the tools.

In open conversation sessions, awardees discussed cohort support and their vision for the future. Key words are captured in the word cloud on the following page.



The primary advice to teams just starting to make their biomedical research data FAIR and AI ready was to develop a data dictionary. The majority of awardees expressed that the AI Supplements kickoff and closeout meetings were beneficial and described their experience with this program as excellent.

### Key Take-Aways

Participants highlighted several specific challenges and opportunities related to their experiences and to NIH's AI programs generally:

- Making data FAIR and AI-ready begins at the research planning phase.
- Workforce needs have two audience types: (1) students and early career researchers who want to develop their skills in making data FAIR and AI-Ready. This was the group targeted by most of the AI-Workforce project. (2) post-docs and PIs who want to prepare grant proposals and plan for research that results in FAIR and AI-Ready data.
- Training focused on a specific area of biomedicine is more likely to result in skills that translate into practice.
- Technology is moving faster than we can train. The field is not producing a sufficient workforce to keep up. The skills that will last and transcend changes in the technology include critical thinking and collaboration skills. Ultimately, trainees must be able to teach themselves.
- Focus on enabling team-based science, rather than expecting people to gain multiple expertise. Focus on the gaps between existing expertise.
- Training would benefit from data, analysis tools, and course materials packaged together for easy use by students and teachers. Consideration of the use of data sets could be integrated in training and curriculum.
- Finding a platform to share examples of model data sets to walk the trainees through the steps (from start to finish) for making data AI-ready is essential.

- Participants wanted a way to share their research and training products: research results, data, training, and models.
- Adequately addressing challenges under the umbrella term of “ethics” requires more than just training and more than just a team approach. Organizational buy-in is also needed.
- Participants found it hard to find ethics experts who can speak to data issues.
- Datasheets are not widely used.
- Collaboration with AI/ML experts increases overall research budgets significantly (and may not be supported through current peer review processes).
- Training support to create a diverse AI-ready workforce is an ongoing need.
- Participants expressed general agreement on the need to reinvent an ethical review process or reinvent an IRB specifically for this space.

## Appendix A

### Closeout Meeting Participants

#### Office of Data Science Strategy (ODSS)

- Laura Biven, Ph.D., ODSS, National Institutes of Health (NIH) (Organizer)\*
- John Gachago, Ph.D., ODSS, NIH
- Samson Gebreab, Ph.D., ODSS, NIH
- Susan Gregurick, Ph.D., ODSS, NIH
- Raphael Isokpehi, Ph.D., ODSS, NIH
- Ashley Hackett, M.S., ODSS, NIH
- Larry Holt, ODSS, NIH
- Bryan Kim, Ph.D., ODSS, NIH
- Alison Lin, Ph.D., ODSS, NIH
- Fenglou Mao, Ph.D., ODSS, NIH
- Elaine Nsoesie, Ph.D., ODSS, NIH
- Belinda Seto, Ph.D., ODSS, NIH
- Michael Spittel, Ph.D., ODSS, NIH (Organizer)\*
- Hsinyi (Steve) Tsang, Ph.D., ODSS, NIH
- Christopher Siwy, Ph.D., ODSS, NIH
- Vivian Ota Wang, Ph.D., ODSS, NIH\*

#### Co-Sponsoring NIH Institutes and Centers/Federal Observers/Invited Guests

- Dorothy Beckett, Ph.D., National Institute of General Medical Sciences (NIGMS), NIH
- Michelle Bond, Ph.D., National Human Genome Research Institute (NHGRI), NIH
- Alicia Chou, M.S., National Institute of Dental and Craniofacial Research, NIH
- Ellen Wright Clayton, M.D., J.D., M.S., Vanderbilt University
- Jennifer Collins, M.S., National Institute of Environmental Health Sciences (NIEHS), NIH\*
- Christine Cutillo, M.M.C.i., National Center for Advancing Translational Sciences, NIH
- Matthew Dancis, Ph.D., Booz Allen Hamilton
- Chris Duncan, Ph.D., NIEHS, NIH
- Deborah Duran, Ph.D., National Institute on Minority Health and Health Disparities, NIH\*
- Laura Dwulet, Ph.D., Booz Allen Hamilton
- René Etcheberrigaray, M.D., National Institute on Aging (NIA), NIH
- James Gao, Ph.D., National Eye Institute, NIH
- Elizabeth Ginexi, Ph.D., Office of Behavioral and Social Sciences Research, NIH\*
- Anne Gershenson, Ph.D., NIGMS, NIH
- Michelle Heacock, Ph.D., NIEHS
- David Leitman, Ph.D., National Institute of Mental Health, NIH\*
- Bradley Malin, Ph.D., Vanderbilt University
- Matthew McAuliffe, Ph.D., Biomedical Research Informatics Computing System, NIH
- Rosemary McKaig, Ph.D., M.P.H., National Institute of Allergy and Infectious Diseases, NIH
- Roger Miller, Ph.D., National Institute on Deafness and Other Communication Disorders, NIH
- Laurie Novak, Ph.D., M.H.S.A., Vanderbilt University Medical Center
- Grace Peng, Ph.D., National Institute of Biomedical Imaging and Bioengineering, NIH
- David Resnik, Ph.D., J.D., NIEHS, NIH\*
- Carol Shreffler, Ph.D., NIEHS, NIH

- Heidi Sofia, Ph.D., NHGRI, NIH
- Yanli Wang, Ph.D., National Library of Medicine, NIH\*
- Martin Were, M.D., M.S., Vanderbilt University Medical Center
- Jonathan Wren, Ph.D., Oklahoma Medical Research Foundation
- Susan Wright, Ph.D., National Institute on Drug Abuse, NIH
- Don Yu, Scholarship America

\* NIH Staff who served as volunteer breakout room moderators

### **AI-Readiness Recipient Team Members**

- David Adalsteinsson, Ph.D., The University of North Carolina at Chapel Hill
- Majid Afshar, M.D., University of Wisconsin–Madison
- Akram Alshawabkeh, Ph.D., Northeastern University
- Fabrisia Ambrosio, Ph.D., University of Pittsburgh
- Alvin Ang, M.D., M.P.H., Boston University
- Steven Bedrick, Ph.D., Oregon Health & Science University
- Nina R. Benway, M.S., Syracuse University
- Jeremy Bigness, M.A., Brown University
- Chris Buswinka, Harvard Medical School
- Chuming Chen, Ph.D., University of Delaware
- Feixiong Cheng, Ph.D., Cleveland Clinic
- Apurva Chitre, M.S., University of California, San Diego
- John Chodera, Ph.D., Memorial Sloan Kettering Cancer Center
- Brandon Conklin, M.S., Rutgers University
- Eric Darve, Ph.D., Stanford University
- Colin Depp, Ph.D., University of California, San Diego
- Rutvik Desai, Ph.D., University of South Carolina
- Matt Dunbar, Ph.D., University of Washington
- Jonathan Elmer, M.D., M.S., University of Pittsburgh
- Gerasimos Fergadiotis, Ph.D., Portland State University
- Benjamin Goudey, Ph.D., The University of Melbourne
- M. Cristina Vazquez Guillamet, M.D., Washington University in St. Louis
- Stephanie Gupton, Ph.D., The University of North Carolina at Chapel Hill
- Katy Haynes, The University of Chicago
- Zihuai He, Ph.D., Stanford University
- Karen Hirsch, M.D., Stanford University
- Yoonmi Hong, Ph.D., The University of North Carolina at Chapel Hill
- Artur Indzhykulian, Ph.D., Harvard Medical School
- Lei Jiang, M.A., University of Missouri
- David Kaeli, Ph.D., Northeastern University
- Richard Kennedy, Ph.D., University of Alabama, Birmingham
- Vijaya Kolachalama, Ph.D., Boston University
- Erica Larschan, Ph.D., Brown University
- Donghoon Lee, ScD., M.S., M.Sc., Harvard T.H. Chan School of Public Health
- Haiqun Lin, Ph.D., Rutgers University
- Boyla Mainsah, Ph.D., Duke University
- Maryann Martone, Ph.D., University of California, San Diego
- Tara McAllister, Ph.D., New York University



- Ofer Mendeleevitch, Ph.D., Syntegra
- Daniel Messinger, Ph.D., University of Miami
- Pietro Michelucci, Ph.D., Human Computation Institute
- Olga Jarrin Montaner, Ph.D., RN, FAAN, Rutgers University
- Senthil Nachimuthu, M.D., Ph.D., Nightingale Open Science
- Madeline Oguss, M.S., University of Wisconsin–Madison
- Abraham Palmer, Ph.D., University of California, San Diego
- Serguei Pakhomov, Ph.D., University of Minnesota
- Sharan Prakash, California Institute of Technology
- Luca Pegolotti, Ph.D., Stanford University
- Lynn Perry, Ph.D., University of Miami
- Benjamin Pritchard, Ph.D., Molecular Sciences Software Institute and Virginia Polytechnic Institute and State University
- Oksana Poleskaya, Ph.D., University of California, San Diego
- Eric Reed, Ph.D., Tufts Medical Center
- Julia Salzman, Ph.D., Stanford University
- Robert Schuler, M.S., University of Southern California
- Paola Sebastiani, Ph.D., Tufts Medical Center
- Nordine Sebkhi, Ph.D., Georgia Institute of Technology
- Amit Sheth, Ph.D., University of South Carolina
- Kerry Smith, M.S., Johns Hopkins University
- Padhraic Smyth, Ph.D., M.S., University of California, Irvine
- Paul Sternberg, Ph.D., California Institute of Technology
- Martin Styner, Ph.D., The University of North Carolina at Chapel Hill
- Jun Wang, Ph.D., The University of Texas at Austin
- Newell Washburn, Ph.D., Carnegie Mellon University
- Dong Xu, Ph.D., University of Missouri
- Fan Zhang, Ph.D., Yale University
- Judy Zhong, Ph.D., New York University School of Medicine

#### **Workforce Development Recipient Team Members**

- Gary Benson, Ph.D., Boston University
- Phil Brown, Ph.D., Northeastern University
- Marino Bruce, Ph.D., M.S.R.C., M.Div., University of Houston
- Karen Butler-Purry, Ph.D., Texas A&M University
- Chuming Chen, Ph.D., University of Delaware
- Linda Clark, Ph.D., Brown University
- Keith Fraser, Ph.D., Rensselaer Polytechnic Institute
- Matthew Gitzendanner, Ph.D., University of Florida
- Catherine Grimes, Ph.D., University of Delaware
- David Julian, Ph.D., University of Florida
- Adam Labadorf, Ph.D., Boston University
- Abiel Roche Lima, Ph.D., University of Puerto Rico Comprehensive Cancer Center
- Craig Marcus, Ph.D., Oregon State University
- Ana Patricia Ortiz, Ph.D., University of Puerto Rico Comprehensive Cancer Center
- Shawn Polson, Ph.D., University of Delaware
- Christina Retzer, M.B.A., Texas A&M University
- Ivan Rusyn, Ph.D., Texas A&M University

- Courtney Shelley, Ph.D., Los Alamos National Laboratory
- Jacob Searcy, Ph.D., University of Oregon
- Jeanette Stingone, Ph.D., M.P.H., Columbia University Mailman School of Public Health
- Katie Wrenn, Ph.D., United Kingdom Health Security Agency
- Hao Ye, Ph.D., M.A., M.S., University of Florida
- Judy Zhong, Ph.D., New York University School of Medicine

## Appendix B

### ODSS AI Supplements Closeout Meeting

NOT-OD-21-079 (AI Workforce)

October 24, 2022

Virtual Meeting

#### Day 1 Agenda

---

11:00 a.m. – 11:10 a.m.	<b>Welcome and Introductions</b> <i>Laura Biven</i> <i>Michael Spittel</i>
11:10 a.m. – 11:40 a.m.	<b>Updates on AI from NIH</b> <i>Laura Biven</i>
11:40 a.m. – 11:45 a.m.	<b>Introduction to Lightning Talks</b> <i>Laura Biven</i>
11:45 a.m. – 11:55 a.m.	<b>Parallel Interactive Lightning Talks – Session A</b> Room 1 Project: FAIR and Practical Data Science Training at the Chemistry– Biology Interface <i>PI Catherine Grimes</i>  Room 2 Project: Adding a FAIR Data Practices Curriculum to UF’s Practicum AI AI/ML Training Workshops <i>PI: David Julian</i>
11:55 a.m. – 12:00 p.m.	<b>Breakout Transition (5 minutes)</b>
12:00 p.m. – 12:10 p.m.	<b>Parallel Interactive Lightning Talks – Session B</b> Room 1 Project: Stackable Trainings in the FAIRification and AI/ML-Readiness of Data with Applications to Environmental Health and Justice <i>PI: Phil Brown</i>  Room 2 Project: AI Training Module for Vision Science <i>PI: Kate Keller</i>  Room 3 Project: Predoctoral Training in Biological Data Management for Advanced Computational Analysis and the Ethical Usage of Biological Data <i>PI: Thomas Tullius</i>

12:10 p.m. – 12:15 p.m.

**Breakout Transition (5 minutes)**

12:15 p.m. – 12:25 p.m.

**Parallel Interactive Lightning Talks – Session C**

Room 1

Project: Maximizing Student Development in Data- and Information Science-Related Disciplines for Biomedical Ph.D. Trainees at Texas A&M University and Beyond

*PI: Karen Butler-Purry*

Room 2

Project: Workforce Training for Making Data FAIR and Compatible with Machine Learning and Artificial Intelligence Applications

*PI: Marcus Craig*

Room 3

Project: Development of Data Science Course and Summer Bootcamp for Alzheimer's Disease and Related Dementia Researchers

*PI: Chunyu Wang*

12:25 p.m. – 12:30 p.m.

**Breakout Transition (5 minutes)**

12:30 p.m. – 12:40 p.m.

**Parallel Interactive Lightning Talks – Session D**

Room 1

Project: Cancer Research Workforce Development in FAIR Artificial Intelligence and Machine Learning

*Presenter: Issam El Naqa*

*PI: William Cress*

Room 2

Project: Making Data FAIR and AI/ML Applications for Cancer Prevention and Control (AI/ML-CAPAC) Research Among Hispanics

*PI: Ana Patricia Ortiz*

Room 3

Project: Demystifying Machine Learning and Best Data Practices Workshop Series for Underrepresented STEM Undergraduate and M.S. Researchers Bound for Ph.D. Training Programs

*Presenter: Pleuni Pennings*

*PI: Raymond Esquerra*

12:40 p.m. – 12:45 p.m.

**Breakout Transition (5 minutes)**

12:45 p.m. – 12:55 p.m.

**Parallel Interactive Lightning Talks – Session E**

Room 1

Project: Making Environmental Health Data FAIR and AI/ML-Ready

*Presenter: Jeanette Stingone*

*PI: Gary Miller*

12:55 p.m. – 1:00 p.m.	<b>Breakout Transition (5 minutes)</b>
1:00 p.m. – 1:10 p.m.	<b>Parallel Interactive Lightning Talks – Session F Room 1</b> Project: Next Generation Sequencing and Biological Imaging in the Era of Machine Learning <i>Presenter: Jacob Searcy</i> <i>PI: Karen Guillemin</i>
1:10 p.m. – 1:20 p.m.	<b>Break (10 minutes)</b>
1:20 p.m. – 1:30 p.m.	<b>Instructions for Breakout Discussion 1</b>
1:30 p.m. – 2:30 p.m.	<b>Discussion (Breakouts)</b>  <i>Questions</i> <ul style="list-style-type: none"> <li>▪ <i>Considering your own project, and the projects you learned about in the lightning talks, what are some important gaps that exist in training in FAIR and AI-ready data?</i></li> <li>▪ <i>Considering your own project, and the projects you learned about in the lightning talks, what are the key challenges in integrating ethics into training for FAIR and AI-ready data?</i></li> <li>▪ <i>What does the future of training in this space look like in 5–10 years?</i></li> <li>▪ <i>What is the value of tools, services, and training opportunities from the private sector? (What are some examples of impactful potential uses for the research community?)</i></li> <li>▪ <i>Considering the projects you learned about in the lightning talks, what are some of the successes or highlights within the projects?</i></li> <li>▪ <i>Which competencies did you aim for in your course? Is “competency” the correct term?</i></li> <li>▪ <i>Where do people go to find the correct expertise, and is it easy or difficult to find that expertise?</i></li> <li>▪ <i>What are other potential career paths for students (i.e., outside of academia) who will use these skills?</i></li> </ul>
2:30 p.m. – 2:35 p.m.	<b>Break (5 minutes)</b>
2:35 p.m. – 3:05 p.m.	<b>Readout from Breakouts</b>
3:05 p.m. – 3:20 p.m.	<b>Open Conversation of Cohort Support and Future Vision</b> <i>Laura Biven</i>
3:20 p.m. – 3:30 p.m.	<b>Thank You and Closeout</b> <i>Laura Biven</i>

## ODSS AI Supplements Closeout Meeting

NOT-OD-21-094 (AI Readiness)

October 31, 2022

Virtual Meeting

### Day 2 Agenda

---

11:00 a.m. – 11:10 a.m.	<b>Welcome and Introductions</b> <i>Laura Biven</i> <i>Michael Spittel</i>
11:10 a.m. – 11:40 a.m.	<b>Updates on AI from NIH</b> <i>Laura Biven</i>
11:40 a.m. – 11:45 a.m.	<b>Introduction to Lightning Talks</b> <i>Laura Biven</i>
11:45 a.m. – 11:55 a.m.	<b>Parallel Interactive Lightning Talks – Session A</b> Room 1 Project: Building a Substance Use Data Commons for Public Health Informatics <i>PI: Majid Afshar</i>  Room 2 Project: A Computational Pipeline to Evaluate AI/ML Readiness in Digital Datasets in the Framingham Heart Study <i>Presenter: Vijaya Kolachalama</i> <i>PI: Lindsay Farrer</i>  Room 3 Project: An AI-Ready Vascular Model Repository for Modeling and Simulation in Cardiovascular Disease <i>Presenter: Luca Pegolotti</i> <i>PI: Alison Marsden</i>  Room 4 Project: Enabling the AI/ML-Readiness of Massive Single-Cell Data for Discovering RNA Regulatory Biology <i>Presenter: Julia Salzman</i>  Room 5 Project: Addressing Class Imbalance and Missingness in the PROTECT Database <i>Presenter: David Kaeli</i> <i>PI: Akram Alshawabkeh</i>  Room 6 Project: De-identified Delirium Data: Finding Delirium to Study Delirium <i>Presenter: Richard Kennedy</i>

11:55 a.m. – 12:00 p.m.

**Breakout Transition (5 minutes)**

12:00 p.m. – 12:10 p.m.

**Parallel Interactive Lightning Talks – Session B**

Room 1

Project: Using Machine Learning and Artificial Intelligence Models to Predict Muscle Stem Cell Biological Age and Regenerative Potential

*PI: Fabrisia Ambrosio*

Room 2

Project: Precision Care After Cardiac Arrest

*PI: Karen Hirsch*

Room 3

Project: Democratizing Machine Learning for Researchers Working in Alzheimer's Space

*Presenter: Benjamin Goudey*

*PI: Colin Masters*

Room 4

Project: Improving AI/ML-Readiness of Data Generated Under the R01: Protein Signatures of APOE2 and Cognitive Aging

*Presenter: Paola Sebastiani or Ofer Mendelevitch*

*PI: Paola Sebastiani*

Room 5

Development and Use of an AI/ML-Ready Dog Aging Project Dataset

*Presenter: Matt Dunbar*

*PI: Daniel Edward Promislow*

12:10 p.m. – 12:15 p.m.

**Breakout Transition (5 minutes)**

12:15 p.m. – 12:25 p.m.

**Parallel Interactive Lightning Talks – Session C**

Room 1

Project: Towards Automatic Transcription of Post-Stroke Disordered Speech

*PI: Steven Bedrick*

Room 2

Project: Cross-Modality Imaging Data Annotations for Deep Learning-Based Analysis Solutions in the Auditory Field

*Presenter: Chris Buswinka*

*PI: Artur Indzhukulian*

Room 3

Project: PERCEPT: A Database of Clinical Child Speech for Automatic Speech Recognition and Classification

*PI: Tara McAllister*

Room 4

Project: Model Organism Neural Circuit Knowledge Graph

*Presenter: Paul Sternberg or Sharan Prakash*

*PI: Paul Sternberg*

Room 5  
Project: Agent-Based Participation of Machine Learning Models in a  
Crowdsourcing System  
*Presenter: Pietro Michelucci*  
*PI: Chris B. Schaffer*

12:25 p.m. – 12:30 p.m.

**Breakout Transition (5 minutes)**

12:30 p.m. – 12:40 p.m.

**Parallel Interactive Lightning Talks – Session D**

Room 1  
Project: Using Artificial Intelligence for Alzheimer’s Disease Drug  
Repurposing  
*PI: Feixiong Cheng*

Room 2  
Project: Developing AI/ML-Ready Aging Trajectory Files  
*PI: Olga Jarrin Montaner*

Room 3  
Project: Harnessing Multimodal Data To Enhance Machine Learning of  
Children’s Vocalizations  
*PI: Daniel Messinger*

Room 4  
Project: Detecting Speech Articulation Patterns Following Laryngeal  
Cancer Treatment Using Artificial Intelligence and Machine Learning  
*Presenter: Jun Wang or Nordine Sebkh*  
*PI: Jun Wang*

Room 5  
Project: Rescuing Missed Longitudinal MRI Scans in the UNC Early  
Brain Development Study  
*Presenter: Martin Styner*  
*PI: John Gilmore*

12:40 p.m. – 12:45 p.m.

**Breakout Transition (5 minutes)**

12:45 p.m. – 12:55 p.m.

**Parallel Interactive Lightning Talks – Session E**

Room 1  
Project: A Novel Dataset for Speech Analysis in Serious Mental Illness  
(Parent Study: Social Cognitive Biases and Suicide in Psychotic  
Disorders)  
*PI: Colin Depp*

Room 2  
Project: Improving AI/ML-Readiness of Data Generated from HABLE  
or Other NIH-Funded Research  
*Presenter: Fan Zhang*  
*PI: Sid O’Bryant*



Room 3  
Project: Machine Learning Development Environment for Single-Cell Sequencing Data Analyses  
*PI: Dong Xu*

Room 4  
Project: Improving AI/ML-Readiness of FaceBase Research Datasets  
*Presenter: Robert Schuler*  
*PI: Carl Kesselman*

Room 5  
Project: Making Data From the Center for GWAS in Outbred Rats FAIR and AI/ML Ready  
*Presenter: Abraham Palmer*

12:55 p.m. – 1:00 p.m.

**Breakout Transition (5 minutes)**

1:00 p.m. – 1:10 p.m.

**Parallel Interactive Lightning Talks – Session F**

Room 1  
Project: AI/ML-Readiness for Neuroimaging of Language  
*PI: Rutvik Desai*

Room 2  
Project: Machine Learning-enabled Comparative Transcriptomic Profiling to Validate Nanoscript-Induced Inner Ear Hair Cells  
*Presenter: Brandon Conklin*  
*PI: Kibum Lee*

Room 3  
Project: Multi-omic Human Brain Immune Cell (HBIC) Resources for AI/ML Applications  
*Presenter: Donghoon Lee*  
*PI: Panagiotis Roussos*

Room 4  
Project: Fair Risk Predictions for Underrepresented Populations Using Electronic Health Records  
*PI: Judy Zhong*

Room 5  
Project: Extending the QCArchive Small Molecule Quantum Chemistry Archive To Support Machine Learning Applications in Biomolecular Modeling  
*Presenter: John Chodera*  
*PI: Michael Shirts*

Room 6  
Project: Assessing HER Data Readiness for ML/AI Algorithms in Sepsis and Antimicrobial Resistance  
*Presenter and PI: M. Cristina Vazquez Guillamet*

1:10 p.m. – 1:20 p.m.

**Break (10 minutes)**

1:20 p.m. – 1:30 p.m.

**Instructions for Breakout Discussion 1**

1:30 p.m. – 2:30 p.m.

**Discussion (Breakouts)**

*Questions*

- *Considering your own project and the projects you learned about in the lightning talks, what are some of the successes or highlights from this program?*
- *Considering your own project and the projects you learned about in the lightning talks, what are some of the open challenges?*
- *What are the key challenges and opportunities for ethical, AI-ready data?*
- *What skills or expertise were critical for your team to make data AI ready?*
- *For which AI applications do you anticipate using your data?*
- *What makes data AI ready?*

2:30 p.m. – 2:35 p.m.

**Break (5 minutes)**

2:35 p.m. – 3:05 p.m.

**Readout from Breakouts**

3:05 p.m. – 3:20 p.m.

**Open Conversation of Cohort Support and Future Vision**

*Laura Biven*

3:20 p.m. – 3:30 p.m.

**Thank You and Closeout**

*Laura Biven*

## ODSS AI Supplements Joint Closeout Meeting

November 1, 2022

Virtual Meeting

### Day 3 Agenda

---

11:00 a.m. – 11:10 a.m.	<b>Welcome and Introductions</b> <i>Laura Biven</i> <i>Michael Spittel</i>
11:10 a.m. – 12:00 p.m.	<b>Recognizing and Integrating Social Good into the AI Development Lifecycle</b> <i>Bradley Malin, Accenture Professor of Biomedical Informatics, Biostatistics, and Computer Science; Vice Chair for Research Affairs</i> <i>Department of Biomedical Informatics, Vanderbilt University</i>
12:00 p.m. – 12:05 p.m.	<b>Break (5 minutes)</b>
12:05 p.m. – 1:00 p.m.	<b>Data Science at NIH</b> <i>Susan Gregurick, Associate Director for Data Science and Director of the Office of Data Science Strategy, NIH</i>
1:00 p.m. – 1:10 p.m.	<b>Break (10 minutes)</b>
1:10 p.m. – 2:10 p.m.	<b>Discussion (Breakouts)</b>  <i>Questions</i> <ul style="list-style-type: none"><li>▪ <i>What does it mean for data to be AI ready? (How do you create AI-ready data? What steps need to be taken?)</i></li><li>▪ <i>What does the ethical development of AI mean?</i></li><li>▪ <i>In this dynamic space of new technologies and managing best practices for upholding privacy and ethics goals, how can NIH keep pace? How can NIH researchers help innovate?</i></li></ul>
2:10 p.m. – 2:15 p.m.	<b>Break (5 minutes)</b>
2:15 p.m. – 3:05 p.m.	<b>Readout from Breakouts</b>
3:05 p.m. – 3:20 p.m.	<b>Open Conversation of Cohort Support and Future Vision</b> <i>Laura Biven</i>
3:20 p.m. – 3:30 p.m.	<b>Thank You and Closeout</b> <i>Laura Biven</i>