Request for Information (RFI) on Implementing the Initial Findings and Recommendations of the National Artificial Intelligence Research Resource Task Force: Response

University of Southern California (USC) Information Sciences Institute (ISI)

DISCLAIMER: Please note that the RFI public responses received and posted do not represent the views or opinions of the U.S. Government nor those of the National AI Research Resource Task Force, and/or any other Federal agencies and/or government entities. We bear no responsibility for the accuracy, legality, or content of the responses and external links included in this document.
These comments are submitted on behalf of the University of Southern California’s Information Sciences Institute.

Response to Recommendation 3-1: Multiple Federal agencies should be funded to cooperatively support NAIRR resources and management, thereby serving the broadest range of research communities and national interests.

There is merit to funding several federal agencies to contribute to NAIRR. This would ensure that all are vested in the effective growth and use of the resource. Indeed, many agencies have served the research community over the years in different ways: the NSF successfully established national supercomputing centers and has supported them for many years to serve the large scientific community; the DoE has provided supercomputing resources for use by the community; and NASA and USGS have been community providers of data. However, a loose independent collection of agencies contributing resources will be a risky approach. Recognizing that selecting a single agency to create the NAIRR would result in a single point of failure for NAIRR, a fully distributed model as stated in the recommendations would be unworkable.

To mitigate this risk, a sensible approach would be to select a small number of agencies to have oversight over the resource. These agencies would have a proven record for serving the community with data and computing resources. An example of a balanced approach could have NSF oversee computing resources for academia, DoE oversee computing resources for government, and NASA oversee data sharing aspects of the NAIRR.

Response to Recommendation 3-4: The day-to-day operations of the NAIRR should be managed by an independent, non-governmental entity with dedicated, expert staff.

Putting NAIRR operations in the hands of a single organization would result in a single point of failure for the entire enterprise. Given that NAIRR is creating a fundamentally new resource like no other, reducing risk should be a major consideration.

To reduce this risk, setting up a consortium of partner organizations would be a reasonable approach. The Open Science Grid (OSG) is an example of a successful consortium to enable sharing of computing resources to serve the community.
Alternatively, a few (3-5) organizations could be responsible for day-to-day operations. Each could be better positioned to serve some segment of the community.

The managing organization(s) of NAIRR should not be in a position of serving many masters. If NAIRR were to be funded by many government agencies then that would make their success more challenging.

Response to Recommendation 3-10: Access to NAIRR resources should be contingent on research project proposal review, be governed by clear use policies and user agreements, and be in compliance with relevant requirements for open sharing of research outputs.

There should be provisions to tie some of the NAIRR allocations more explicitly to research awards. Since research proposals require a description of the resources available to do the work, it would be difficult to obtain funding for ambitious research that ultimately would require resources on the NAIRR. Perhaps a process of pre-approval for using NAIRR for a research proposal if funded would be warranted.

Response to Recommendation 4-1: The NAIRR should coordinate a network of trusted data and compute providers and hosts for a robust, transparent, and responsible data ecosystem.

NAIRR planning should include funding to support AI research on data sharing and data integration. Data sharing has been traditionally challenging, and AI research offers many possibilities to address those challenges by automating data modeling, data integration, and data analysis. In other words, AI presents an opportunity to address traditional challenges in data sharing that have been found in similar efforts at NIH and other agencies. Explicit allocations of funding to support AI research for data sharing should be part of the planning for NAIRR. These funds would not necessarily need to be managed by NAIRR, and could be allocated to funding agencies such as NSF that traditionally fund basic AI research.

NAIRR planning should also include provisions for building on successful AI approaches for data sharing and data integration. WikiData is one such approach that should be considered.

Response to Recommendation 4-5: The NAIRR ecosystem should make the most of community
access by incentivizing the contribution of high-quality data for AI R&D to the federated system.

The Obama administration’s directive for open government data with no additional resources to federal and local agencies resulted in repositories with thousands of datasets that were not very usable. NIH and NSF efforts to create community repositories of shared data have also shown that the effort to share high-quality well-documented data is not always affordable for data providers. It is also not clear that data providers have the knowledge or skills to do the necessary work properly. Therefore, relying on “incentives” for doing what is an enormous amount of work that requires special skills is repeating the same mistakes from the past. It is time to create better approaches for sharing and integrating data through new AI research, so that data sharing becomes cheaper and more scalable. When human effort is needed, there should be a plan to fund the work that includes consideration of the costs and benefits of the target datasets.

The University of Southern California’s Information Sciences Institute (ISI) carries out basic and applied research in artificial intelligence, networks and cybersecurity, high-performance computing, microelectronics, and quantum information systems. Its $100M annual external funding comes from the NSF, DoD, IC, NIH, DoE, industry, foundations, and other sponsors. ISI is home to the first quantum computer in academia. Part of the USC Viterbi School of Engineering, ISI has more than 400 personnel that includes 28 faculty that advise 65 PhD students. ISI’s Artificial Intelligence Division is one of the largest AI research groups in the U.S. ISI’s AI systems for machine translation, online misinformation detection, and data-centric AI are first-rate and have been deployed to support many parts of DoD and hundreds of law enforcement agencies. Some of ISI’s commercial spinoffs were acquired for tens of millions and contribute to a vibrant innovation ecosystem in Southern California.