Federal Register Notice 86 FR 46278, https://www.federalregister.gov/documents/2021/08/18/2021-17737/request-for-information-rfi-on-an-implementation-plan-for-a-national-artificial-intelligence, October 1, 2021.

Request for Information (RFI) on an Implementation Plan for a National Artificial Intelligence Research Resource: Responses

Executive Summary

NiyamIT, Inc. (Niyam) greatly appreciates the opportunity to submit this Response to Request for Information for a National Artificial Intelligence Research Resource (NAIRR). Niyam is an **SBA Certified 8(a)** and **HUBZone** Small Disadvantaged Business, founded in 2007 by a group of consultants who shared a unique vision: a technology company steeped in an orderly process, yet driven by passion and innovation. Today, Niyam leads the way in developing and delivering mission-critical technologies using Artificial Intelligence (AI), Data Science/Analytics and Management, Geospatial Information Systems (GIS), Software Development, Agile/DevSecOps. Our solutions are proven to increase efficiency, streamline process flows, accelerate collaboration, and consistently provide breakthrough results - all while adapting and responding to the realities of shifting timelines and budgets. Niyam is ISO 27001:2013, ISO 9001:2015, and ISO 20000-1:2018 certified and CMMI-DEV Level 3 appraised. We strive to be a trusted partner to National Science Foundation (NSF) and the Office of Science and Technology Policy (OSTP), offering our commitment to delivering positive, measurable outcomes in support of your mission.

1. Task Force Considerations

A. Goals for Establishment and Sustainment

Large-scale digitization of information resulting from IT modernizations has leveraged data as a very valuable resource. Harnessing large volumes of data, processing it, and generating quick actionable insights surpasses human capabilities. Machine learning (ML) and Artificial Intelligence (AI) can extend human capabilities to model “never before” scenarios. Gaining a strategic advantage with AI will help the county create proactive solutions to large scale unprecedented complex problems like pandemics, loss of lives and property due to intensifying natural disasters, shifting geopolitical polarization, cyber and physical security threats, and volatile economies. In recognition of this fact, most government agencies have established AI advisory boards or similar inter-agency groups to study the potential of AI. Funding has been provided to research groups from multiple universities to develop prototypes.

The **primary goal** of the National Artificial Intelligence Research Resource (NAIRR) should be to mobilize the AI groups in government agencies to create a centralized, unified “think tank”. Such joint effort will be cost-effective, prevent duplication of efforts, and enable collaboration on a larger scale. The table in appendix A provides details on the probable transformative role of AI in each government agency. AI research is a data-driven effort, requiring specialized compute infrastructure, ML expertise, IT support staff, and strong governance support. NAIRR’s unified research platform will serve as a convergence of the best minds in the academia and varied data sets from member government agencies, compute infrastructure purpose-built for AI/neural/deep learning computations in a highly secure environment and governed by top policymakers. **Long term sustainment** of NAIRR largely depends on the collaborative efforts of government agencies and their collective drive to succeed.

Quantifying AI research success with metrics is challenging and nuanced, except in cases of genuine breakthroughs. Some platform-centric metrics of NAIRR can be indicative of success:

- No. of individuals, institutional memberships, and workspaces on NAIRR, indicating user engagement and growth.
- No. of deployable AI modules produced, or no. of milestones, indicating the progress of research objectives.
- Feedback scores from grantors for research projects indicating fair and appropriate use of the platform.

Niyam's AI/ML Solution

- Niyam developed custom-built AI/ML tool Flood Assessment Structure Tool (FAST) to compute flood risk, increasing speed and accuracy.
- FAST analyzes NYC 800,000 structures for 100-year flood losses in less than 8 seconds (previously it was >10 hours), 275x faster than existing traditional software.

B. Ownership and Administration

An AI working group established by the NAIRR Task Force should oversee the operations of AI subgroups, established in each participating government agency. These agency AI subgroups should be responsible for research initiatives piloted by the parent government agency. The AI subgroup can be owned by the participating agency. This agency subgroup will define research problems and invite new ideas/proposals for transformative AI to be published as grant opportunities that will be made available to accredited universities. After successful evaluation, the grants administration should be handed over to an agency similar to grants.gov.

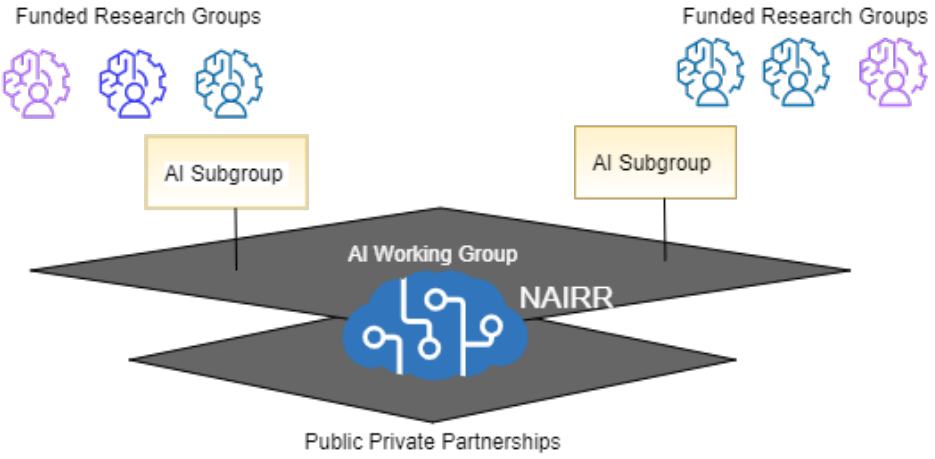

Figure 1. Ownership and Administration of NAIRR

Responsibilities of agency-specific AI subgroup:

- Publish research opportunities on the NAIRR platform for aspiring research groups
- Evaluate the feasibility of submitted proposals and grant approvals
- Disburse funds for NAIRR to the AI working group with expenses calculated based on usage requirement of NAIRR platform that includes data sets, algorithms, and compute times. (This replaces the need for government agencies to fund individual research groups)
- Track research progress report to AI working group

- Identify platform insufficiencies such as compute requirements and unavailability of data, connectivity, availability, and suggest improvements to the NAIRR working group.

C. Model for Governance and Oversight

Governance and oversight objectives of the AI working group should be to enable the use of high-quality data for ethical AI research on a secure and robust platform while protecting privacy and adhering to the compliances as applicable in each business domain. Governance should not be a blocker, but an enabler to innovation. A balance between governance and enablement must be established to foster innovation. Research projects on NAIRR should be given autonomy for exploration and experimentation. In turn, they need to provide 100% visibility into the data sets being utilized, the purpose of the research, and adherence to privacy and compliance laws. The 5 anchors for oversight and governance are:

- High-quality data: The U.S Open Government Directive required that all agencies post at least 3 high-value data sets online and register them on data.gov within 45 days. As an extension of this directive, the NAIRR working group should procure fresh data sets from participating government agencies regularly.
- Ethical AI: Research projects on NAIRR would have the potential to have a high impact on the daily lives of the public. Ongoing scrutiny is crucial to ensure that the NAIRR is committed to enabling unbiased and trustworthy AI research.
- Compliance & Privacy: A unified charter of privacy and compliances requirements (e.g., HIPAA, SOC, PCI) should be jointly maintained by member government agencies. All AI research projects on NAIRR should fulfill these requirements for continued access.
- Security: Security threats to NAIRR exist at dual levels, at data repository level and programmatic/model level. Many AI systems could be composed of open-source libraries with security vulnerabilities. There exists a possibility of “inversion attacks” where an AI model can be hacked, revealing information about itself and the data that it was trained on. Such threats should be identified and monitored to prevent misuse of the NAIRR.
- Platform Infrastructure: Factors that lead to high-performing AI platforms are high computing capacity, storage capacity, and networking infrastructure. NAIRR working group should monitor and adapt the platform infrastructure to meet the needs of the research community.

D. Creating and Maintaining Shared Computing Infrastructure

Resources required for data storage/preparation are different in scale and nature from those required for AI computation. Therefore, infrastructure for storage and computation can be decoupled from each other.

Data storage and processing requirements: The NAIRR platform will have to plan for a massive storage repository of data needed for ML model training, and a high velocity of incoming data streams for model inference and predictive analytics. Data sets for ML and AI can reach hundreds of terabytes to petabytes. Data consolidation from multiple

sources, selection, and preprocessing, such as filtering, categorization, and feature extraction which are the primary factors contributing to a model's accuracy and predictive value are significant factors that will influences storage and processing infrastructure.

Computing resources: The technology that powers AI is unique and always evolving. Specialized hardware, such as GPUs is critical for machine learning (ML) and subsequently AI. NAIRR should evaluate the pros and cons of a "build vs buy" approach. Enterprise AI has multiple success stories that feature high-level ML and Deep learning (DL) services like AWS SageMaker, Azure Cognitive Services, Google Cloud Machine Learning Engine, H2O.ai, IBM Watson Studio, and ML, etc. Among the multiple choices available for architecture and deployment for AI, hyper-converged infrastructure (HCI) systems offer the most density, scalability, and flexibility. Most AI systems run on Linux VMs or as Docker containers. Some popular AI frameworks and many sample applications are available as prepackaged container images from Nvidia and other vendors. Some of these applications include Computer vision such as image classification, object detection, image segmentation, and image restoration, speech and natural language processing, recommendation systems that provide ratings or products based on prior user activity, content analysis, filtering and moderation, pattern recognition, and anomaly detection. These applications can be used in a variety of business domains like fraud analysis, surveillance systems for physical security, geologic analysis for oil, gas, and other forms of energy, cybersecurity, and automation, etc.

The system components most critical to AI performance are:

- **CPU -** Responsible for operating the VM or container subsystem, dispatching code to GPUs, and handling I/O. E.g. Second-generation Xeon scalable platinum or gold processor, AMD Epyc CPUs.
- **GPU -** Handles ML or DL training and inferencing (ability to automatically categorize data based on learning) E.g., Nvidia P100(Pascal), V100(Volta), or A100(Ampere) for GPU training and V100, A100, or T4 (Turing) for inference.
- **Memory -** AI operations run from embedded high bandwidth GPU memory that is much faster than conventional DRAM.
- **Network -** AI systems are often clustered together to scale performance and are connected with 10Gbps or higher Ethernet interfaces. They can also include InfiniBand or dedicated GPU (NVLink) interfaces for intra-cluster communications.

Portal-based educational Tools and services: NAIRR can provide members access to prebuilt algorithms. The portal can also support virtual events to host prototype presentation opportunities regularly. Prototypes can be scored based on technical merit and award grant opportunities to winning scores. "Citizen scientist programs" can be initiated to generate interest among aspiring entrepreneurs. Educational seminars and certification tracks can be developed as revenue-generating resources. Video tutorials, wikis, blogs, and chatbot-based help and forums powered by an expert knowledge base will help researchers achieve their goals.

E. Barriers to High-quality Government Data Sets

Government agencies are using data dissemination and related communication strategies to extend the utility of their data to a wide audience. The effectiveness of these

strategies depends on the characteristics of data, target audiences, channel of dissemination, and data formats. A successful approach would need to consider the different needs and sophistication of data requirements of the research communities. This approach should also define the data sharing formats and channels for impactful use in AI research.

Barriers to dissemination are:

Knowledge barriers- A common barrier to the use of publicly held data sets and sharing of privately-held data is a lack of knowledge and awareness about the data available and methods of accessing it. It can be overcome by publishing the availability of data sets and accompanying lineage.

Technical obstacles- Technical obstacles are due to incompatibilities in data formats and preparation methods. To overcome these, data must be transformed, organized, and restructured into a commonly acceptable format using a common methodology. Though technically feasible, the process is elaborate and time-consuming.

Documentation barriers – Differences in data documentation, such as undocumented codes, coding conventions, or missing documentation can result in data that can only be used with difficulty or is completely unusable. This can be overcome by adopting standards for cataloging and documentation.

Conflicting values and obligations – Intellectual property and confidentiality concerns are often cited as reasons for not sharing data. Also, premature release of data might allow some other organization to publish first, and any sharing could deprive the original data collector of longer-term opportunities to mine the data. Enforcing data sharing standards across all NAIRR members will ease data access.

F. Security Requirements and Access Controls

Researchers of accredited and approved universities, which are members of NAIRR should be granted access to the NAIRR platform on a project-by-project basis. They should be required to undertake training in privacy and ethics to become eligible for access, receive certification and meet security requirements. To reduce security risks, the AI subgroups should formally review the research outline and build data sets for analysis using minimum no. of variables and provision minimum compute resources required for the project. Any articles, papers, or materials developed during the research should be examined and cleared by the subgroup prior to publication. The platform should easily allow for the discovery of users, roles, permissions, and access logs.

G. Privacy and Civil Rights and Civil Liberties

Anonymization of data sets is the first measure to ensure the privacy of data subjects. There could also be several other indirect identifiers that could lead to deductive disclosure such as small sample sizes or unusual characteristics of occurrences. Samples taken from specific sub-populations, geographic areas, and linked data sets can be a challenge when trying to protect subject identities. NAIRR can consider data suppression strategies to minimize privacy risks. As data and AI research findings take a variety of formats during the project from system to system, maintaining an audit trail will increase privacy confidence in the NAIRR platform.

H. Sustainment of NAIRR

Partnerships with the corporates can be considered as a source of sustainment for NAIRR. Access to the NAIRR platform can be configured as two types, for government agencies and the commercial sector. Government agencies or member subgroups will fund NAIRR from a portion of research grants. The commercial sector will fund NAIRR for using the platform for data-as-a-service and compute-as-a-service.

I. Parameters for Establishment and Sustainment of NAIRR

Under the AI working group, we propose agency roles for 5 anchors of governance:

Chief Data Officer (CDO): To ensure data quality and promote data governance. Duties include:

- Maintain high availability and quality of data.
- Oversee data governance.
- Create and maintain a data management system to secure data collection and preprocessing of data to promote usability.
- Foster data sharing across the government and industry and establish open data initiatives.
- Identify potential AI research areas and develop implementation plans to acquire or collect data for future research

Chief Ethics Officer (EO): To promote responsible and ethical use of AI on the NAIRR platform. The duties of the EO include:

- Ensure ethical procedures are implemented and adhered to by NAIRR users.
- Examine ongoing research on the platform to enable unbiased and trustworthy AI research.
- Raise ethical considerations for proposed research areas.
- Install an ethical and responsible AI culture interagency as well as across the government and industry.

Chief Privacy & Compliance Officer (CPO): To uphold compliance and privacy requirements. The duties of the CPO include:

- Assess and manage risk related to privacy considerations and compliance.
- Examine ongoing research and future research for information privacy.
- Oversee and monitor implementation of a compliance program within NAIRR.
- Promote adherence to privacy laws within NAIRR.

Chief Information Security Officer (CISO): To protect NAIRR users, assets, and IT systems from security threats. The duties of the CISO include:

- Prevention, investigation, and handling of security threats.
- Establish security practices as well as implement security systems.
- Uphold adherence to governance related to security.
- Responsible for disaster recovery and continuity of operations.

- Assess and handle the risk of cyber threats, data loss, and fraud.

Chief Infrastructure Architect (IA): To maintain computing resources and IT infrastructure. The duties of the IA include:

- Design and implementation of NAIRR enterprise infrastructure.
- Promote and maintain platform infrastructure quality and high availability.
- Improve customer experience of NAIRR IT systems.
- Explore and lead customization and modernization efforts.

2. Prioritizing NAIRR Capabilities and Services

AI being a data-driven initiative, NAIRR should prioritize the creation, collection, preservation, storage, retrieval, and distribution of machine-readable data that can fuel a wide spectrum of AI research.

Niyam's Unified Data for FEMA

- We managed over 480 TB of structured and unstructured data from various sources for combined business intelligence for Risk Management Directorate (RMD).
- We developed ETL workflows for data ingestion and advanced analytics capabilities utilizing modern data science techniques such as clustering, Bayesian, and other statistical models.

Creating a unified data platform as a first step will give an idea of the variety and volume of data workloads, which will help form a strong foundation for architecting a robust compute platform for AI and ML processes. Domain experts should be onboarded in this stage for data sourcing, management and cataloging curated data sets. NAIRR should publish the cataloged metadata for every data set. Regular refresh and archival cycles on existing data sets (and metadata) should be scheduled, to maintain relevance.

3. Reinforcing Ethics and Responsible Research

NAIRR and its components can reinforce principles of ethical and responsible research and development of AI by enforcing the use of the Responsible AI framework for all projects implemented on the platform. Within this framework, all ML models should be comprehensive, explainable, ethical, and efficient.

- Comprehensiveness: The AI model has clearly defined testing and governance criteria
- Explainability: The purpose, rationale, and decision-making process of the AI model can be understood by the average end-user
- Ethical: The AI initiative has processes in place to seek out and eliminate bias in ML models
- Efficient: The AI model can run continually and respond quickly to changes in the operational environment.

At the pre-design stage, the agency subgroups should evaluate research requests on NAIRR by a scrutiny of the problem documentation that includes:

a. Business context of the research problem undertaken
b. Business justification for the algorithm to be developed

c. Model parameters are used for tuning the model to maximize its performance without overfitting or creating a high variance
d. Feature choices (inputs) and definitions (outputs)
e. Any customizations to the algorithm if it was reused
f. Instructions for reproducing the model
g. Examples for training the algorithms and datasets used
h. Examples for making predictions from the algorithm

After successful evaluation, the research projects should be developed within the guidelines of Responsible AI as below:

Shared code repositories: Shared code repositories facilitate efficiency by eliminating rework and reducing the processing overheads of the compute platform. Researchers can reuse existing models/algorithms as stepping stones to further their research on solving newer problems.

Approved model architectures: New model architectures should be approved by the NAIRR working group by evaluating them on explainability and interpretability. This is an important factor to eliminate issues related to fairness, bias, transparency, and accountability.

Sanctioned variables: Datasets made available of research should not contain any personally identifiable information (PII) directly or indirectly. Each dataset should be tagged with its summary statistics indicating the distribution of values, to eliminate bias.

Established bias testing methodologies to uphold fairness, civil rights, gender equity in the models created for AI systems.

Stability standards for active machine learning models to make sure AI programming works as intended and does cause memory leaks and performance bottlenecks for the platform.

Implementing Responsible AI: The most important catalyst for solid governance for implementing Responsible AI is model validation and reproducibility. Model validation is the process of ensuring that the AI model is performant, statistically, sound delivers statistically significant benefits, and meets the definition of “success” put forward by the AI project.

Researchers should group their model by the project. Each attempt to train a model for that project is called a “run,” with all the runs for that project being rolled up into an “experiment.” Putting forth a simple metadata framework centered on the concept of an experiment yields increased visibility and auditability for any AI project. Metadata necessary to reproduce an experiment or a run of an experiment:

a. Type of algorithm used for the development of the model
b. Features and transformations used in the model
c. Data snapshot or identify the data set used
d. Model tuning parameters
e. Model performance metrics
f. Verifiable code location from source control management

g. Training environment setup used for model training

To test the validity of the models, the model should be tested on these behaviors:

- It achieves acceptable statistical performance for a sensible offline metric (accuracy)
- It achieves a statistically significant improvement when compared to control on some online metric or key performance indicator (KPI) (Clicks, conversions, purchases)
- It is statistically sound, there is no data leakage, and the supervised ML problem was framed correctly.
- The performance of the model can be successfully explained based on available features.

4. Existing AI Building Blocks for NAIRR

As the potential of AI research is being recognized, many government organizations are developing programs, resources, and services. NAIRR can establish partnerships to use these as building blocks.

Open data initiatives that provide data sets for download:

- Data.gov – federal government data repository for publicly available federal, state, local, and tribal government information.
- Census.gov – provides United States Census data publicly.
- Data.gov.uk – United Kingdom's published data by their government.
- Open Knowledge Foundation Global Open Data Initiative – provides a repository of the world's open government data publications.
- Federal Reserve FRED Economic Data – provides economic data for research.

Compute and data analytics management resources and AI platforms:

- National Science Foundation's (NSF) Advanced Computing Systems and Services (ACSS) program.
- IBM Watson Studio, RapidMiner, Alteryx, MATLAB, Tableau Server, RStudio, Qlik Sense, Google Cloud AI Platform, Azure Machine Learning Studio.
- Veritone aiWARE Government platform.

Organizations promoting AI through advocacy and innovation efforts:

- Department of Energy (DOE) Artificial Intelligence and Technology Office (AITO).
  - First Five Consortium – a collaboration with industry to promote AI capability, Niyam is a partner of the First Five Consortium.
- United States Department of Agriculture (USDA) AI Institute for Next Generation Food Systems (AIFS) – using AI to drive food systems in the USA.
- USDA National Institute of Food and Agriculture (USDA-NIFA) and NSF.
- Department of Commerce National Artificial Intelligence Advisory Committee (NAIAC).

5. Public-Private Partnerships for NAIRR and Exemplars

NAIRR should provide a separate public and private tier of services, for sustainability. For the public platform, partnerships should be developed with prominent research institutions and industry. Examples of such collaboration include:

a. Partnership between Niyam and Pacific Disaster Center (PDC). We collaborate with the University of Hawaii and other research institutions to research mitigation steps for risks during natural disasters
b. Federal Emergency Management Agency (FEMA) provides publicly available flood products to reduce flood risk under the Risk Mapping, Assessment, and Planning (Risk MAP) program.
c. DOE has partnered with AI-capable companies in the industry, including Niyam, to focus on Humanitarian Assistance and Disaster Response.

The private tier of services of NAIRR can host opportunities like coding competitions to raise funding for sustainment. For example, Kaggle has a business model based on partnerships with private companies to host competitions.

6. Limitations in Democratize Access to AI R&D

Observability, Visibility, and Control: Democratizing AI across a broad spectrum of government agencies, researchers, and citizen scientists will lead to heavy usage and can cause performance degradation, network congestion, data store deadlocks and contention, and other resource allocation complexities. These issues can be addressed by a holistic continuous performance monitoring of the platform and individual models served in real-time. This requires first-class integration with dashboards and visualization software, that can generate usage reports, alerts, and risk mitigation steps.

Intellectual Property rights: Researchers using the NAIRR platform will be on a path to develop powerful AI tools, with novel ideas. The perceived benefits of democratization may not be achieved without decisions about who owns the intellectual property rights. NAIRR's working group should have a strong legal framework that addresses rights and responsibilities around patents, copyrights, and trade secrets.

Cost visibility and management: The ownership and administration structure of NAIRR as explained in Question 1-B, detailed financial reporting for federal stakeholders will be required. Costs associated with the development and maintenance of NAIRR are likely to fluctuate based on a no. of different factors. Therefore, building visibility of costs using responsible accounting strategies is important. Chargeback and show-back strategies can be used to shift responsibility to participating agency subgroups or members and encourage them to become more aware of costs. They will also be helpful in budgeting, planning, and forecasting.

Compliance attestation (SOC 2, HIPAA): Any compliance failure could put the NAIRR platform at massive PR and financial risks. The agency subgroup should ensure that all AI research within their jurisdiction is compliance attestable and there are no violations.