DRAFT – SUBJECT TO TASK FORCE APPROVAL ON DECEMBER 13, 2021

MEETING SUMMARY
National Artificial Intelligence Research Resource Task Force
Meeting #3

October 25, 2021

Meeting Summary

The third meeting of the National Artificial Intelligence Research Resource Task Force (NAIRR TF) was held online via Zoom on October 25, 2021, 11:00 AM–5:00 PM EDT.

Welcome and Administrative Remarks

The meeting started at 11:04 AM EDT.

Dr. Erwin Gianchandani (National Science Foundation), NAIRR TF Co-Chair, opened the meeting. Dr. Gianchandani entertained a motion to approve the minutes from the prior NAIRR TF meeting; the motion passed, and Dr. Gianchandani noted that the minutes will be updated online to reflect their approval.

Dr. Gianchandani then introduced the agenda, detailing the primary topics that would be taken up in the meeting, i.e., discussion of proposed working-group recommendations for compute capabilities and governance structure; and consideration of the decision space for data capabilities, access to testbeds and test resources, and provision of user resources such as educational tools and portal interfaces within the NAIRR framework.

Dr. Gianchandani highlighted that the Request for Information to gather public input on a NAIRR implementation plan closed on October 1 and that the 84 responses received are now available publicly on https://ai.gov/nairrtf/.

Dr. Lynne Parker (White House Office of Science and Technology Policy), NAIRR TF Co-Chair, provided an overview of the role of the TF Working Groups (WGs). The WGs are meeting in the weeks between TF public meetings to develop proposed recommendations within specific topic areas. Once agreed to by the full TF, these recommendations will provide the foundation for the TF interim report. Dr. Parker introduced TF member Dr. Dan Stanzione as the leader of the Compute Resource WG and TF member Dr. Fred Streitz as the leader of the Governance, Ownership, and Administration WG. She noted that Drs. Stanzione and Streitz would present the conclusions and proposed recommendations of their respective WGs and moderate separate conversations among TF members in the following sessions, then modify the recommendations accordingly following the meeting. If unresolved issues remain, a small group meeting will be convened to develop solutions to be presented at the December meeting.

The session ended at 11:18 AM EDT.

Readout and Discussion of Draft Recommendations: Compute Working Group

The session started at 11:19 AM EDT.

Dr. Stanzione presented the proposed recommendations of the Compute WG. The presentation introduced definitions of key terms, including production system, experimental system, research on AI, and research using AI. The WG proposed that:

1) The NAIRR be a federation of compute resources, and deployment of resources should be phased;
2) The scale of compute needed should be based on a determination of capability and capacity goals, as well as support for user-provided edge devices; and
3) The WG or full TF should develop metrics of performance and impact.

TF members discussed the presentation, responding positively to the set of proposed recommendations. Members discussed how capability and capacity goals could be in tension: designing a resource to support tackling the hardest problems in AI can decrease the ability of the resource to serve broader users. Members additionally discussed the role of user choice over which cloud resource to use, ways of leveraging existing federally-funded computing resources, and how to build effective metrics to measure impact.

The session ended at 12:21 PM EDT.

Readout and Discussion of Draft Recommendations: Governance Working Group

The session started at 12:21 PM EDT.

Dr. Streitz presented the proposed recommendations developed by the Governance WG. Dr. Streitz detailed pros and cons of structuring the NAIRR as either a public-private partnership or consortium, an NSF-style center awarded to one or more universities, or a new division or element under an existing government agency. The WG proposed that the NAIRR be created as a separate entity operated as a Federally Funded Research and Development Center (FFRDC) supported by the Department of Energy, the National Institutes of Health, the National Science Foundation, and perhaps other agencies. Dr. Streitz presented recommendations on policies, possible governance structures of a NAIRR operational entity, and approaches to access and allocate resources.

TF members discussed the recommendations. Members discussed how having multiple agencies chair a NAIRR could affect the appropriations process and how to design a hybrid process for researchers' access to the resource that would not add undue administrative burden for researchers with current federal grants and create pathways to access for researchers without existing federal grants. Also considered was the role that agencies providing researchers with grants should play in allocating NAIRR resources.

The TF members appreciated the thoughtful discussions leading to the recommendations presented and also agreed there would need to be further refinement as other aspects of the NAIRR are discussed in forthcoming meetings.

The session ended at 1:19 PM EDT.

Break: 1:19-1:40 PM EDT

Scene Setter: Data Needs of the AI Community

The session started at 1:42 PM EDT.

TF member Dr. Daniela Braga presented on the data needs of the AI community. Most data are generated by big corporations and public corporations, and much of those data are unstructured. Small and medium enterprises and citizen users do not currently have the resources to access third-party services as needed to make use of the data. Data needs include: access to trusted data, creation of data standards to facilitate interoperability, certifications for responsible AI, access to inclusive datasets, and standards for data privacy.

The session ended at 1:57 PM EDT.

Panel Discussion: Data Resources

The session started at 1:58 PM EDT.

The panel comprised the following individuals:

- Ian Foster, Director, Data Science and Learning Division, Argonne National Laboratory; Professor of Computer Science, University of Chicago;
- Robert Grossman, Founder, Open Data Commons Consortium; PI, NCI Genomic Data Commons;
- Ron Hutchins, Vice Provost for Academic Technologies, University of Virginia;
- Anita Nikolich, Research Scientist and Director of Research and Technology Innovation, School of Information Sciences, University of Illinois at Urbana-Champaign;
- Nancy Potok, CEO, NAPx Consulting; former Chief Statistician of the United States; and
- Andrew Trask, Leader, OpenMined.

TF member Dr. Julia Lane introduced the panel, and each panelist provided five minutes of opening comments. Panelists touched on: support for careful curation of large datasets; mechanisms to decide which data to prioritize; inclusion of users when tailoring the design of a data commons; support for creation and understanding of data policies at universities; establishment of clear legal authority over data collection, privacy, and confidentiality; creation of a sustainable set of policies and methods to access dispersed datasets; and leveraging privacy-enhancing technologies to create a federated structure of private datasets.

Following remarks from each panelist, Dr. Lane moderated a discussion. Panelists commented that much research can be conducted with de-identified and non-sensitive data. Panelists also discussed that data cleaning requires funding, and that people may clean and structure data in different ways, preventing interoperability. Discussion included the need for a process to decide which data to use and curate and mechanisms to incentivize collaboration and data cleaning.

The session ended at 2:49 PM EDT.

Briefing: Testbeds as Component of the NAIRR

The session started at 2:50 PM EDT.

Dr. Lisa Van Pay and Morgan Livingston (Science and Technology Policy Institute) presented findings from an examination of AI testing resources that could be made accessible via a NAIRR. The research showed that a range of testbeds and testing tools exist across the private sector, academia, and Federal Government. Resources include physical, virtual, and hybrid environments. Each type of environment introduces different considerations for privacy protection, security, intellectual property, user access, and use of open-source code. Incorporating existing testbeds into a NAIRR would require aligning incentives, policies, and technical designs with the testbed owners and administrators. Dr. Van Pay and Ms. Livingston also noted that testbeds can support and advance NAIRR objectives, including: supporting research on trustworthy AI; expanding the AI workforce; and enabling work on community-driven challenges such as climate change, access to datasets, and growing a diverse workforce.

The session ended at 3:10 PM EDT.

Discussion: Testbeds

The session started at 3:10 PM EDT.

TF member Dr. Andrew Moore moderated a discussion of the full TF on the role of the NAIRR in connecting researchers to testing and experimentation environments. Questions for the TF to consider include the role of federally-funded user facilities, testbed maintenance and upkeep, using physical facilities safely, the need for staffing, and the need for and availability of scaling resources. TF members discussed questions of intellectual property protection and ownership over the results of tests conducted using testbeds accessed via the NAIRR. Members also discussed how testbeds fit into the broader purpose of the NAIRR, the need to determine which testbeds would require compute and funding, and whether the NAIRR should itself support the creation of additional testbeds.

The session ended at 3:33 PM EDT.

Break: 3:33-4:00 PM EDT

Panel: User Resources – Portal Interface, Educational Tools

The session started at 4:00 PM EDT.

The panel comprised the following individuals:

- Tiziana Ferrari, Director, EGI Foundation;
- Kimberly Greene Starks, Global Lead, Infrastructure and Technology Strategy, IBM University Programs;
- Ana Hunsinger, Vice President for Community Engagement, Internet2; and
- Ed Lazowska, Professor and Bill & Melinda Gates Chair Emeritus, Paul G. Allen School of Computer Science & Engineering University of Washington.

TF member Dr. Fei-Fei Li introduced the panel, and each panelist provided five minutes of opening comments. The panelists discussed maximizing the usability of portal interfaces and providing effective educational resources. Lessons learned from past efforts include: meeting the users where they are, directing outreach efforts to groups historically unable to access these types of resources, enabling log-in

using university credentials, having two-factor authentication, providing cloud cost monitoring, leveraging existing trainings on how to use cloud infrastructure, and providing a community forum and help desk.

Dr. Li moderated a discussion among panelists. Suggestions offered included: building user personas to develop tailored education, providing data sovereignty to data scientists to ensure they control their data, bringing data closer to compute, continuously updating educational resources, and investing in cyberinfrastructure expertise (i.e., cyberinfrastructure professionals).

The session ended at 4:50 PM EDT.

Questions from Public and Meeting Close

The session started at 4:51 PM EDT.

Dr. Gianchandani addressed questions submitted by attendees via the Zoom Q&A interface, including about considerations for the technical architecture of the NAIRR; the NAIRR's integration with existing federal research infrastructure investments; data governance and algorithmic accountability within the NAIRR; and how the NAIRR can ensure responsible development and use of trustworthy AI.

Dr. Gianchandani announced a transition of his role as NSF Co-Chair to Dr. Manish Parashar (Office Director, Office of Advanced Cyberinfrastructure, National Science Foundation) due to Dr. Gianchandani's recent transition to a new role within NSF's Office of the Director.

Dr. Gianchandani concluded the session, thanking members of the TF, NSF, OSTP, STPI, and the public, and reminding everyone that meeting summaries, slide presentations, and details about upcoming meetings can be found at https://www.ai.gov/nairrtf/. He noted that the next meeting is scheduled for Monday, December 13, and details are posted to the Federal Register.

The meeting adjourned at 5:03 PM EDT.

Appendix I: Attendance for NAIRR TF Meeting 3

TF Members Present:

Erwin Gianchandani, National Science Foundation (Co-Chair)

Lynne Parker, White House Office of Science and Technology Policy (Co-Chair)

Daniela Braga, DefinedCrowd

Mark Dean, retired (formerly IBM and University of Tennessee, Knoxville)

Oren Etzioni, Allen Institute for AI

Julia Lane, New York University

Fei-Fei Li, Stanford University

Andrew Moore, Google

Dan Stanzione, University of Texas, Austin

Frederick Streitz, Department of Energy

Elham Tabassi, NIST

TF Members Absent:

Michael Norman, University of California, San Diego