

MEETING SUMMARY

National Artificial Intelligence Research Resource Task Force

Meeting #9

September 12, 2022

MEETING SUMMARY

Meeting Summary

The ninth meeting of the National Artificial Intelligence Research Resource (NAIRR) Task Force (TF) was held online via Zoom on September 12, 2022, 11:04 AM–5:02 PM EDT.

Welcome and Administrative Remarks

The meeting started at 11:04 AM EDT.

Dr. Manish Parashar, NAIRR TF Co-Chair, opened the meeting. He introduced the new TF Co-Chair, Ms. Tess deBlanc-Knowles, who has been working with the TF since its inception and is taking over for Dr. Lynne Parker. As planned, Dr. Parker successfully completed her four-year rotation at The White House Office of Science and Technology Policy (OSTP) on August 19th.

Dr. Parashar then provided an overview of the meeting agenda, noting that it would include presentations and discussions from TF Working Groups (WGs) on: Security, Access Controls and User Portal; Resource Allocation, Usage Policies, and Evaluation; and Environmental Sustainability, International Collaboration, and Legal Considerations. The agenda also included a panel on related Federal efforts to provide data and computational resources to the research community and a TF discussion of open issues on which consensus might still be needed. Dr. Parashar noted that time would be reserved at the end of the meeting for public questions, and that no questions had been held over from the previous meeting.

Dr. Parashar called for a motion to approve the summary from the eighth NAIRR TF meeting held on July 25, 2022; the motion passed.

The session ended at 11:14 AM EDT.

Readout and Discussion of Proposed Plans: Security, Access Controls, and User Portal

The session started at 11:14 PM EDT.

Dr. Elham Tabassi, NAIRR TF member, presented recommendations of the WG focused on defining the implementation details associated with NAIRR security protocols, access controls, and the user portal. The WG recommended the creation of two zones within the NAIRR: NAIRR-Open, a lower-security zone devoted to open science and open data, and NAIRR-Secure, a higher-security zone for research involving protected or restricted data. These zones would have separate user portals, identity management systems, security policies, and access management controls.

The user portal of NAIRR-Open would be widely used for resource discovery and selection, job submission and monitoring, data management, and usage monitoring, as well as serving as a common access point for training materials, technical assistance, and user forums. The user portal of NAIRR-Secure would perform similar functions while allowing access to a secure enclave. Within the secure enclave, user access

would be controlled and monitored through a virtual desktop to ensure compliance with relevant data controls. In addition to NAIRR-Open and NAIRR-Secure, NAIRR would include a “sandbox” that would provide flexible access and security for researchers’ needs not met by the other two zones.

Following the presentation, TF members discussed whether and how to accommodate data protected by non-U.S. law, the potential for alignment between the activities of the NAIRR and the National Secure Data Service (NSDS), possible timelines for standing up NAIRR-Open and NAIRR-Secure, and how the sandbox would function.

The session ended at 12:09 PM EDT.

Readout and Discussion of Proposed Plans: Resource Allocation, Usage Policies, and Evaluation

The session started at 12:09 PM EDT.

Dr. Fei-Fei Li, NAIRR TF member, opened the session, handing off to NAIRR TF members Dr. Dan Stanzione and Dr. Julia Lane to walk through the proposed recommendations developed by the WG. The WG suggested that resource allocation and usage policies could be adapted from existing models, especially from computing and data facilities, though they would need to have some unique aspects because of the specific needs of AI R&D. With respect to resource allocation, the WG suggested that the NAIRR should incorporate three resource tracks: allocations to researchers funded through participating agencies; allocations by the NAIRR Operating Entity (OE), with a “startup” track for limited credits and a peer-reviewed track for larger requests; and allocations for fees (i.e., a “fee-for-service” approach). The WG also suggested that all NAIRR use should be through credits for computing time at NAIRR facilities rather than funding transfers that would be subject to indirect costs.

With respect to usage policies, the WG suggested that the NAIRR once again should draw upon existing models. Both the OE and the individual resource providers (RPs) would be responsible for operational monitoring. Researchers, data providers, and data users would all be subject to NAIRR policies intended to promote collaboration and shared use. The intellectual property, liability, and accuracy of data would be the responsibility of the NAIRR’s data providers rather than the NAIRR or the NAIRR OE itself.

Dr. Lane presented on the evaluation of the NAIRR. The WG suggested that the NAIRR should adopt a standard evaluation framework and should trace standard measures, such as publications and patents, but also new measures, such as student/postdoc placement, startups enabled, productivity, and growth. At startup, the WG suggested that the NAIRR should establish initial benchmarks in the context of a counterfactual. The WG also suggested that evaluation should occur at four levels: 1) the NAIRR as a whole; 2) the NAIRR OE; 3) each RP; and 4) each agency’s portfolio of research funded to use the NAIRR. Finally, the WG noted that an external entity should be responsible for designing, gathering the data for, and implementing the evaluation.

TF members discussed the need for developing mechanisms for monetizing credits that allow the OE to accept research, budget compute use, and reimburse public cloud providers. A second topic discussed was the mechanism by which the “agency track” would fund research. The WG recommended that agencies should include a review of NAIRR compute credits along with the research review to avoid a double-review requirement, while recognizing that some agencies might instead instruct researchers to apply directly to NAIRR for compute resources. A third topic of discussion focused on the benchmarking portion of the evaluation, especially with respect to identifying a comparison group and the need to identify the baseline for the NAIRR and AI as a field.

The session ended at 1:10 PM EDT.

Break: 1:10–1:40 PM EDT

Readout and Discussion of Proposed Plans: Environmental Sustainability, International Collaboration, and Legal Considerations

The session started at 1:40 PM EDT.

NAIRR TF member, Dr. Mark Dean, focused on proposed ideas for how the NAIRR might minimize environmental impact, support international collaboration, and address legal considerations including statutory authorities and data use agreements (DUAs).

To mitigate the impact of NAIRR resources on the environment, the WG proposed that the NAIRR have an environmental policy that establishes and tracks metrics for total cost to the environment and E-waste management. The NAIRR should also promote the importance of AI research areas relevant to the environment, and track and report the percentage of time that the NAIRR infrastructure is used for this research. TF members also discussed the need to improve the efficiency of software and provide relevant tools and efficiency criteria, as well as the ability to monitor improvements in this area.

The WG recommended that the NAIRR follow existing guidelines for international collaboration and suggested that a specific outreach effort may be needed to coordinate across stakeholder groups. The TF discussed where this effort could be housed within the proposed NAIRR governance structure, concluding that the partnerships and outreach group within the OE could serve this role if sufficiently resourced.

The WG explored the idea of crafting a single DUA in collaboration with a key data provider such as the NSDS, but acknowledged this has been a challenge and that establishing a simple DUA, or pair of DUAs for use of open and protected data, respectively, might be more easily accomplished. Addenda that define the additional terms necessary to share protected data may also be needed.

In terms of statutory authorities, the WG recommended that the NAIRR be established with similar statutory authorities and authorizing legislation as the High Performance Computing and Communications Program created by the High Performance Computing Act of 1991 (renamed to the Networking and Information Technology Research and Development [NITRD] Program via the American Innovation and Competitiveness Act of 2017).

The session ended at 2:11 PM EDT.

Questions from the Public

The session started at 2:11 PM EDT.

Dr. Parashar moderated a discussion among TF members addressing questions submitted by public attendees via Zoom's Q&A portal, noting that the AI Bill of Rights effort is being run by OSTP, not the NAIRR TF, and discussing privacy and copyright issues associated with datasets.

The session ended at 2:20 PM EDT.

Open Issues and Implementation Plan Development Process: Session 1

The session started at 2:20 PM EDT.

With the close of the previous readout and discussion session, the NAIRR TF began a discussion of open issues at 2:20 PM, moderated by Ms. deBlanc-Knowles. The session began with a discussion of the NAIRR's role in hosting data. Dr. Lane provided the overarching position that the NAIRR will host curated data through its RPs, and that such data could include public and open datasets, seeded with high-value/high-

use open data, and data resulting from NAIRR-supported research. Decisions on what data to host would be made by the NAIRR OE using a transparent allocation process based on the availability of resources. Any data hosted would be decommissioned using cost/benefit analysis. The TF then discussed several ideas related to data hosting, including:

- The role of the NAIRR OE in data curation and the extent to which data will need to be “AI-ready” before being hosted by the NAIRR;
- Metadata standards for these NAIRR datasets;
- Checklists for assessing whether datasets are “AI-ready”;
- The nature of incentives for data sharing;
- Ownership of research and datasets produced through NAIRR research; and
- Proactive work by the NAIRR OE to identify high-value datasets and to make them ready and available for research.

The session ended at 2:42 PM.

Panel: Associated Federal Efforts for Provision of Data and Computing Resources

The session started at 2:43 PM EDT.

Dr. Erwin Gianchandani, Assistant Director for Technology, Innovation and Partnerships at the National Science Foundation and prior NAIRR TF co-chair, introduced the panel comprising the following speakers:

- Shelly Martinez, Senior Statistician, Office of Management and Budget;
- Vipin Arora, Deputy Director, National Center for Science and Engineering Statistics, National Science Foundation;
- Dr. Parashar;
- Kamie Roberts, Director, National Coordination Office for the Networking and Information Technology Research and Development Program; and
- Jerry Sheehan, Deputy Director for Policy and External Affairs, National Library of Medicine, National Institutes of Health.

Panelists provided backgrounds on their efforts related to provision of computational and data resources and discussed their visions for working with the NAIRR. Initiatives and concepts discussed included: the National Secure Data Service, America’s Data Hub, the Future Advanced Computing Ecosystem roadmap and implementation plan including a National Strategic Computing Reserve, National Discovery Cloud, and open science infrastructure.

Following remarks from each panelist, Dr. Gianchandani moderated a discussion with TF members. TF members asked panelists to share ways to build collaboration into the initial NAIRR infrastructure, which produced the following comments:

- NSDS could engage with NAIRR as an RP;
- Some of the datasets of interest for NAIRR are not core datasets for NSDS;
- Consider existing infrastructure elements upon which the NAIRR could be overlaid;

- NAIRR design and deployment efforts should be coordinated with related cyberinfrastructure efforts, which will require some flexibility as these efforts proceed and evolve; and
- The need for such coordination should be emphasized in the final report, even if details cannot yet be prescribed.

The session ended at 3:45 PM EDT.

Break: 3:45–4:10 PM EDT

Open Issues and Implementation Plan Development Process: Session 2

The session started at 4:10 PM EDT.

Ms. deBlanc-Knowles continued to lead the discussion of open issues, aimed at gaining consensus on aspects of the NAIRR implementation plan. The discussion began with a review of the NAIRR organizational structure, which would include an interagency Steering Committee headed by a lead agency; a Project Management Office to be hosted within the lead agency; a NAIRR OE (which could be led by a university, non-profit, or a company) that will directly manage the NAIRR's core functions; and a User Committee and three Advisory Boards (Ethics, Science, and Technology) to provide oversight, advice and direction. Dr. Lane also walked through the initial steps recommended for standing up the NAIRR governance structure: 1) Congress funds the NAIRR; 2) the Steering Committee, lead agency, and Project Management Office are created; 3) solicitations are issued to identify the OE and RPs; and 4) the NAIRR OE, overseen by the Project Management Office and advised by the Advisory Boards, begins to operate the NAIRR. The NAIRR TF discussed several topics related to governance, including:

- Whether the NAIRR OE should conduct research independently;
- Whether the NAIRR OE should build the user portal;
- Whether the NAIRR OE should be allowed to bid to be an RP, especially for secure data; and
- How metrics and evaluation will be incorporated into this governance framework.

Dr. Stanzone led the next topic of discussion focused on capabilities that the NAIRR should make available at the start of its pilot operations (its "Day One" capabilities). He identified the following required capabilities at "Day One": 1) a user access portal with allocation and identity management systems; 2) a data publication system that allows datasets to be added; 3) linked indices of resources (e.g., testbeds, education/training materials); 4) Compute RPs accepting computational jobs including a mix of on-premises and cloud computing options; 5) a high-speed network; and 6) sufficient memory capacity. The TF discussed and agreed on inclusion of the following in the definition of "Day One" capabilities: the software environment and tools, responsible AI guidance, a user help system and support staff. The TF also discussed and agreed whether initial capabilities would primarily consist of extensions of existing resources or newly-established resources, as well as the amount of time it would take the NAIRR OE to reach "Day One" operations.

A final topic, also framed by Dr. Stanzone, concerned the NAIRR's user capacity. The TF has identified that the NAIRR's "Day One" capacity should be 48–60 million hours on quad-GPU nodes, which could support 50,000 student/research projects (of 1,000 hours/user) or 25,000 student/research projects plus 40 projects at the scale of the GPT-3 benchmark. At steady state, the NAIRR's capacity should be 140–180 million hours on quad-GPU nodes, which would permit triple the "Year One" capacity. The TF discussion addressed:

- The need for the NAIRR TF to also specify capacity of testbeds and more experimental environments (e.g., Internet of Things, cyber-physical systems);
- The balance between capacity and capability systems within the NAIRR; and
- How to compare this number of users with the current baseline of U.S. AI students and investigators.

This session ended at 4:55 PM EDT.

Questions from Public and Meeting Close

The session started at 4:55 PM EDT.

The TF discussed comments from members of the public related to approaches to develop a baseline for the number of AI researchers in the United States. Three questions from the public were answered in real time in the preceding sessions.

The next meeting is scheduled for October 21, 2022, to discuss the draft NAIRR TF Final Report and Implementation Plan. Details will be posted in the Federal Register.

The meeting adjourned at 5:02 PM EDT.

Appendix I: Attendance for NAIRR TF Meeting 9

TF Members Present:

Manish Parashar, National Science Foundation (Co-Chair)

Tess deBlanc-Knowles, White House Office of Science and Technology Policy (Co-Chair)

Daniela Braga, Defined.ai

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

Julia Lane, New York University

Fei-Fei Li, Stanford University

Michael Norman, University of California, San Diego

Dan Stanzione, University of Texas, Austin

Frederick Streitz, Lawrence Livermore National Laboratory

Elham Tabassi, National Institute of Standards and Technology

TF Members Absent:

Andrew Moore, Google

Oren Etzioni, Allen Institute for AI

Other Contributors

Erwin Gianchandani, National Science Foundation

Lynne Parker, White House Office of Science and Technology Policy (Consultant)

Science and Technology Policy Institute Staff Present:

Emily Grumbling

Lisa Van Pay

Brian Zuckerman

Matt Ishimaru

Matt Christman