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

Data Foundation

Data Coalition
1100 13th Street NW
Suite 800
Washington, DC 20005

📞 312.493.7533
🐦 @DataCoalition
✉ info@datacoalition.org
🏠 datacoalition.org

June 30, 2022

RE: Feedback on the findings and recommendations put forward in the NAIRR Task Force's Interim Report

To Whom It May Concern:

The Data Foundation is a non-profit organization that seeks to improve government and society by using data to inform public policymaking. Our Data Coalition Initiative is America's premier voice on data policy, advocating for responsible policies to make government data high-quality, accessible, and usable. Ensuring reasonable, responsible, and ethical practices are implemented in legislative and administrative activities is a priority for the Data Coalition, thus we work to promote strategies for meaningful artificial intelligence (AI) advancements in government.

The Data Coalition applauds the NAIRR Task Force for the publication of its interim report. The report is a productive step in the long journey toward making access to data for AI research equitable, secure, and effective for use across sectors. We fully support the goals, objectives, and overall vision of the NAIRR. The following response expands on three areas identified in the request for feedback, emphasizing aspects from the interim report that we consider critical to include in the final report, and highlights areas where additional elaboration may be helpful.

B. Establishment and sustainment of the NAIRR. Including agency roles, resource ownership and administration, governance and oversight, resource allocation and sustainment, and performance indicators and metrics. (Chapter 3)

The identified management and governance structure of the NAIRR as a federated cybersecurity ecosystem is conducive to ensuring the proper independence, oversight, and transparency of a national AI R&D tool. Incorporating a tiered access system and performance measures will both encourage use and highlight the value of a shared data infrastructure.

The more NAIRR data is used, the more there will be opportunities to identify flaws and solutions, strengthening the functions and quality of the data and research products stemming from a NAIRR. Recommendations such as 3-13 to incentivize use, and recommendation 3-17 to evaluate the impact of NAIRR use, are important aspects to ensure NAIRR data is functioning as envisioned and contributing meaningfully to future AI R&D.

Further, as seen in recommendation 3-16, calling for external evaluation is important for transparency, accountability, and ensuring effectiveness of a NAIRR. Without identifying external evaluators, evaluations may be limited to a small number of evaluators close to the

project, potentially influencing the findings or direction of evaluations that should be objectively conducted.

Also relevant to sustainability and governance, while the report acknowledges the need for Congress to fund NAIRR through appropriations and suggests the management entity can explore other revenue streams for long-term sustainability, it would be beneficial to provide a cost estimate. Including an estimate of how much funding NAIRR will need from Congress can facilitate further advocacy around NAIRR implementation to ensure it has the funding to achieve the Task Force's vision.

C. NAIRR resource elements and capabilities. Including data, government datasets, compute resources, testbeds, user interface, and educational tools and services. (Chapter 4)

The report highlights that "[i]ncreased access and diversity of perspectives would, in turn, lead to new ideas that would not otherwise materialize and set the conditions for developing AI systems that are inclusive by design." We see this as a priority for the NAIRR and we support the explicit recommendations expanding access to and use of data resources across the government.

Seen in recommendations 4-24 through 4-26, the Task Force's provisions for software, training, and educational resources to support a diverse set of users with varying levels of proficiency is important to achieving the vision of a NAIRR. Supporting a career pathway – rather than a career pipeline – can help expand and better ensure inclusion, access, and equity in a data workforce. Incorporating these educational resources and activities are key factors for a NAIRR to broaden the contributions to AI R&D.

The search and discovery portal for data is also an important function of the NAIRR, seen in recommendation 4-8 as well as 2-3. Providing a single access portal that incorporates existing data repositories from across the government, rather than duplicating efforts, can expand data access and use, leverage existing data capacity, improve overall government efficiency, and facilitate contributions to AI R&D that may have previously been excluded. We would like to reiterate the need for a streamlined, standardized, searchable, and accessible portal to achieve the full potential of a NAIRR as identified in the goals and objectives. We also encourage the task force to connect with other single application portal initiatives taking place within the federal statistical system, as well as the Advisory Committee on Data for Evidence Building.

Similarly, coordination of open data plans, among other efforts, seen in Recommendation 4-11 is crucial to leverage current advancements in AI R&D and government-wide data use and data sharing. A NAIRR that aligns with the Foundations for Evidence-based Policymaking Act requirements will bolster the federal data infrastructure and can serve as an important model that may be applied in other research areas. All of these steps to ensure accessible and open data will improve transparency and help promote trustworthy AI initiatives.

E. Privacy, civil rights, and civil liberties requirements. (Chapter 6)

The role of the NAIRR management entity to implement steps to ensure transparency, access for diverse users, and proper oversight as well as demonstrate how research using the NAIRR is being reviewed, approved, and performed are all critical to protecting civil rights and liberties. We support these steps and believe by expanding upon NAIRR data curation processes and identifying standards for documenting bias both within the research process and the data itself, the NAIRR can further bolster responsible use of AI.

Overall, we agree with with the interim report that a "NAIRR would transform the U.S. national AI research ecosystem by strengthening and democratizing foundational, use-inspired, and translational AI R&D in the United States." As the Task Force prepares its final report, we encourage continued emphasis on expanding data access and use, developing a search and discovery portal, and ensuring data and research processes are without bias or inaccuracies.

Thank you for the opportunity to provide comments on this very important issue, and we hope to continue to support your efforts to support ethical and useful AI advancements in government. Please contact me at corinna@datafoundation.org if you have any questions or would like to discuss the Data Coalition's interest in this matter further.

Sincerely,

Corinna Turbes
Managing Director
Data Coalition