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)

**Foundational Discussions in Artificial Intelligence–Readiness and
Workforce Development: Joint Kickoff Meeting**

November 1, 2, and 15, 2021
Virtual

Summary Report

Description

This series of three meetings, sponsored by the NIH Office of Data Science Strategy (ODSS), brought together recipient team members (i.e., 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 Administrative Supplements to Support Collaborations to Improve the AI/ML-Readiness of NIH-Supported Data (NOT-OD-21-094). Artificial intelligence and machine learning (AI/ML) are a collection of data-driven technologies with the potential to significantly advance biomedical research. Much of this potential is unrealized, however, because biomedical data are not collected and prepared in ways that would allow them to be used efficiently and effectively by AI/ML applications. The task of making data FAIR and AI/ML-ready is not only algorithmic. It requires multi-disciplinary expertise, experimentation and, often, iterative feedback from AI/ML applications and experts. AI/ML-readiness should be guided also by attention to ethical, legal, and social implications.

With the NOT-OD-21-079 (AI Workforce Development) and NOT-OD-21-094 (AI Readiness) NOSIs, the NIH intends to support collaborations that assemble multidisciplinary expertise in biomedicine, data management, and AI/ML to improve the AI/ML readiness of data generated from NIH-funded research and shared through repositories, knowledge bases, or other data-sharing resources. This cohort of NIH-funded investigators is laying the foundation to support the next wave of biomedical discovery empowered by AI. The purpose of the meeting was to bring together these experts to network, share ideas, and potentially inspire collaborations or further sharing. It was also an opportunity for NIH to better understand details of the planned activities and challenges in the training and research environment to build AI-ready data sets and associated tools; learn about methods to recruit the expertise needed; and better understand the available resources across the AI Readiness and AI Workforce Development teams.

Three main sessions composed the joint kickoff event. Meeting 1—Foundational Discussions in AI Readiness and Workforce Development: Joint Kickoff—on November 1, 2021, consisted of bidirectional conversations between the AI Readiness and AI Workforce Development groups, with a focus on ethics. Meeting 2—Foundational Discussions and Networking for AI Readiness—on November 2, 2021, convened the AI Readiness NOSI recipient team members. Meeting 3—Foundational Discussions and Networking for AI Workforce Development—on November 15, 2021, convened the AI Workforce Development NOSI recipient team members. The agendas included presentations from the ODSS and other experts in the field; speed networking sessions; breakouts to better understand common interests,

challenges, and themes; and discussions on functioning as a cohort. ODSS staff and other representatives from co-sponsoring NIH Institutes and Centers moderated the breakout sessions. More than 100 participants, including recipient team members and federal observers, registered and attended.

Overview and Highlights

The ODSS Director, Dr. Susan Gregurick, presented on "Connecting Data, Enhancing Software, and Creating a Data Ecosystem," intended to initiate a discussion about data science initiatives at the NIH. The ODSS Director noted the motivations for the AI NOSIs. Researchers remain challenged to make data FAIR (findability, accessibility, interoperability, reusability) and AI ready. In addition, data scientists or AI engineers are tasked to understand vast quantities of data. The AI NOSIs address these challenges and align with the Final NIH Policy for Data Management and Sharing and overarching strategic goals and objectives outlined in the *NIH Strategic Plan for Data Science*.

Dr. Laura Biven, also from ODSS, conveyed that ethics is a guiding principle in leading government-wide initiatives to accelerate AI/ML and reported on AI ethics from the perspective of the NIH. It was expressed that to address biomedical data appropriately, AI/ML readiness should be guided by a concern for human and clinical impact. This therefore requires attention to ethical, legal, and social implications of AI/ML, including, but not limited to—

- Biases in data sets, algorithms, and applications
- Concerns related to privacy and confidentiality
- Effects on disadvantaged or marginalized groups and health disparities
- Unintended adverse social consequences of research and development

Invited experts, Prof.. Katie Shilton, Associate Professor, College of Information Studies, University of Maryland, and Prof. James Zou, Assistant Professor, Biomedical Data Science, Stanford University, elaborated on trustworthiness in AI data and shared insight on data-centric lessons for trustworthy biomedical AI research, respectively. Participants were provided with opportunities to engage in discussions of AI readiness and AI workforce development and to network and learn from one another.

Discussion Topics

After the opening presentations, the meetings focused on discussing key topics. A number of themes, outcomes, and next steps emerged from the discussions and are summarized here.

Leadership, Workforce, and Readiness

Participants gathered into small breakout groups and discussed the steps needed for data preparedness for AI/ML, training, and collaboration. A member of each group provided a brief overview of the topics and themes covered during the session, noting the following key points raised during their breakout discussion.. Key points discussed by the participants included:

- The need for a common AI/ML language, noting the importance of annotating data and how annotations change and need to change over time.
- Preparing data is impacted and challenged by heterogeneity across data types, institutions, and time.
- Data set "cleaning" to remove problematic data currently requires significant effort. Automating this process would save a great amount of work but might be especially challenging for voice and imaging data.
- Attracting and retaining AI experts in biomedical science is a challenge.

- Regulatory process can present a time-consuming challenge. Institutional review board (IRB) approval, Data Use Agreements (DUAs), and data owner approvals, which can take months. Cohort and data element discovery and anonymization require extensive planning as larger data sets are built from various sources which multiply the time needed to obtain the appropriate approvals and agreements.
- Outstanding questions surrounding data and scientific training include: *Should everyone who generates data learn how to share those data, or should a subset of experts working with investigators needed to make data AI/ML ready and shareable? Who pays for these experts? What is the career track for these expert data scientists in academia?*
- Complex research projects often require someone who can communicate in both areas from the outset (i.e., before the design phase to determine necessary formatting and identify what data to collect).
- A vetted system of metadata and systems for validating and assessing the completeness of metadata is needed. Minimum standards of metadata are needed so that data are informative and can be shared.
- Researchers and domain experts should work with modelers to ensure that models align with research goals and domains. They also should understand the limitations of ML models.

Considering Ethics in the Data-to-AI/ML Research Cycle

Considering ethics in AI/ML data and research aligns with the priorities of the NIH Advisory Committee to the Director Working Group on AI[1]. ODSS AI Program Coordinator, Dr. Allison Dennis, provided an overview of NIH interest on this topic, and Prof. Katie Shilton and Prof. James Zou, two leading experts in the field, discussed trustworthiness of AI data and provided data-centric lessons for trustworthy biomedical AI. Participants again gathered into breakout groups, noting the following key points raised during their breakout discussion. Key points discussed by the participants included:

- There is a need to examine data and identify biases, identify missing groups, identify and engage ethics experts, examine the funding and review process, examine publication challenges, and include ethical considerations in research budgets.
- There can be a tension between open data and protecting participants' rights.
- A possible mechanism for incentivizing data sharing could be encouraging job seekers to share data sets they have worked to make FAIR or open as line items on CVs and encouraging hirers to view this as a valuable contribution.
- As data models become more complex, better curation is needed to identify the scope of ethical problems.
- Academic institutions often incentivize avoiding risk, which can de-incentivize the development of creative, new solutions.
- The development of models for communities affected indirectly in research studies poses ethical challenges.
- Existing IRB protocols, information related to bias, and the grant mechanisms that address ethical questions in AI/ML studies are currently limited.

---

1 https://www.acd.od.nih.gov/documents/presentations/12132019AI_FinalReport.pdf

Crosscutting Themes

A. Readying Diverse Data Types for AI/ML Applications

AI Readiness recipient team members were assigned to breakout rooms (moderated by NIH staff) based on the data types with which they work. Data types included -omics, imaging, health systems, and speech. Common themes emerging from the discussions are summarized here.

- Training plans and use cases for AI/ML data are needed.
- Mechanisms to addressing standards for preparing diverse data types for different purposes, as well as current databases and data acquisition methods, will be necessary.
- Maintenance and curation remains challenging; thus, new methods and standards are needed.
- Facilities for data sharing are limited, and the role of the NIH in supporting this effort is unclear. Developing a roadmap to data sharing that addresses the issues, including regulatory hurdles might be explored.
- Analysis of heterogeneous data sets requires special considerations.
- Challenges regarding use of metadata standards included a lack of a single metadata standard or data model to cover every data type or class. Implementing a hierarchy of models could be one solution.
- Challenges regarding use of data include legal and regulatory agreements, lack of a clear centralized solution, blanket memorandum of understanding
- Training data sets are limited in the field.

B. Building Core Competencies in Data Science for AI/ML for Trainees at Diverse Career Stages

AI Workforce Development recipient team members were assigned breakout rooms (moderated by NIH staff) based on the career-stage (late, early, middle, broad) focus of their projects. Common themes emerging from the discussions are summarized here:

- A training gap remains between biomedical domain experts and data scientists. One solution could be to train "translators" to assist with communication between the two disciplines.
- Incentive models for training on all levels (e.g., individual researcher, community) are lacking.
- Sustainability (updating and maintenance of materials) beyond the current initiatives is needed. A central website would be useful for collating all the information.
- Use of broad, asynchronous training material applicable to multiple domains and taught by non-computational experts will be beneficial to addressing current workforce needs. Teaching assistants will be necessary to guide and supervise trainees making their way through the material.
- Responsibility for AI/ML training efforts, inclusion and ethics, and community standards for developing and teaching relevant curricula are not well understood.
- Ethics of AI/ML should be a core competency in developing AI-ready data.
- Core competencies should be broad to enable the training and inclusion of diverse participants.
- Data competency of domain scientists remains an issue.
- Training evaluation that encompasses use of advisory boards and pre- and post-training assessment (e.g., hosting a project week) will be essential. The field needs harmonization of evaluations and assessments for increased efficiency and standardization.

- Challenges to curriculum development exist. Teaching AI/ML conceptually without instructing users in AI/ML techniques could address the gap in knowledge and the intimidation barrier for those not trained in data science.

Appendix A

Kickoff Meeting Participants

Office of Data Science Strategy (ODSS)

Laura Biven, Ph.D., ODSS, National Institutes of Health (NIH) (Organizer)*
Allison Dennis, Ph.D., ODSS, NIH (Organizer)*
Allissa Dillman, Ph.D., ODSS, NIH*
Susan Gregurick, Ph.D., ODSS, NIH
Kristin Hook, Ph.D., ODSS, NIH (Organizer)*
Fenglou Mao, Ph.D., ODSS, NIH
Alyssa Tonsing-Carter, Ph.D., ODSS, NIH*
Vivian Ota Wang, Ph.D., ODSS, NIH*

Co-Sponsoring NIH Institutes and Centers/Federal Observers

- Michelle Bond, Ph.D., National Human Genome Research Institute (NHGRI), NIH
- Eric Brunskill, Ph.D., National Institute of Diabetes and Digestive and Kidney Diseases, NIH*
- Jennifer Collins, M.S., National Institute of Environmental Health Sciences (NIEHS), NIH*
- Valerie Cotton, *Eunice Kennedy Shriver* National Institute of Child Health and Human Development, NIH
- Valentina Di Francesco, M.S., NHGRI, NIH
- Deborah Duran, Ph.D., National Institute on Minority Health and Health Disparities, NIH*
- Lisa Federer, Ph.D., MLIS, National Library of Medicine (NLM), NIH*
- Michele Ferrante, Ph.D., National Institute of Mental Health, NIH*
- Elizabeth Ginexi, Ph.D., Office of Behavioral and Social Sciences Research, NIH*
- Nina Hsu, Ph.D., National Institute of Neurological Disorders and Stroke, NIH*
- Francesca Macchiarini, Ph.D., National Institute on Aging, NIH
- 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
- Carolina Mendoza-Puccini, M.D., National Institute of Neurological Disorders and Stroke, NIH
- Christopher Miller, M.D., National Heart, Lung, and Blood Institute (NHLBI), NIH*
- Grace Peng, Ph.D., National Institute of Biomedical Imaging and Bioengineering, NIH
- Chanel Press, Ph.D., NHLBI, NIH*
- David Resnik, Ph.D., J.D., NIEHS, NIH*
- Asif Rizwan, Ph.D., NHLBI, NIH*
- Carol Shreffler, Ph.D., NIEHS, NIH
- Helen Thompson, NHGRI, NIH
- Yanli Wang, Ph.D., NLM, NIH*
- Maryam Zaringhalam, Ph.D., NLM, NIH*

* 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
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
Jonathan Elmer, M.D., M.S., University of Pittsburgh
Gerasimos Fergadiotis, Ph.D., Portland State University
Benjamin Goudey, Ph.D., The University of Melbourne
Stephanie Gupton, Ph.D., The University of North Carolina at Chapel Hill
M. Cristina Vazquez Guillamet, M.D., Washington University in St. Louis
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
Vijaya Kolachalama, Ph.D., Boston University
Erica Larschan, Ph.D., Brown University
KiBum Lee, Ph.D., Rutgers University
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
Daniel Messinger, Ph.D., University of Miami
Pietro Michelucci, Ph.D., Human Computation Institute
Olga Jarrin Montaner, Ph.D., RN, FAAN, Rutgers University
Madeline Oguss, M.S., University of Wisconsin–Madison
Abraham Palmer, Ph.D., University of California, San Diego
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
Daniel Promislow, Ph.D., University of Washington
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
Judy Zhong, Ph.D., New York University School of Medicine

Workforce Development Recipient Team Members

Bettina Beech, Dr.P.H., M.P.H., University of Houston
Gary Benson, Ph.D., Boston University
Gundula Bosch, Ph.D., M.Ed., Johns Hopkins Bloomberg School of Public Health
Marino Bruce, Ph.D., M.S.R.C., M.Div., University of Houston
William Cress, Ph.D., Moffitt Cancer Center
Nick Duffield, Ph.D., Texas A&M University
Issam El Naqa, Ph.D., M.A., Moffitt Cancer Center
Raymond Esquerra, Ph.D., San Francisco State University
Matthew Gitzendanner, Ph.D., University of Florida
Catherine Grimes, Ph.D., University of Delaware
Karen Guillemin, Ph.D., University of Oregon
Michelle Hribar, Ph.D., Oregon Health & Science University
Ilona Jaspers, Ph.D., The University of North Carolina at Chapel Hill
David Julian, Ph.D., University of Florida
Ioannis Kakadiaris, Ph.D., University of Houston
Huining Kang, Ph.D., The University of New Mexico Health Sciences Center
David Kennedy, Ph.D., University of Massachusetts Medical School
Richard Kennedy, M.D., Ph.D., University of Alabama at Birmingham
Marianthi-Anna Kioumourtzoglou, D.Sc., Columbia University Mailman School of Public Health
Erin Kobetz, Ph.D., M.P.H., University of Miami Miller School of Medicine
Adam Labadorf, Ph.D., Boston University
Angie Laird, Ph.D., Florida International University
Craig Marcus, Ph.D., Oregon State University
Ofer Mendelevitch, M.S., Syntegra, Inc.
Blaine Mooers, Ph.D., The University of Oklahoma Health Sciences Center
Frank Penedo, Ph.D., University of Miami, Florida
Pleuni Pennings, Ph.D., San Francisco State University
Shawn Polson, Ph.D., University of Delaware
Julia Rager, Ph.D., The University of North Carolina
Travis Redd, M.D., Oregon Health & Science University
Jacob Searcy, Ph.D., University of Oregon
Katie Shilton, Ph.D., University of Maryland, College Park
Jeanette Stingone, Ph.D., M.P.H., Columbia University Mailman School of Public Health
Jian Tao, Ph.D., Texas A&M University
Swaroop Vedula, Ph.D., Johns Hopkins University
Ryan Weaver, Rensselaer Polytechnic Institute
Jennifer Westendorf, Ph.D., Mayo Clinic
Martin Zanaj, M.S., Oregon State University

Foundational Discussions in Artificial Intelligence–Readiness and Workforce Development: Joint Kickoff

November 1, 2021, 12 p.m. – 4:50 p.m. EDT

Attendees: NOT-OD-21-079 and NOT-OD-21-094 Recipient Team Members

Agenda

12:00 p.m. – 12:05 p.m.   **Welcome**
Dr. Allison Dennis, Program Coordinator, Artificial Intelligence (AI)
Programs, Office of Data Science Strategy (ODSS), NIH

12:05 p.m. – 12:45 p.m.   **NIH Office of Data Science Strategy Overview**
Dr. Susan Gregurick, Associate Director for Data Science and Director,
ODSS, NIH

12:45 p.m. – 1:00 p.m.   **ODSS AI Activities Overview**
Dr. Laura Biven, Branch Chief, Data Science Ecosystem, ODSS, NIH

1:00 p.m. – 1:10 p.m.   **AI-Readiness and Workforce Development Portfolio Showcase**
Dr. Alyssa Tonsing-Carter, Analysis and Evaluation Officer, ODSS, NIH

1:10 p.m. – 2:20 p.m.   **Breakout Discussion: Leadership, Workforce, and Readiness**

2:20 p.m. – 2:40 p.m.   **Break**

2:40 p.m. – 4:50 p.m.   **Breakout Discussion: Considering Ethics in the Data to AI/ML Research Cycle**

2:40 p.m. – 2:50 p.m.   Overview of NIH/ODSS Interest
Dr. Allison Dennis, Program Coordinator, AI Programs, ODSS, NIH

2:50 p.m. – 3:40 p.m.   Invited Speakers and Q&A
Dr. Katie Shilton, Associate Professor, College of Information Studies,
University of Maryland, College Park
Dr. James Zou, Assistant Professor, Biomedical Data Science,
Stanford University

3:40 p.m. – 4:35 p.m.   Group Breakout Discussion

4:35 p.m. – 4:50 p.m.   Recap and Conclusion

4:50 p.m. – 5:00 p.m.   **Close**

Foundational Discussions and Networking
for Artificial Intelligence–Readiness

November 2, 2021, 1 p.m. – 4:00 p.m. EDT

Attendees: NOT-OD-21-094 Recipient Team Members Only

Agenda

1:00 p.m. – 1:10 p.m.	**Introduction** *Dr. Allison Dennis, Program Coordinator, Artificial Intelligence (AI) Programs, Office of Data Science Strategy (ODSS), NIH*
1:10 p.m. – 1:50 p.m.	**Group Speed Networking**
1:50 p.m. – 2:00 p.m.	**Break**
2:00 p.m. – 3:30 p.m.	**Breakout Discussion: Readying diverse Data Types for AI/ML Applications**
3:30 p.m. – 3:50 p.m.	**Open Conversation of Cohort Support and Future Vision** *Dr. Laura Biven, Branch Chief, Data Science Ecosystem, ODSS, NIH*
3:50 p.m. – 4:00 p.m.	**Thank You and Closeout** *Dr. Laura Biven, Branch Chief, Data Science Ecosystem, ODSS, NIH*

Foundational Discussions and Networking
for Artificial Intelligence–Workforce Development

November 15, 2021, 12:00 p.m. – 3:00 p.m. EST

Attendees: NOT-OD-21-079 Recipient Team Members Only

Agenda

12:00 p.m. – 12:10 p.m.	**Introduction** *Dr. Allison Dennis, Program Coordinator, Artificial Intelligence (AI) Programs, Office of Data Science Strategy (ODSS), NIH*
12:10 p.m. – 12:50 p.m.	**Group Speed Networking**
12:50 p.m. – 1:00 p.m.	**Break**
1:00 p.m. – 2:30 p.m.	**Breakout Discussion: Building Core Competencies in Data Science for AI/ML for Trainees of Diverse Career Stage**
2:30 p.m. – 2:50 p.m.	**Open Conversation of Cohort Support and Future Vision** *Dr. Laura Biven, Branch Chief, Data Science Ecosystem, ODSS, NIH*
2:50 p.m. – 3:00 p.m.	**Thank You and Closeout** *Dr. Laura Biven, Branch Chief, Data Science Ecosystem, ODSS, NIH*