Agenda

11:00-11:10	**Welcome and Administrative Remarks,** Manish Parashar & Tess deBlanc-Knowles
11:10-12:10	**Readout and Discussion of Proposed Plans: Security, Access Controls, and User Portal,** Elham Tabassi
12:10-1:10	**Readout and Discussion of Proposed Plans: Resource Allocation, Usage Policies, and Evaluation,** Fei-Fei Li
1:10-1:40	**Break**
1:40-2:40	**Readout and Discussion of Proposed Plans: Environmental Sustainability, International Collaboration, and Legal Considerations,** Mark Dean
2:40-3:40	**Panel: Associated Federal Efforts for Provision of Data and Computing Resources** Moderator**:** Erwin Gianchandani • Vipin Arora, *Deputy Director, National Center for Science and Engineering Statistics, National Science Foundation* • Shelly Martinez, *Senior Statistician, Office of Management and Budget* • Manish Parashar, *Office Director, Office of Advanced Cyberinfrastructure, National Science Foundation* • Kamie Roberts, *Director, National Coordination Office for Networking and Information Technology Research and Development* • Jerry Sheehan, *Deputy Director, National Library of Medicine, National Institutes of Health*
3:40-4:10	**Break**
4:10-4:50	**Open Issues and Implementation Plan Development Process,** Tess deBlanc-Knowles
4:50-4:58	**Questions from Public,** Manish Parashar
4:55-5:00	**Closing Remarks,** Manish Parashar

NAIRR WG 4
Security/access controls and user portal

Task Force Members:

Andrew Moore, Mike Norman, Elham Tabassi

Charge to the WG

1. How should the NAIRR management entity construct and manage the user portal? What level of user support should be provided through it?

2. How will the NAIRR security policies be developed and built into the cyberinfrastructure and governance processes?

3. How will the management entity manage user access and permissions?
   - How will it adjudicate instances of users violating use agreements?

4. For resource providers/partners, will the entire center have to abide by NAIRR security policies and standards even if only a subset of the Center resources is part of the NAIRR?

5. Beyond FEDRAMP (which is specific to cloud infrastructure) what security standards will the NAIRR need to develop for other kinds of resources it will include?

6. Should egress of data from the NAIRR infrastructure be restricted due to security concerns?

7. Will the NAIRR require an Authorization to Operate (ATO) if connecting into existing Federal resources?
   - Would the NAIRR manage its own ATO process for non-governmental resources it will include?

1. How should the NAIRR management entity construct and manage the user portal? What level of user support should be provided through it?

Open science

- User portals are widely used for resource discovery (compute, data, applications), resource selection, job submission and monitoring, data management, account creation/management, and usage reporting.

- They also serve as a clearing house for training materials and possibly an entry point to user forums (ACCESS, CloudBank examples).

Sand boxes

operated within a secure environment but with relaxed rules to enable experimentation

Protected/restricted data

- User access and the application suite is centrally controlled and monitored for compliance with the relevant controls (e.g., HIPAA, FISMA).

- Access to a secure enclave via a virtual desktop which presents a bespoke application suite customized for each user community.

- From that virtual desktop, one *cannot* launch arbitrary applications or services in the cloud or on-prem, as these applications must be integrated into the adopted security architecture.

1. How should the NAIRR management entity construct and manage the user portal? What level of user support should be provided through it?

NAIRR-Open

- The NAIRR OE should bid out the design, construction, and maintenance of an integrated user portal interfacing to all resources (compute, data, AI testbeds) that are part of the NAIRR-Open that includes
    - built-in help functions to assist in the use of the portal itself.
    - maintain an up-to-date catalog of resource provider user documentation and training materials . Finally, the portal should have a help desk ticketing system integrated into it.

Sand boxes

operated within a secure environment but with relaxed rules to enable experimentation

NAIRR-Secure

- The NAIRR OE should bid out the design, construction, and operation of NAIRR-Secure, a managed secure enclave in multiple public clouds (to avoid vendor lock-in) with application suites tailored for AI/ML workloads in the targeted research domains
    - User support includes requirements gathering, solution architecting, implementation of the solution in the cloud, patching/updating software, and ongoing user support (add/delete users), manage data access controls, managed by a team of 6-8 people.
    - public clouds which have implemented fine-grain access controls would be preferred platforms for NAIRR-Secure compared to on-prem solutions.

2. How will the NAIRR security policies be developed and built into the cyberinfrastructure and governance processes?

The working group recommends the NAIRR operating entity establish two separate zones with separate identity management, security policies, and access management controls.

NAIRR-Open

- Adopt the best practices developed over two decades in the open science community, drawing from experiences and approaches used by XSEDE/ACCESS, the Open Science Grid, and the National Research Platform.

Sand boxes

NAIRR-Secure

- One or more secure enclaves adhering to a common set of security controls such as FedRAMP, FISMA moderate, or CMMC level 3. Example: Sherlock Secure Cloud operated by the San Diego Supercomputer Center.

3. How will the management entity manage user access and permissions? How will it adjudicate instances of users violating use agreements?

NAIRR-Open

- Using single-sign-on (SSO) authentication and a resource allocation mechanism.

  - In XSEDE allocations were tied to specific resources. The ACCESS program is implementing portable credits which can be allocated independent of the resources and used on any resource. The CloudBank project uses real US dollars as portable credits to purchase public cloud resources.

  - A possible merger of these approaches into a single NAIRR credit currency that would be allocated to approved projects -- based on identity and allocation -- and that could be spent anywhere within the NAIRR.

NAIRR-Secure

- A separate identity and access management system than NAIRR-Open

- The NAIRR operating entity will need to either build or contract out a suite of managed services that help them comply with the adopted security policies and access controls (e.g., FedRAMP).

  - services include identity and access management, vulnerability management, configuration management, patch management, anti-virus/malware protection, system monitoring and hardening, log management, secure file transfers, IDS/IPS, authentication, domain controllers, firewall and VPN management, backup archiving, and certificate authority.

3. a. How will it adjudicate instances of users violating use agreements?

– Compliance with acceptable use policies of the NAIRR requires both user education and user activity monitoring.

– Many monitoring tools are built into public cloud consoles. These notifications and alerts can be captured by a multi-cloud user portal such as the CloudBank User Portal and forwarded to both user and NAIRR administrator.

– Non-public cloud RPs will be required to deploy a standardized set of security monitors reporting to a centralized database managed by NAIRR OE or its subcontractor.

– Violations of acceptable use agreements will be handled with a range of measures, the severity of which will depend on the frequency and seriousness of the infraction. These range from friendly warnings at the low end to account suspension at the high end.

4. For resource providers/partners, will the entire center have to abide by NAIRR security policies and standards even if only a subset of the Center resources is part of the NAIRR?

– No. Only the NAIRR-affiliated resource and its operators will have to abide by the NAIRR security policies and standards.

– A center may provide one or more open science resources, one or more secure data resources, or a combination of resource types.

– Each resource is governed by the relevant policies and standards depending on its type.

5. Beyond FEDRAMP (which is specific to cloud infrastructure) what security standards will the NAIRR need to develop for other kinds of resources it will include?

– Controlled Unclassified Information (CUI)

– The database of Genotypes and Phenotypes (dbGaP)

– Health Insurance Portability and Accountability (HIPAA)

– Federal Information Security Management (FISMA)

Recommendation: NAIRR-Secure must be FISMA compliant, and have the ability to support related security requirements arising from HIPAA and CUI regulations.

6. Should egress of data from the NAIRR infrastructure be restricted due to security concerns?

NAIRR-Open

- Open, available for egress for external uses.

NAIRR-Secure

- Protected data shall stay at the secure enclave which houses it, or securely transferred to other secure enclaves at the same or more severe security level and subject to data use agreements.

- Related security controls can be implemented as a policy or enforced by exfiltration detection. The working group recommends explicit exfiltration prevention mechanisms.

- Enclave to enclave data transfers would be done using APIs that respect the access controls established by the data administrators.

7. Will the NAIRR require an Authorization to Operate (ATO) if connecting into existing Federal resources?
   Would the NAIRR manage its own ATO process for non-governmental resources it will include?

– For governmental resources: the working group assumes NSDS will handle all administrative data originating from federal agencies and need an ATO

– NAIRR-Secure will focus on research data including but not limited to data owned by federal agencies (e.g., NASA, NOAA, USGS). To the extent that the data owners (the agencies) require an ATO, then it will be the responsibility of NAIRR-Secure to obtain it.

– For non-governmental resources: NAIRR-Secure will manage its own ATO process.

NAIRR and NSDS

- The CHIPS+ Act includes a provision to establish the National Secure Data Service (NSDS) demonstration project at the National Science Foundation. Authorized in legislation, as a pilot project, with the ability to scale, the NSDS will allow some federal data to be combined temporarily in a privacy-protected environment.

- WG recommends coordinating and collaboration with the NSDS, and others as appropriate, on specifying security and user access controls to restricted (confidential) government and third-party data. To NAIRR OE, NSDS could look like another specialized resource provider.

Discussion

Governance, part II

resource allocation, including initial access for those without prior federal funding, usage policies, and metrics/evaluation processes

Members: Fei-Fei Li, Dan Stanzione, Julia Lane, Manish Parashar

Reminder - NAIRR goals; How should our governance policies reflect them?

- Spur innovation
- Increase diversity of talent
- Improve capacity
- Advance trustworthy AI

15

Part 1: Resource Allocation Charge Questions

1. What would be the process for allocating access to resources that are part of the NAIRR? For example, how would the management entity process requests for access to resources from potential users with funding from federal agencies? From potential users without any such funding? (Process)
2. Will researchers seeking funding from Federal entities need to add NAIRR fees to their proposals? (Fees)
3. What role would the NAIRR management entity have in operationally monitoring the resources that it offers? (Monitoring)

16

Resource Allocation: Findings

1. There are multiple excellent models for resource allocation from other federal research infrastructure on which the NAIRR can draw, including telescopes, light sources, and computing facilities.
2. Computing and Data facilities are the closest analogs for the NAIRR to emulate.
3. There are certain unique characteristics of AI that will need slightly different handling than existing processes.

Resource Allocation: Process

What would be the process for allocating access to resources that are part of the NAIRR?

RECOMMENDATION 1:

There should be multiple processes, based on size and scope of the request:

One track should be driven by participating agencies.   Credits awarded directly in grant programs, or directly to their awardees within the projected budget.

A second track should be driven by the management entity:
- Startup or educational requests: reviewed by staff at management entity, turnaround less than 1 month, caps at say, $1,000 worth of compute time/storage.
- Research: Peer-reviewed, Quarterly, panel organized by management entity, up to a certain size cap.
- Purchases
  - Allocation over the research cap.
  - Any other entities that want access (industry, etc.).

18

Resource Allocation
End User View

How do I get access to NAIRR?

Apply through your funding Agency

Apply directly to NAIRR
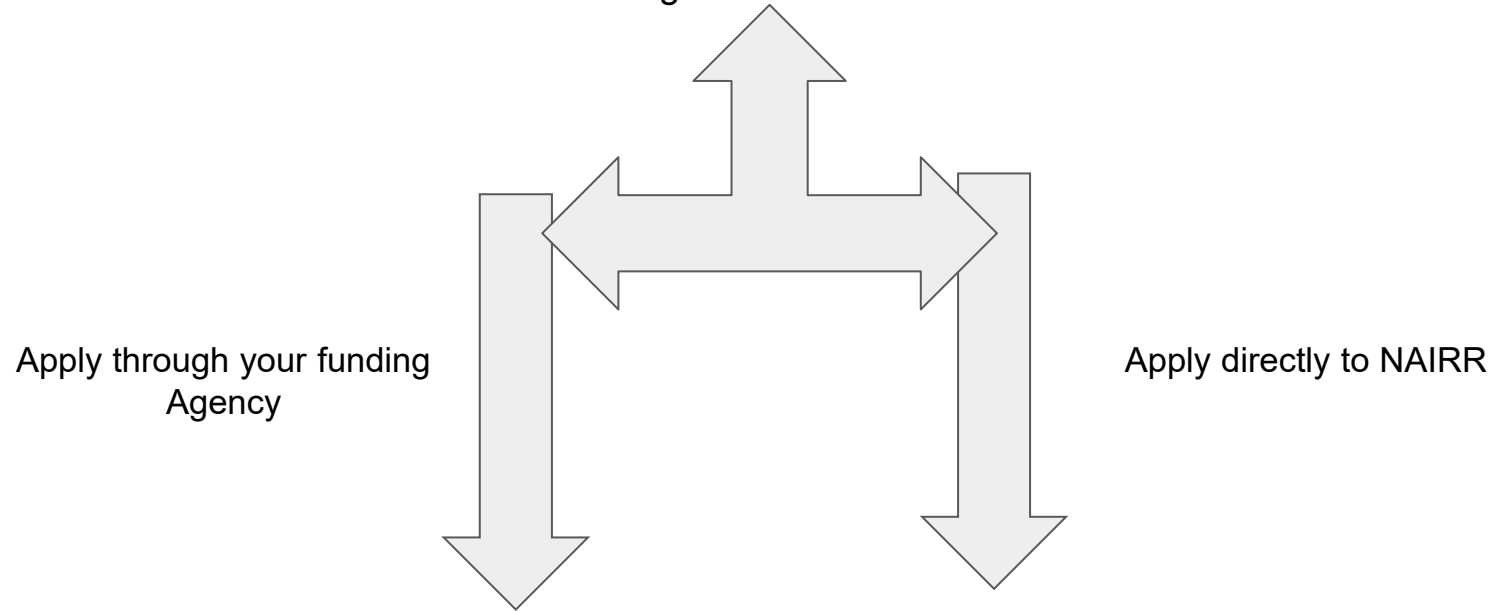
Talk to your program officer for agency-specific rules

See the Application Guide at http://NAIRR.gov/Apply

19

Resource Allocation: Process

Recommendation 2: Split allocation between tracks:

- Agency contributions should be "taxed" (or central funding is available) to create the "Peer Review" track.
- Approximately 70-30(??) split in "base" funding between Peer-Review and Agency tracks. ("Agile re-balancing").
- Additional funding provided by agencies goes 100% to time for that agency.
- *Need to create space for the broader access mission beyond agency-picked awardees.*

20

Resource Allocation: Process

Recommendation 2 (continued)

- Allocations will happen in "Credits"
  - Base rates will be **Compute Time** or **Data Storage**
  - Some things, like downloading data/models from repository, would not require any credits.
- Credits could also be used for more advanced or derived quantities:
  - Access to support time could be provided for a rate in credits (beyond basic tickets/user support).
  - More advanced workflows (e.g., interpret documents through a natural language model) could have a rate derived from the consumption of compute, storage, and any "value add".
- A single unit greatly simplifies the allocation process!
  - But can make it complex to estimate. . .
- Some flexibility will be needed, as AI workloads are notoriously difficult to estimate
  - Typically, this can be dealt with by policies for supplements, advances, transfers, etc.
- Care must be taken to match awarded credits with available resources!

Resource Allocation: Process

Recommendation 2 (ctd)

Tracks should be structured with different Criteria/Process for Selection

1. "Peer Review" Track
- Basic principle: As the size of the request gets larger, the bar for review/access goes up.
- At the startup/student/classroom level:
  - Students/Staff/Faculty at any US institution of higher education are eligible within the constraints of US export control law
  - Students must have a responsible faculty member
  - Application would be a simple form, validation of enrollment and eligibility, description of project.
- At the research level:
  - Proposal describing the work.
  - Possible outcomes include full acceptance, full rejection, cuts in amount awarded, or re-directing to different resources within NAIRR.

Resource Allocation: Process

2. "Agency Track"

- Agencies would have broad latitude in making assignments.
- We *suggest* they award NAIRR access with the funding awards, but other in-house processes or possible.
- They could turn the process back over to the NAIRR allocations team to run through the Peer-review track.
- NAIRR Management's responsibilities are to keep the agencies within their allocation caps.

Resource Allocation: Charging

Will researchers seeking funding from Federal entities need to add NAIRR fees to their proposals?

FINDING:

- Universally applying fees would run counter to the NAIRR's mission to democratize access to AI Resources

RECOMMENDATION: Different cases

- No fees for agency-driven awards
- "Typical" research allocation would not require fees, nor would startups or allocations.
- Fees required for guaranteed time or large work
- Private sector partners would need $$$ (unless eligible small business or similar circumstance).

Administration

- For grants admin simplification, fees shouldn't be cash – they should be "credits" on NAIRR from a funding-agency managed budget (a credit can be a dollar, or an hour on a typical computer, a quantity of data storage, support time, etc.).

Resource Allocation: Monitoring

3. What role would the NAIRR management entity have in operationally monitoring the resources that it offers?

Recommendation:
Monitoring should be shared responsibility of OE/Resource Providers
- The governing principle should be:
  - An RP should be responsible for the user experience, and attracting users to, their services Make sure usage is appropriate, and within terms of use (e.g. no bitcoin mining). RPs would be responsible for uptime, security, maintenance, upgrades, problems supporting users distinct to their services, etc.
  . The OE should in turn monitor the RPs for effectiveness, and attract/train users to NAIRR as a whole, independent of the specific resources available.
    - Assure each allocation is having the appropriate impact before renewal, collecting measures of impact, etc.
    - Evaluating user experience and other metrics on RPs, to evaluate eventual RP renewals.
    - General and front line user support.
    - Collect data for a central dashboard showing operational status of resources.

25

Resource Allocation: Additional Notes

- A suggested/strawman resource allocation policy is included as part of the report.
  - Covers process, eligibility, criteria, etc.
  - Also suggestions for participating agencies.

- A bibliography of other Resource Allocation models is also provided with the report.

26

Part 2: Usage policies
Charge Questions

1. What kind of use policies should the NAIRR management entity develop? How should they be enforced?
2. For users who are accessing NAIRR resources directly through the NAIRR and not via prior Federal funding, should the NAIRR intellectual property policies mirror relevant Federal guidelines? What should they be?

27

Usage Policy

- FINDINGS:
  - As with Resource Allocations, there are many excellent models to draw from (shown in the report and including international activities).
  - Several different policies to consider; e.g. different requirements for users of NAIRR vs. those publishing data on NAIRR.
  - The Advisory Committee for Data on Evidence Building (ACDEB) has developed usage policies in parallel for confidential, CIPSEA protected data
  - The CHIPS plus act established a National Secure Data Service (NSDS) to provide access to confidential data both via secure facilities and with privacy preserving technologies
- RECOMMENDATIONS:
  - The management entity should develop a general user policy, as well as a set of data publishing and use policies.
  - The NAIRR's approach should be informed by the activities of the NSDS

28

Recommendation for Compute/General Use Policy
(see report/appendices for full policy recommendations)

- ● All users
  - ○ Lawful use of NAIRR, respect of IP rights, confidentiality, compliance (e.g. HIPAA), etc.
  - ○ Proper conducting, including access, etc.
  - ○ Security
  - ○ Account responsibility
  - ○ Crediting NAIRR, publication citations, etc.

- ● PIs and Project leaders
  - ○ Represent proper institutional usage
  - ○ Responsible selection of NAIRR users and usage allocation
  - ○ Inform NAIRR RPs and Mangement issues when staff/students leave the institution.

29

Data Use Policies for Data *Consumers*

Recommendations

- Use $$ as a currency instead of time (e.g. $ instead of 10-hours)
- Incentives put in place to promote collaboration and shared use
  - E.g. Can earn shadow $$ by contributing to curation and code (like Stack Overflow)
- Data set citations - example:
  - Buesseler, K. (2006) "VERTIGO Cruise Event Log." National AI Research Resource (NAIRR). Dataset version: 5 September 2023. [hyper link] [access date]
- No warranty:
  - All materials on NAIRR are made available to NAIRR and in turn to you as-is. There is no warranty, expressed or implied, to these materials, their title, accuracy, non-infringement of theird party rights, or fitness for any particular purpose, or the performance or results you may obtain from their use.
- Limitation of liability:
  - Under no circumstances shall NAIRR be liable for any direct, incidental, special, consequential, indirect, or punitive damages that result from the use of, or the inability to use, either this Website or the materials available via this Website. If you are dissatisfied with any Website feature, content or terms of use, your sole and exclusive remedy is to discontinue use.
- Requests for Trademarks Use:
  - NAIRR is registered trademarks and may not be used for any purpose without the express written permission of the [NAIRR office]
- Access to confidential data
  - Governed by legal requirements of data provider
  - National Secure Data Services is developing protocols and Standard Application Process

30

Data submission policies for Data *Producers* (Part 1)

Recommendation

All datasets must, to the extent possible, include the following

1. DOI and citation information
2. Metadata about what is in the dataset that provides the following information
   a. **Data dictionary using standardised** terminology
   b. Source data and any applicable code used to create the submitted dataset
   c. Provenance and version
3. Access rules
4. Usage information
5. Pay attention to FAIR principles (Findable, Accessible, Interoperable, Reusable)

31

# Data submission policies for Data *Producers* (Part II)

Recommendations

- By using NAIRR, you're agreeing to be bound by the Terms and Conditions
- All datasets and associated information are licensed under a [Placeholder: "creative commons attribution 4.0 International License."] Per the CC By 4 license, it is understood that any use of the data set will properly acknowledge the PI.
- NAIRR recommends that you contact the original PI should you require additional information about the data beyond what is available on this site. In the even that the data originator is not available, please contact [NAIRR office] for guidance.
- By submitting data, metadata and any other contents to NAIRR, submitteres warrant that they own the rights to the contents and are authorized to do so under original copyright. Items in NAIRR are protected under original copyright with all right reserved, where applicable (e.g., documents such as related publications, etc.).
- By submitting data and information, the submitter grants NAIRR the rights needed to copy, store, redistribute and share data, metadata and any other content. By validating and making their data publicly available, submitters gran NAIRR and any other users the right to reuse their data according to the term.

32

Data usage tracking policies for the NAIRR OE

Information on data is **constantly** being generated. Depending on the phase in the data lifeline, the type of information **differs**

Recommendation: NAIRR OE should constantly track and report access and use
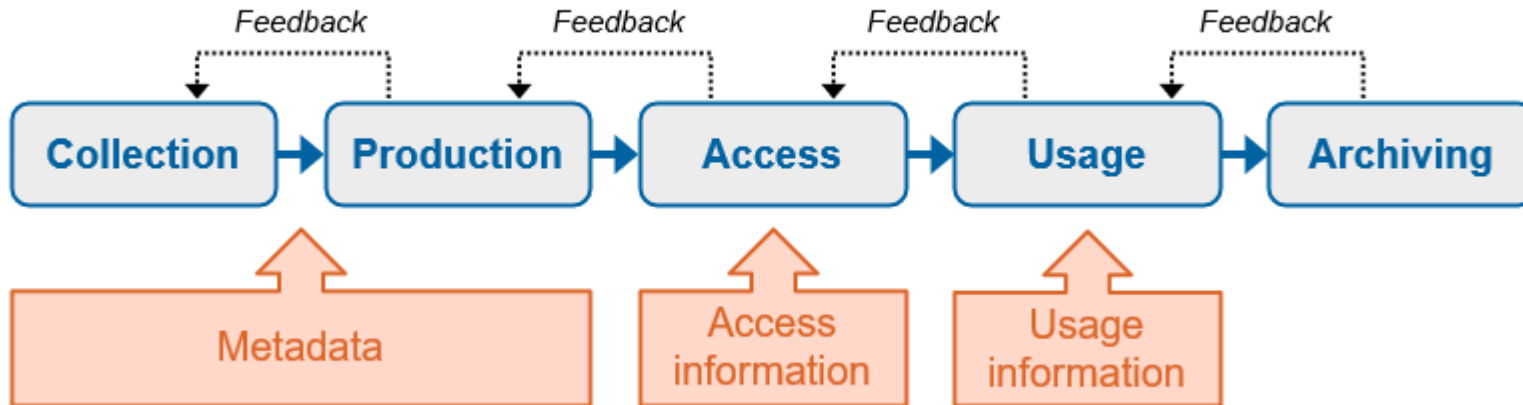
Source: New Data need New Processes: Stefan Bender, Jannick Blaschke and Christian Hirsch (Deutsche Bundesbank)

Data usage Guidelines for papers published that used NAIRR Data

Recommendation

The policy strongly encourages authors reporting experiments on a new dataset to conform to the following rules:

a. A link to the dataset is provided in the paper (at submission time in anonymized format).
b. The dataset has a persistent identifier such as Digital Object Identifier or Compact Identifier.
c. The dataset adheres to NAIRR metadata standards.
d. The license and/or any data access restrictions are described in the paper.
e. The paper includes a convincing justification of the special nature of the dataset that makes it impossible to conform to these suggestions.

34

3. Metrics/evaluation

1. What specific indicators should the NAIRR management entity measure and track? (using prior list as a starting point.)
2. How often should the NAIRR be externally evaluated? Is this constant or should it ramp up as the resource matures?

35

Initial Findings

36

Findings (ctd) Evaluations span multiple dimensions and time periods

·   "Inputs. Resources at the disposal of the project, including staff and budget.

·   Activities. Actions taken or work performed to convert inputs into outputs.

·   Outputs. The tangible goods and services that the project activities produce; these are directly under the control of the implementing agency.

·   Outcomes. Results likely to be achieved once the beneficiary population uses the project outputs; these are usually achieved in the short to medium term and are usually not directly under the control of the implementing agency.

·   Final outcomes. The final results achieved indicating whether project goals were met. Typically, final outcomes can be influenced by multiple factors and are achieved over a longer period of time"
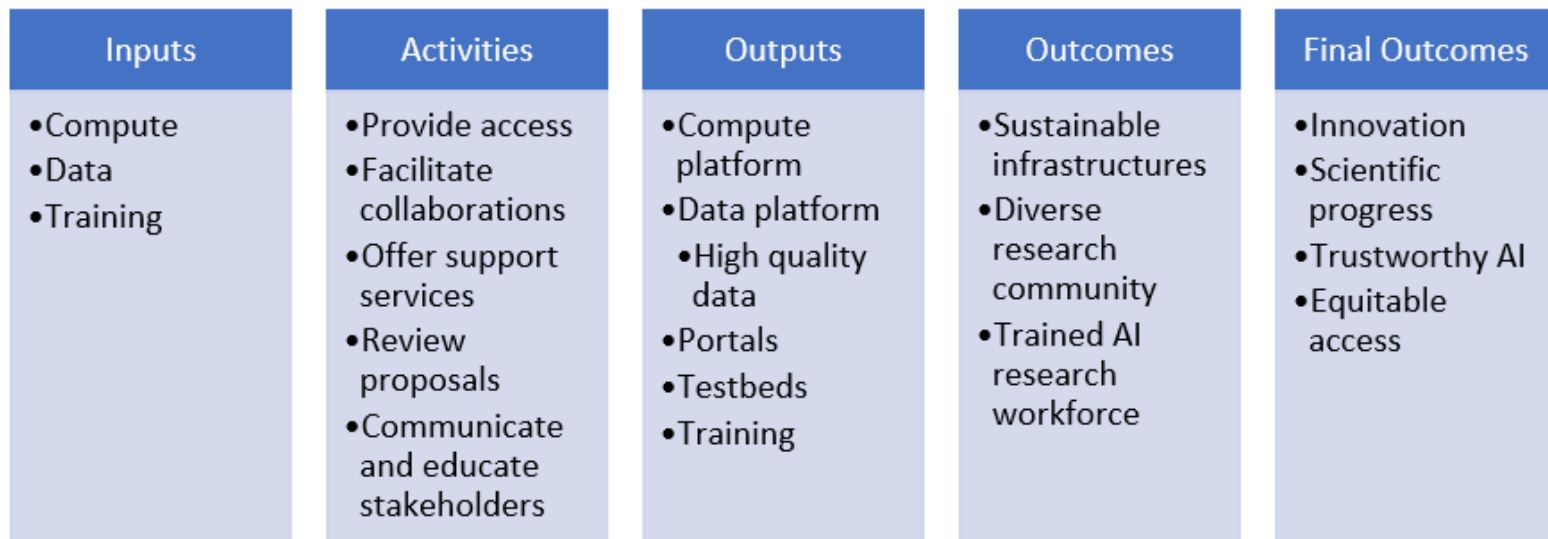
Initial Recommendations

The NAIRR should adopt a standard evaluation framework

The NAIRR should trace standard measures, such as publications and patents, but also new measures, such as student/postdoc placement and firm startups, productivity and growth. This will require establishing an evaluation dataset

The NAIRR OE should develop and publish appropriate KPIs based on that framework (see appendix)

The initial evaluations (years 1-3) should be operational implementation (see appendix)

The three-year evaluation should evaluate progress towards Congressional goals

Inputs	Activities	Outputs	Outcomes	Final Outcomes
•Compute •Data •Training	•Provide access •Facilitate collaborations •Offer support services •Review proposals •Communicate and educate stakeholders	•Compute platform •Data platform •High quality data •Portals •Testbeds •Training	•Sustainable infrastructures •Diverse research community •Trained AI research workforce	•Innovation •Scientific progress •Trustworthy AI •Equitable access

Detailed Finding 1: There are important challenges in developing KPIs

Identifying operational definition of AI

Note that AI is neither an industry nor a scientific field, but some work has been done
(CSET; National Network for Critical Technology Assessment; industry definitions ; OECD;OSTP)
Appropriate evaluation datasets exist and can be expanded
Grant information (relevant agencies); Publications and patents (CSET); Workforce measures (IRIS at University of Michigan); Economic impact (IRIS; state data)
Identifying operational definition of a counterfactual/related field
Some work has been done; draw on National Network for Critical Technology
Assessment
Appropriate datasets exist and can be expanded
Grant information (relevant agencies); Publications and patents (CSET); Workforce measures (IRIS at University of Michigan); Economic impact (IRIS; state data)

40

Detailed Recommendation 1

KPIs should be developed that mirror the evaluation methodology (see appendix for examples)

The performance metrics for the NAIRR resource providers and OE should be a limited set of high-level metrics that the PMO can initially use to monitor and evaluate the operational effectiveness of the research resources managed and the services provided by OE to the user community. Those metrics should be clearly stated and published

41

Detailed Finding 2: There are a number of best practices that can be followed

- The evaluations should be designed and implemented by an external entity
- A substantial budget should be dedicated to designing, gathering the data for and implementing the evaluation (about 2% of cost of intervention)
- KPIs should be SMART (Specific, Measurable, Attributable, Realistic and Targeted)
- The data underpinning the KPIs should be transparent, accessible, and covers the appropriate population in a timely and reproducible manner

42

Detailed Recommendation 2

NAIRR should immediately establish initial benchmarks using appropriate datasets (see appendix)

NAIRR should publish levels and trends in each KPI relative to initial benchmark and counterfactual on an annual basis (see appendix)

Detailed Finding 3: There are different levels of evaluation required

Entities of interest

- NAIRR itself
- NAIRR OE
- Each resource provider
- Each program

Type of evaluation

- Operational efficiency
- Progress towards goals

44

Appendix

Metrics:

· "Inputs. Resources at the disposal of the project, including staff and budget (first year).

· Activities. Actions taken or work performed to convert inputs into outputs.

· Outputs. The tangible goods and services that the project activities produce; these are directly under the control of the implementing agency.

· Outcomes. Results likely to be achieved once the beneficiary population uses the project outputs; these are usually achieved in the short to medium term and are usually not directly under the control of the implementing agency.

· Final outcomes. The final results achieved indicating whether project goals were met. Typically, final outcomes can be influenced by multiple factors and are achieved over a longer period of time"

Inputs

- How much funding has been allocated to each area (compute, data, training, testbeds)?

- How many program staff have been allocated from each agency?

- How much support has been provided by philanthropic organizations?

- How much support has been provided by industry?

47

Metrics:

· "Inputs. Resources at the disposal of the project, including staff and budget.

· Activities. Actions taken or work performed to convert inputs into outputs (Years 1-3)

· Outputs. The tangible goods and services that the project activities produce; these are directly under the control of the implementing agency.

· Outcomes. Results likely to be achieved once the beneficiary population uses the project outputs; these are usually achieved in the short to medium term and are usually not directly under the control of the implementing agency.

· Final outcomes. The final results achieved indicating whether project goals were met. Typically, final outcomes can be influenced by multiple factors and are achieved over a longer period of time"

Innovation: Operational

- Strategic plans for each resource developed and publicly published
- Relevant portals developed for each core service area (compute, data, training, testbeds)
- Advisory committees established with appropriate charters
- Guidelines for which individuals or groups are eligible to use the resource, how to allocate their resources among interested users, and how the users request and gain access in support of their research published and posted
- At least five RFPs drafted with core agencies that provide access to NAIRR core services
- At least five cross agency competitions launched
- Funding line established to support investigator-initiated workshops to scope emerging areas of science and technology
- Established definition of AI as field and identified set of counterfactual fields

49

Innovation: Baseline Measurement

- Number of AI publications and patents relative to counterfactual
- Number of "groundbreaking" AI publications and patents relative to counterfactual
- Number of AI "research involved" individuals (defined as individuals paid on AI grants) relative to counteractual
- Earnings of AI "research involved" individuals when placed relative to counterfactual
- Number of AI intensive firms relative to counterfactual
- Employment in AI intensive firms relative to counterfactual
- Mean and median earnings of workers in AI intensive firms relative to counterfactual

50

Diversity: Operational

- Diversity strategic plan developed and publicly published

- Access barriers identified and a plan established to lower barriers

- Effective support system put in place to promote access for a diverse community

- Communication plan established to promote access for a diverse community

51

Diversity: Baseline Measurement

- Number of AI publications and patents published by underrepresented minorities
- Number of "groundbreaking" AI publications and patents published by underrepresented minorities
- Number of minority AI "research involved" individuals (defined as individuals paid on AI grants)
- Earnings of minority AI "research involved" individuals when placed
- Number of publications and patents published by underrepresented minorities in counterfactual fields
- Number of "groundbreaking" publications and patents published by underrepresented minorities in counterfactual fields
- Number of minority counterfactual "research involved" individuals (defined as individuals paid on AI grants)
- Earnings of minority counterfactual "research involved" individuals when placed

Capacity: Operational

- - Training strategic plan developed and publicly published
- - Training program staff hired
- - Training provider(s) identified
- - Training curriculum developed
- - Training program established and delivered

Capacity: measurement

- Number of AI training programs

- Number of participants in AI training programs

- Number of counterfactual training programs

• Number of participants in counterfactual training programs

54

Ethics: Operational

- Ethics Officer appointed
- Operational plan for PCRCL issues developed and publicly posted
- Ethics consulting service established
- Scientific Integrity Office established
- Operational plan for scientific integrity developed and publicly posted
- Project registry of all funded projects established and publicly posted

Metrics:

· "Inputs. Resources at the disposal of the project, including staff and budget.

· Activities. Actions taken or work performed to convert inputs into outputs.

· Outputs. The tangible goods and services that the project activities produce; these are directly under the control of the implementing agency (year 3)

· Outcomes. Results likely to be achieved once the beneficiary population uses the project outputs; these are usually achieved in the short to medium term and are usually not directly under the control of the implementing agency.

· Final outcomes. The final results achieved indicating whether project goals were met. Typically, final outcomes can be influenced by multiple factors and are achieved over a longer period of time"

56

Innovation:compute

Has the NAIRR's computer provider(s) provided greater access to computing resources?

Metrics

- Availability, speed, size

- Percentage of resource utilization (% delivered as compared to planned annual hours of availability)

- Number and diversity of individuals accessing the compute

- Value: the number of sponsoring agencies

Cost/benefit

- What was the cost of the provision?

- What was the benefit/cost relative to an appropriate counterfactual?

Innovation: Data
Has the NAIRR's data provider(s) provided greater access to computing resources?

Metrics

- How many existing and new datasets have been accessed and reused in search and discovery platform
- Data coverage, as measured by expansion of types of testbed data to new populations, domains, etc.
- Data variety
- Access efficiency, as measured by the efficiency of workflow for users, including the distribution of task turnaround times (securing data agreements for new data sets, data ingest, data cleaning, granting permissions for existing datasets, coding, compute, analysis of results, and disclosure review)
- Diversity, as measured by number and diversity of individuals accessing the data
- Value, as ,easured by the number of sponsoring agencies

Cost/benefit

- What was the cost of the provision?
- What was the benefit/cost relative to an appropriate counterfactual?

58

Innovation: Measurement

- Growth in Number of AI publications and patents relative to counterfactual

- Growth in Number of "groundbreaking" AI publications and patents relative to counterfactual

- Growth in Number of AI "research involved" individuals (defined as individuals paid on AI grants) relative to counterfactual

- Growth in Earnings of AI "research involved" individuals when placed relative to counterfactual

- Growth in Number of AI intensive firms relative to counterfactual

- Growth in Employment in AI intensive firms relative to counterfactual

- Growth Mean and median earnings of workers in AI intensive firms relative to counterfactual

Diversity

- Growth in number of AI publications and patents published by underrepresented minorities relative to counterfactual

- Growth in number of "groundbreaking" AI publications and patents published by underrepresented minorities relative to counterfactual

- Growth in number of minority AI "research involved" individuals (defined as individuals paid on AI grants)

- Growth in earnings of minority AI "research involved" individuals when placed

- Growth in number of minority counterfactual "research involved" individuals (defined as individuals paid on AI grants)

- Growth in earnings of minority counterfactual "research involved" individuals when placed

Cost/benefit

- What is the cost of the program?

- What i the benefit/cost relative to an appropriate counterfactual?

60

Capacity

- Growth in number of AI training programs throughout the country (relative to the appropriate counterfactual)

- Growth in number of individuals trained in AI throughout the country (relative to the appropriate counterfactual)

- Growth in number of minority individuals trained in AI throughout the country (relative to the appropriate counterfactual)

Cost/benefit

- What is the cost of the program?

- What i the benefit/cost relative to an appropriate counterfactual?

61

General comments

- *Resource providers*: Should be evaluated for operational efficiency on the following high-level performance metrics: Customer support, Queue times, Consultant response time, Computational time, Computational services, Allocated Time limits, and quality and completeness of resource documentation.

- *Individual programs*: The characteristics of provisioned resources may vary by user community. Additional specific metrics for each major category of provisioned resources should complement the high-level performance metrics above.

62

Supplemental materials: Compute

Compute user policy - All users

- Only use NAIRR-allocated resources to perform work and transmit/store data consistent with the stated allocation request goals and conditions of use as defined by your approved NAIRR project, this NAIRR use policy and any local service provider policies
- NAIRR allocations are awarded for open research intended for publication, but you will respect IP rights and observe confidentiality agreements.
- You will not use NAIRR resources for unauthorized financial gain or any unlaweful purpose, nor attempt to breach or circumvent any NAIRR administrative or security controls. You will comply with all applicable laws, working with your home institution and the specific NAIRR service providers utilized to determine what constraints may be placed on you by any relevant regulations such as export control law or HIPAA.
- You will protect your access credentials (e.g., private keys, tokens & passwords) which are issued for your sole use. This includes:
  - Using a unique password for your NAIRR User Portal account
  - Only entering your NAIRR password into NAIRR sites
  - Not knowingly allowing any other person to impersonate your NAIRR user identity

- You will immediately report any known or suspected security breach or misuse of NAIRR access credentials
- Use of resources and services through NAIRR is at your own risk. There are no guarantees that resources and services will be available, that they will suit every purpose, or that data will never be lost or corrupted. Users are responsible for backing up critical data.
- Logged information, including information provided by you for registration purposes, is used for administrative, operational, accounting, monitoring and security purposes. This information may be disclosed, via secured mechanisms, only for the same purposes and only as far as necessary to other organizations cooperating with NAIRR.
- You will acknowledge use of NAIRR, supported by [blah blah] and award number [blah blah], in manuscripts submitted for publication. In addition, you are expected to acknowledge use of the specific resource(s) utilized (see blah blah)
- Violations of NAIRR policies and/or service provider policies can result in loss of access to resources. Activities in violation of any laws may be reported to the proper authorities for investigation and prosecution.

64

Compute user policy - PIs and Project Leaders

- Access-granting organizations, your allocation's Principal Investigator (PI), and service providers are entitled to regulate, suspend or terminate your access, and you will immediately comply with their instructions.
- PIs are responsible for properly vetting users on their allocations and by doing so they are attesting that the NAIRR User Portal username belongs to the intended person. PI's will also ensure that users who have access to NAIRR-allocated resources on the PI's NAIRR allocation follow this AUP.

65

Supplemental materials: Data

Stefan Bender, Deutsche Bundesbank
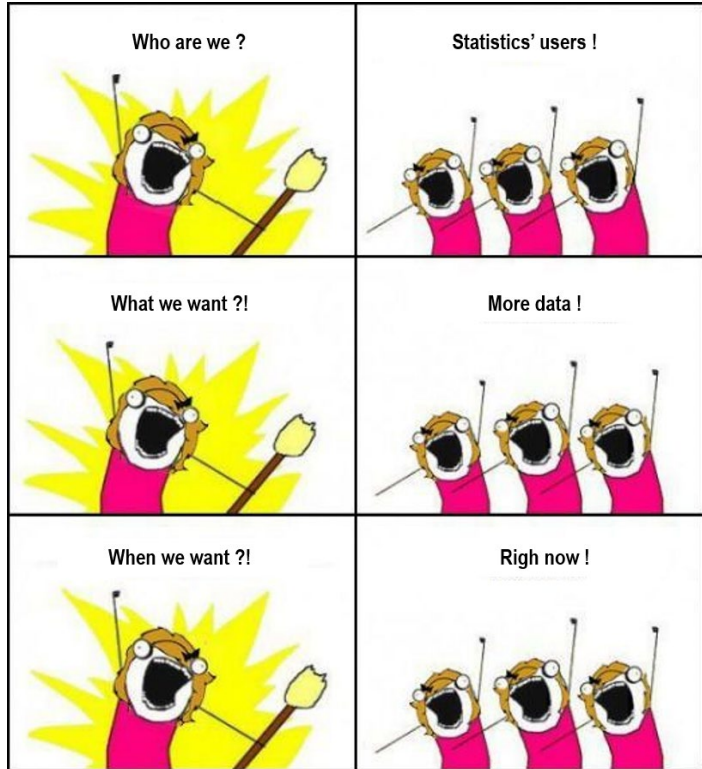
[2] Data use: What information is **needed**?

67

55

A two-step procedure: Collect information and transform it into knowledge

How to transform **information to knowledge** (2)

What information does this dataset contain?

➜ **Metadata**

How can I access this dataset?

➜ **Access information**

What can I learn from **other users** of this dataset?

➜ **Usage information**

How to **collect information** on datasets (1)

Metadata, Annodata and Usage Information
14 October 2022
Page 13

68

56
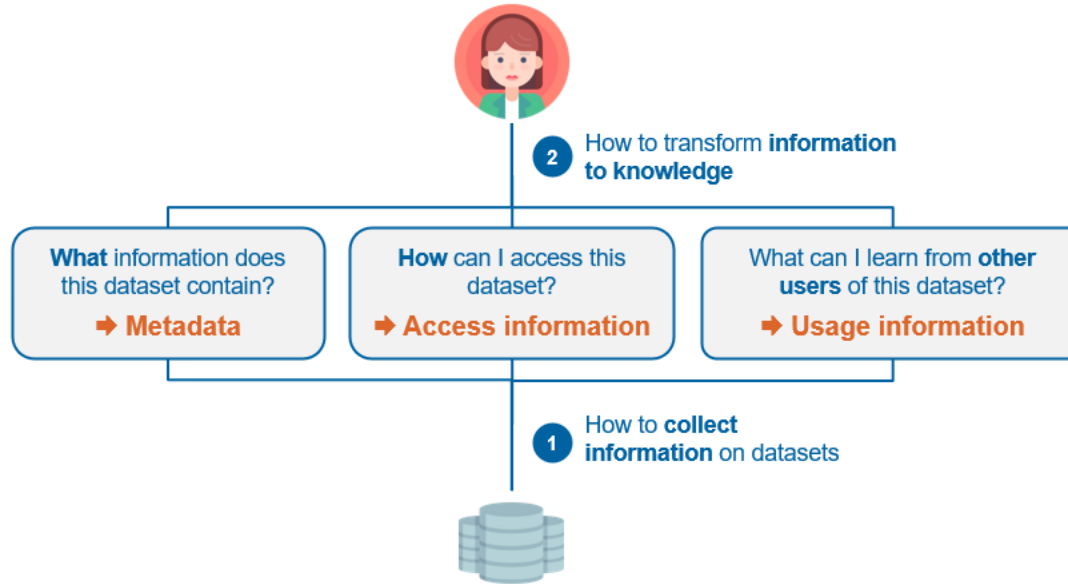
WG 6: Associated issues – environmental sustainability, international collaboration, and legal considerations

Members: Mark Dean (lead), Fred Streitz, Daniela Braga, Oren Etzioni

Objective

Develop recommendations for steps the NAIRR management entity could take to mitigate impact on the environment, promote international collaboration, and address legal issues.

Key Questions

1. How can the NAIRR mitigate detrimental impact of the resources it provides on the environment?
   2 Recommendations

2. How could the NAIRR support international collaboration?
   3 Recommendations

3. Will establishment of the NAIRR require any new statutory authorities from Congress?
   1 Recommendation

4. How could the NAIRR act as a facilitator to ease challenges around data use agreements? Could the NAIRR propose a common data use agreement, and if so, what would that look like?
   1 Recommendation

Q1 - How can the NAIRR mitigate detrimental impact of the resources it provides on the environment?

General Statement - NAIRR should express environmental consciousness in everything it does. NAIRR's two primary activities that directly impact the environment include:1) the management and allocation of a computing infrastructure for AI research and 2) the fostering of AI technologies that will aid in the study and impact of natural occurring events and human activity on the environment.

Recommendation 1: NAIRR should have an environmental policy that establishes metrics for 1) Total Cost to the Environment (TCE) and 2) E-Waste management.

Example TCE metrics: Power Usage Effectiveness (PUE), Data Center Performance Efficiency (DCPE), Deployed Hardware Utilization Ratio (DH-UR), Deployed Hardware Utilization Efficiency (DH-UE), and HVAC System Effectiveness (HSE).

Example E-Waste Management actions: Maximizing life-cycle and usability of systems, an electronic waste recycling and tracking plan, and a systems and equipment repurposing and reselling hardware plan.

Q1 (cont.) - How can the NAIRR mitigate detrimental impact of the resources it provides on the environment?

Recommendation 2: NAIRR should promote the importance of the study of the environment through its support of relevant AI research areas. It should track and report on the percentage of time the NAIRR infrastructure is used for environment research.

Example areas of study: Environmental systems modeling and analysis, climate modeling, bio-systems modeling, water shed modeling and analysis, energy systems management, and waste management. Predictive maintenance and sensor systems learning are other areas of AI research.

NAIRR should also identify computing technologies that are energy efficient, carbon neutral, and have little or no negative effect on water quality, air quality, waste accumulation, soil contamination or the USA's carbon footprint.

Q2 - How could the NAIRR support international collaboration?

Predecisional

Recommendation 3: NAIRR should follow the guidelines for international collaboration set by key government research agencies: DOE, NIH, NSF, DHS, OSTP, EPA, etc. NAIRR is required to comply with the laws and regulations issued by the U.S. government related to the export of both goods and services, which includes certain technologies, software, and hardware for reasons of national security, foreign policy, and for competitive trade reasons.

Recommendation 4: NAIRR should leverage the International Science Council's (ISC) Committee on Data (CODATA - https://codata.org/), through the National Academies' US National CODATA Committee national membership. With its ties to the ISC, CODATA has a built-in way for coordinating with 200 international scientific unions and associations as well as national and regional scientific organizations, including academies and research councils it represents.

Q2 (cont.) - How could the NAIRR support international collaboration?

Recommendation 5: NAIRR should establish a lightweight coordination office, similar to the NSF-funded EarthCube Office, to represent the organization at other international conferences where stakeholders gather including: AAAI, NeurIPS, KDD, and ICML. The Office would document successes in science stories, ensure NAIRR opportunities are disseminated broadly through international networks, and coordinate presentations and outreach in key forums.

Q3 - Will establishment of the NAIRR require any new statutory authorities from Congress?

Recommendation 6: NAIRR should be established with similar statutory authorities and authorizing legislation as the US's High Performance Computing and Communications (HPCC) Program created by the High Performance Computing (HPC) Act of 1991 (renamed to the Networking and Information Technology Research and Development (NITRD) Program via the American Innovation and Competitiveness Act of 2017.)

For reference, the other congressional legislations of note that supported the US's HPCC initiative include: The Next Generation Internet Research Act of 1998, America COMPETES Act of 2007, and the Cybersecurity Enhancement Act of 2014.

Q4 - How could the NAIRR act as a facilitator to ease challenges around data use agreements? Could the NAIRR propose a common data use agreement, and if so, what would that look like?

Recommendation 7: **It would be desirable to craft a single DUA in collaboration with the NSDS[†] around Open Data terms but with addenda that define the additional terms necessary to share Protected Data. Should this prove too difficult in practice, NAIRR and NSDS could instead establish a separate DUA to meet the need for documentation of use of open data.**

Although creating a simple yet broadly applicable template for DUAs has proven elusive to many organizations, the goal is to establish as simple a DUA or pair of DUAs as possible. NAIRR should investigate and leverage best practices from existing DUAs at universities and other government research organizations that have proved to be effective and efficient in the support of safely, securely, and legally sharing of data.

[†] National Secure Data Service

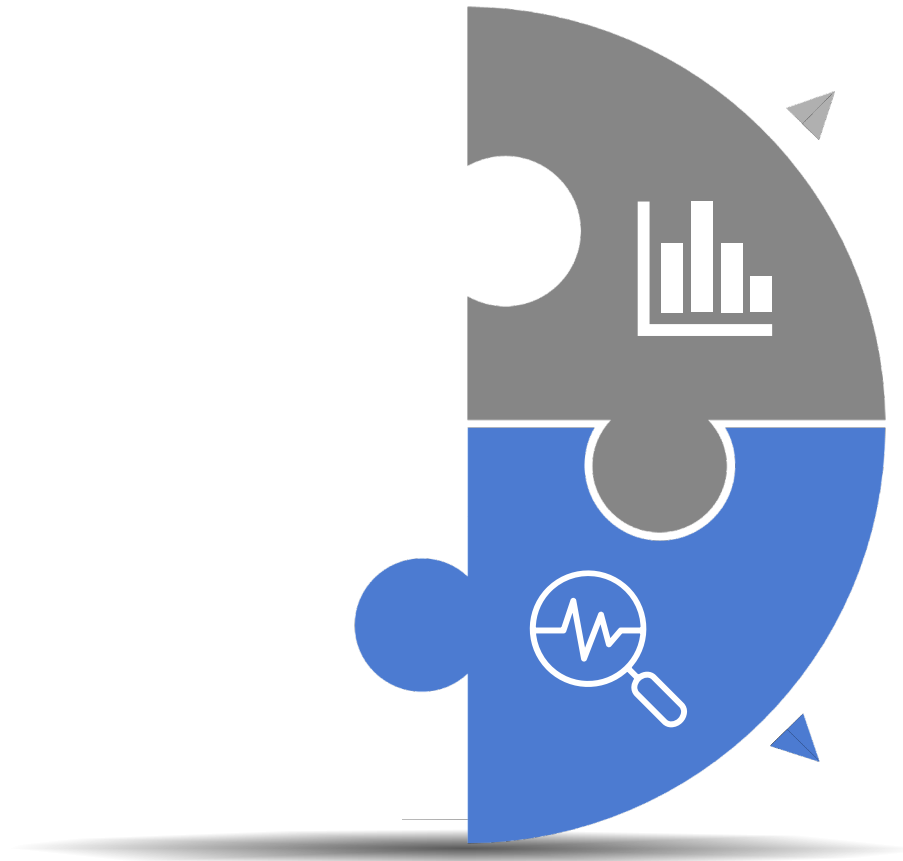# Associated Federal Efforts for Provision of Data and Computing Resources

Associated Federal Efforts for Provision of Data and Computing Resources

NATIONAL SECURE DATA SERVICE (NSDS) + AMERICA'S DATA HUB

Providing a centralized, secure means of accessing restricted data

Associated Federal Efforts for Provision of Data and Computing Resources

NATIONAL SECURE DATA SERVICE (NSDS) + AMERICA'S DATA HUB

Providing a centralized, secure means of accessing restricted data

NATIONAL DISCOVERY CLOUD

Advancing research on the technical foundations of a discovery computational environment

Associated Federal Efforts for Provision of Data and Computing Resources

Predecisional

NATIONAL SECURE DATA SERVICE (NSDS) + AMERICA'S DATA HUB

Providing a centralized, secure means of accessing restricted data

SUBCOMMITTEE ON OPEN SCIENCE

Establishing an open science infrastructure

NATIONAL DISCOVERY CLOUD

Advancing research on the technical foundations of a discovery computational environment

Associated Federal Efforts for Provision of Data and Computing Resources

FUTURE ADVANCED COMPUTING ECOSYSTEM

Integrating data, compute, software, and educational resources for the science and engineering community

NATIONAL SECURE DATA SERVICE (NSDS) + AMERICA'S DATA HUB

Providing a centralized, secure means of accessing restricted data

SUBCOMMITTEE ON OPEN SCIENCE

Establishing an open science infrastructure

NATIONAL DISCOVERY CLOUD

Advancing research on the technical foundations of a discovery computational environment

Associated Federal Efforts for Provision of Data and Computing Resources

FUTURE ADVANCED COMPUTING ECOSYSTEM
Integrating data, compute, software, and educational resources for the science and engineering community

NATIONAL SECURE DATA SERVICE (NSDS) + AMERICA'S DATA HUB
Providing a centralized, secure means of accessing restricted data

SUBCOMMITTEE ON OPEN SCIENCE
Establishing an open science infrastructure

NATIONAL DISCOVERY CLOUD
Advancing research on the technical foundations of a discovery computational environment

83

Associated Federal Efforts for Provision of Data and Computing Resources

NATIONAL SECURE DATA SERVICE (NSDS) + AMERICA'S DATA HUB

Providing a centralized, secure means of accessing restricted data

Shelly Martinez

Senior Statistician, Office of Management and Budget

The White House

The Federal Statistical System & NSDS

Stat System blueprint for future:

- **Building from the trusted statistical infrastructure**

- Relying **heavily on the framework provided by national statistical law (CIPSEA 2018)**

- **Operating as a seamless system, including**
  - **to simplify evidence-building process by providing a centralized, secure means of applying for access to restricted data**

NSDS:

- **FY 2023/2024 demonstration/pilots**

- **Likely focus on capacity, services, and coordination complementary to existing Stat System efforts**
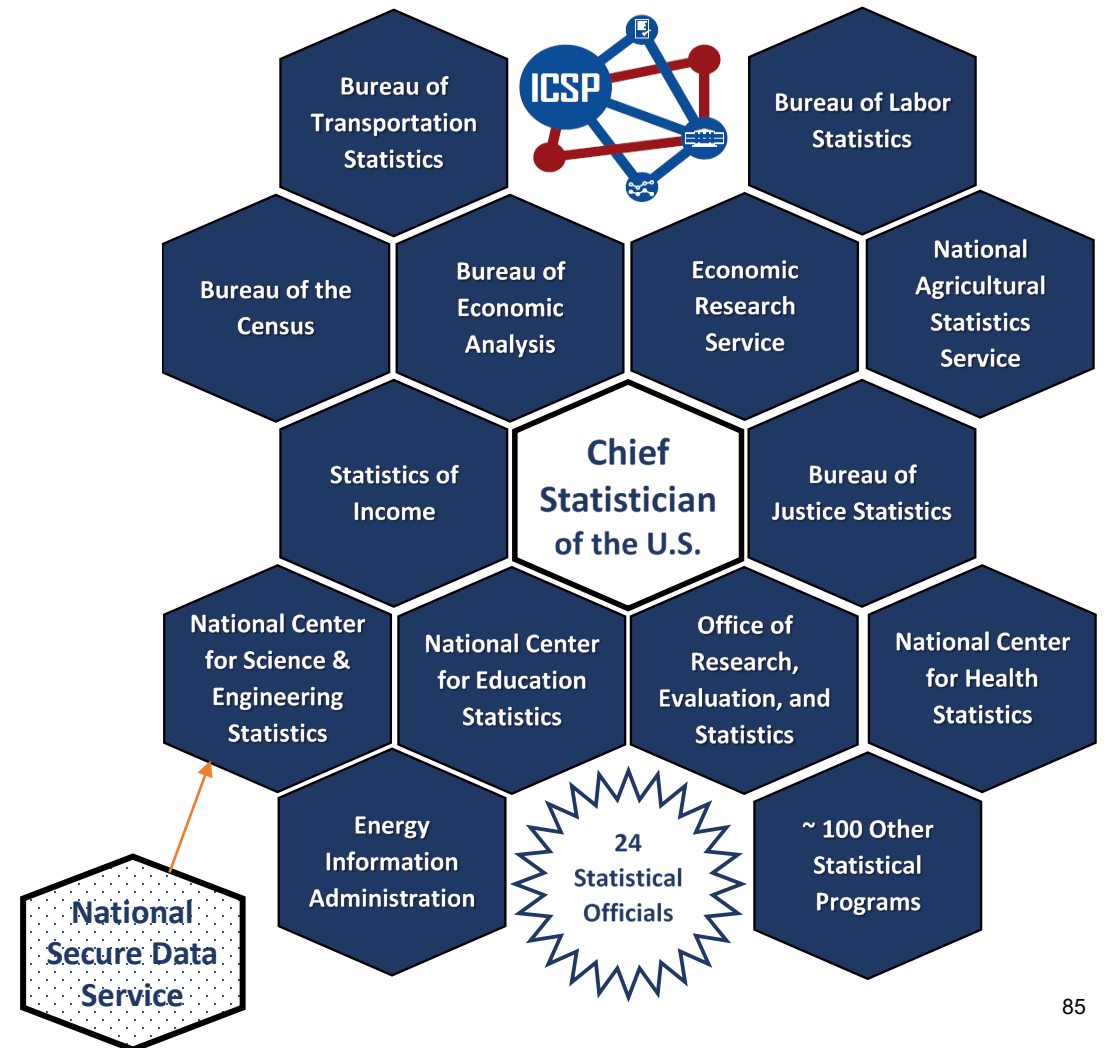
Associated Federal Efforts for Provision of Data and Computing Resources

NATIONAL SECURE
DATA SERVICE (NSDS) +
AMERICA'S DATA HUB

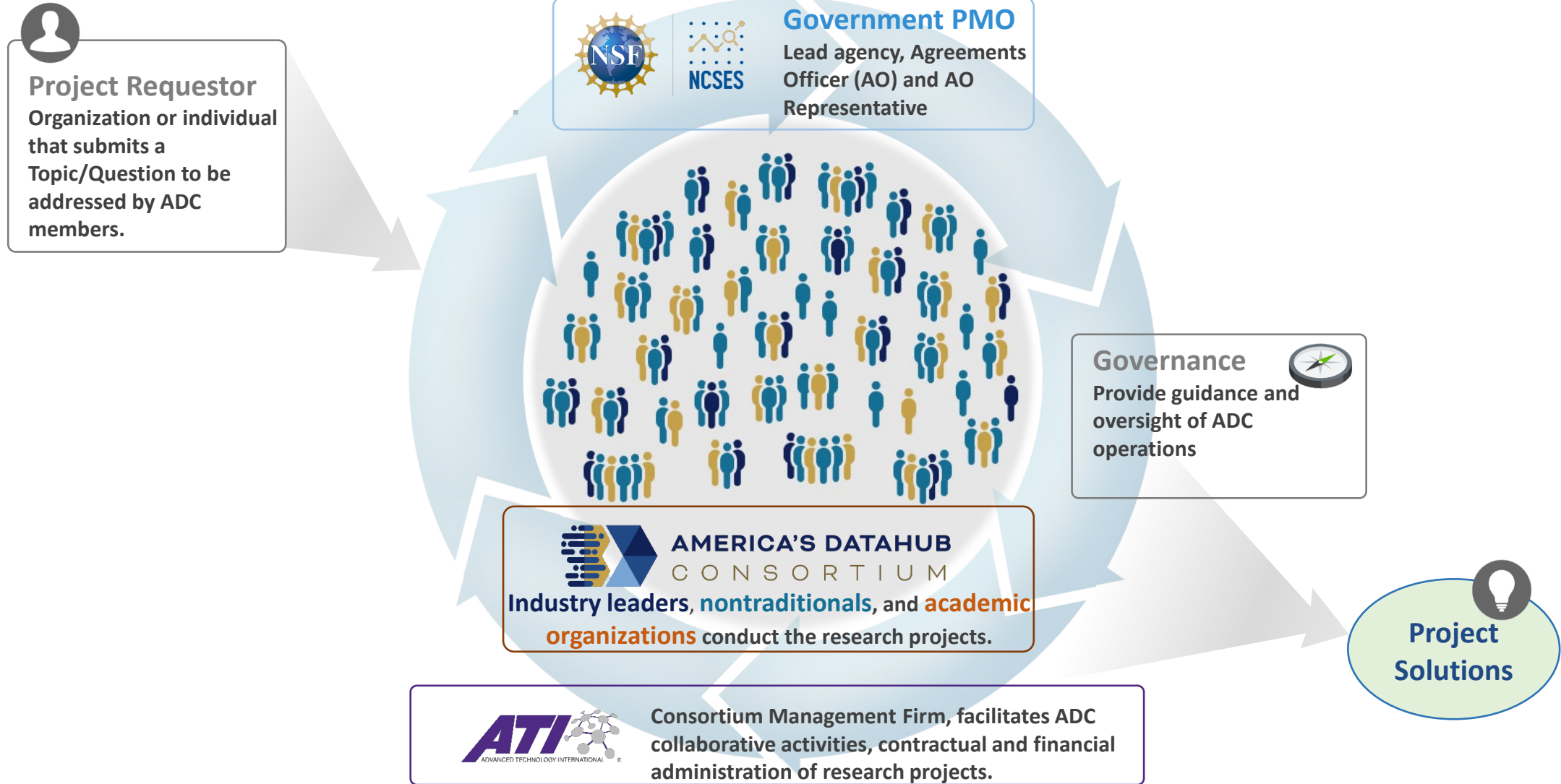Providing a centralized, secure means of accessing restricted data

Vipin Arora

Deputy Director, National Center for Science and Engineering Statistics
National Science Foundation

Our vision for America's DataHub

America's DataHub will be an enduring national asset, where eligible people and secure data come together for collaborative research and decision-making that will benefit the American public.

The current structure of America's DataHub

Project Requestor
Organization or individual that submits a Topic/Question to be addressed by ADC members.

Government PMO
Lead agency, Agreements Officer (AO) and AO Representative

Governance
Provide guidance and oversight of ADC operations

AMERICA'S DATAHUB CONSORTIUM
Industry leaders, nontraditionals, and academic organizations conduct the research projects.

ATI ADVANCED TECHNOLOGY INTERNATIONAL
Consortium Management Firm, facilitates ADC collaborative activities, contractual and financial administration of research projects.

Project Solutions

Associated Federal Efforts for Provision of Data and Computing Resources

FUTURE ADVANCED COMPUTING ECOSYSTEM

Integrating data, compute, software, and educational resources for the science and engineering community
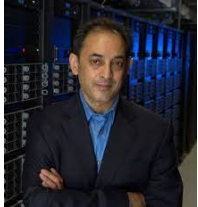
Manish Parashar

Office Director, Office of Advanced Cyberinfrastructure

National Science Foundation

Pioneering the Future Advanced Computing Ecosystem: A Strategic Plan

- Realize
  system
  suppo
  defens

- Key str
  - Ut
    go
  - Es
  - Su
    fu
  - Ex
    su

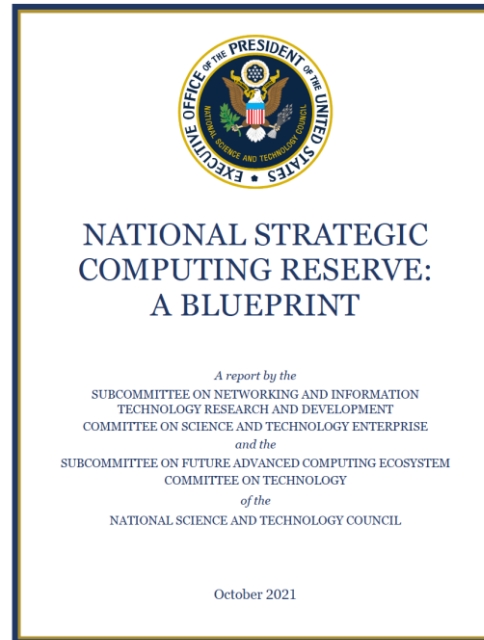**Strategic Objective 1: Advanced computing ecosystem as a strategic national asset**

Vision: A federated set of resources and services that are heterogeneous in architecture, resource type, and usage mode; support the seamless alignment between the varied resource offerings and application use cases

- Partnerships across government, academia, nonprofits and industry essential

- Complements the objectives and activities of other initiatives and national priorities, including NQI, AAII, IoTF, and envisions synergistic relationships.

https://www.nitrd.gov/pubs/Future-Advanced-Computing-Ecosystem-Strategic-Plan-Nov-2020.pdf 90

Advanced Computing Ecosystem as a Strategic National Asset

## National Strategic Computing Reserve (NSCR)

- A coalition of experts and resource providers that could be mobilized quickly to provide critical computational resources in times of urgent need
  - Build on experiences from the COVID-19 HPC Consortium, responses to RFI
  - Aligns with the FACE Strategic plan

NATIONAL STRATEGIC COMPUTING RESERVE: A BLUEPRINT

A report by the
SUBCOMMITTEE ON NETWORKING AND INFORMATION TECHNOLOGY RESEARCH AND DEVELOPMENT
COMMITTEE ON SCIENCE AND TECHNOLOGY ENTERPRISE
and the
SUBCOMMITTEE ON FUTURE ADVANCED COMPUTING ECOSYSTEM
COMMITTEE ON TECHNOLOGY
of the
NATIONAL SCIENCE AND TECHNOLOGY COUNCIL

October 2021

https://www.whitehouse.gov/wp-content/uploads/2021/10/National-Strategic-Computing-Reserve-Blueprint-Oct2021.pdf

NSF's Advanced Cyberinfrastructure Ecosystem

- Network of advanced systems and services
  - Leadership and capacity systems, testbeds
- Federation (PATh) and coordination services (ACCESS)
- Scalable user support networks
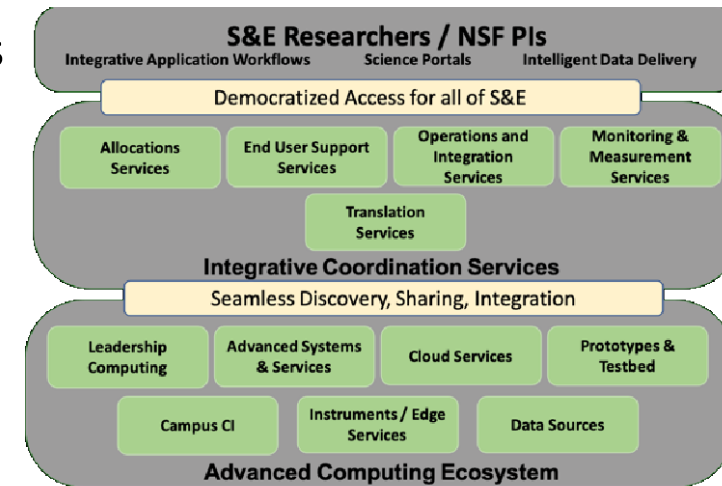
Democratized access to an advanced CI Ecosystem

Associated Federal Efforts for Provision of Data and Computing Resources

Kamie Roberts

Director, National Coordination Office
for the Networking and Information Technology Research and Development Program

NATIONAL DISCOVERY CLOUD

Advancing research on the technical foundations of a discovery computational environment

Associated Federal Efforts for Provision of Data and Computing Resources

Jerry Sheehan

Deputy Director for Policy and External Affairs, National Library of Medicine

National Institutes of Health

SUBCOMMITTEE ON OPEN SCIENCE

Establishing an open science infrastructure

NSTC Subcommittee on Open Science (SOS)

CHARTER
of the
SUBCOMMITTEE ON OPEN SCIENCE
COMMITTEE ON SCIENCE
NATIONAL SCIENCE AND TECHNOLOGY COUNCIL

A.   Official Designation

The Subcommittee on Open Science (SOS) is hereby re-established by action of the Committee on Science of the National Science and Technology Council (NSTC).

B.   Purpose and Scope

The purpose of the SOS is to coordinate and advance efforts related to open science across Federal agencies. Such efforts include, but are not limited to, making the products and processes of Federally funded research, including scholarly publications and data, more equitably findable, accessible, interoperable, and reusable to the public, policymakers, industry, and the scientific community.

C.   Functions

The SOS shall (no order of priority is implied):

- Provide continuity in functions, as appropriate, from its predecessor deliberative bodies including the Interagency Public Access Committee and the Interagency Working Group on Open Science;
- Promote the exchange of information regarding open science initiatives, policies, practices, and programs between and among Federal agencies, including but not limited to improving consistency and sharing good practices for implementation;
- Facilitate interagency coordination and cooperation on topics of common interest related to open science;
- Advise on and contribute to OSTP's reporting requirements to Congress on public access of Federally funded research and data;
- Facilitate coordination of agency efforts to support training, education, and Federal workforce development related to open science policies, practices, and programs; and,
- Engage with national and international stakeholders to advance open science.

The purpose of the SOS is to coordinate and advance efforts related to open science across Federal agencies. Such efforts include, but are not limited to, making the products and processes of Federally funded research, including scholarly publications and data, more equitably findable, accessible, interoperable, and reusable to the public, policymakers, industry, and the scientific community.

SOS Subgroup on Open Science Infrastructure

Definition

Open science infrastructure means digital services and technologies (physical and virtual) that enable and provide equitable access to open science research outputs.

Charge

Improve coordination of ongoing agency and interagency efforts to establish digital infrastructures to support open science. **Identify cross-agency approaches for digital infrastructure** that promote open science objectives and the inter-linking of research objects.

Vision

Interoperable computing, data, and analytical tools that opens research to everyone. The SOS Open Science Infrastructure sub-group seeks to advance an equitable, evolving research computing ecosystem that maximizes the benefits of open science and advances knowledge for the United States and its international partners.

FINAL REPORT DISCUSSION

TESS DEBLANC-KNOWLES, SENIOR POLICY ADVISOR, NATIONAL AI INITIATIVE OFFICE,
WHITE HOUSE OFFICE OF SCIENCE AND TECHNOLOGY POLICY

Proposed Discussion Topics

1) Data hosting
2) Governance structure
3) "Day One" capabilities
4) NAIRR user capacity
5) Role for NAIRR in AI workforce development
6) Others?

Data Hosting

- NAIRR will host curated data through its resource provider(s)
- Data hosted by the NAIRR can include:
  — Public and open datasets
    ◦ Seeded with high-value/high-use open data
  — Data resulting from NAIRR-supported research
- Decisions on what data to host will be made by the NAIRR operating entity using a transparent allocation process and the availability of resource
- Data will be decommissioned using cost/benefit analysis

NAIRR Governance Structure
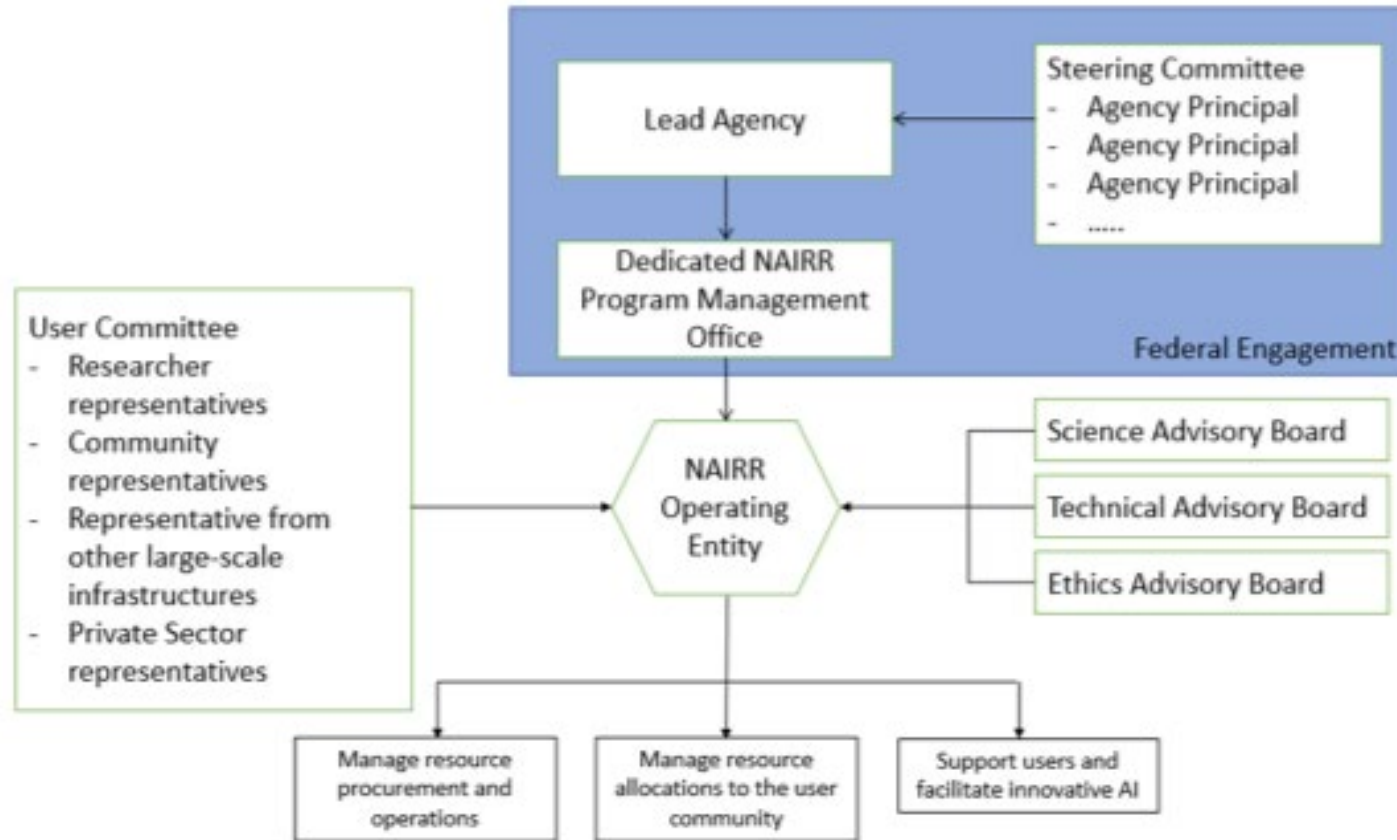
Standing up NAIRR Governance Structure

1. **Fund NAIRR**
   a. Congress funds the NAIRR through appropriations to supporting agencies, with appropriation for lead agency sufficient to run core operations
2. **Set up governance structure**
   a. Steering Committee (SC) created
   b. Lead Agency identified
   c. The NAIRR Program Management Office (PMO) created at Lead Agency
3. **Establish the NAIRR Operating Entity (OE)**
   a. SC defines parameters/characteristics for OE
   b. PMO works with SC to create a solicitation and select OE
   c. Lead Agency funds core operations and makes coordinated investments with other Agencies
   d. PMO awards contract to selected OE and provides oversight
   e. Boards/advisory committees, etc., created

3. **Fund NAIRR Resource Providers (RP)**
   a. OE works with PMO (with guidance from SC) to develop a multi-agency solicitation for RPs (compute, data, software/tools, support, instructional content, portal, etc.).
   b. SC reviews submissions with PMO and OE and select awardees
   c. OE negotiates contracts with slate of RP identified (and funded) by agencies
4. **Operate NAIRR**
   a. OE integrates RPs and initiates NAIRR operations
   b. PMO oversees OE operations based on the award T&Cs
   c. Boards/advisory committees provide oversight and guidance

"Day One" Capabilities

Proposed "Day One" capabilities for the NAIRR in its pilot phase:

- User access portal with allocation and identity management system

- Data publication system that allows datasets to be added

- Linked indices of resources (e.g., testbeds, education/training materials)

- Compute resource providers (established via solicitation/USG funding) accepting computational jobs

  — Mix of on-premises (HPC) and cloud (CPU/GPU options with multiple accelerators per node)

  — Preferred (desired but not required): at least one "experimental" compute resource and one ML supercomputer to train 10^{12} parameter models

- High-speed network

- Sufficient memory capacity

NAIRR User Capacity

- In year 1 of operations, 48-60 million hours on quad-GPU nodes:
  — 50,000 students/researchers x 1,000 hours/user, **or**
  — 25,000 students/researchers x 1,000 hours/user PLUS 40 projects at scale of GPT-3 benchmark
- By steady-state operations, 140-180 million hours on quad-GPU nodes:
  — 150,000 students/researchers x 1,000 hours/user, **or**
  — 75,000 students/researchers x 1,000 hours/user PLUS 60 projects at scale of GPT-3 benchmark
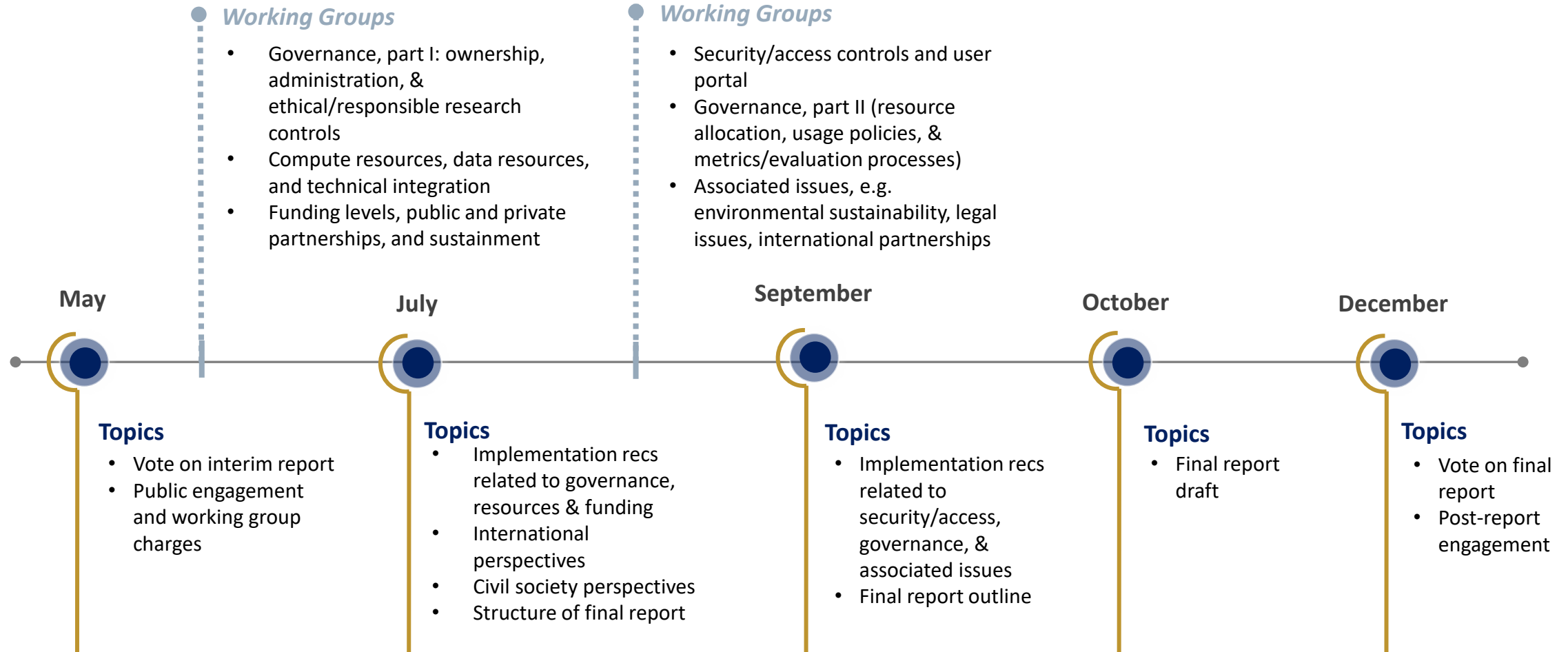
Role for NAIRR in AI Workforce Development

- Most of our discussions have focused on NAIRR as a means to give AI technical resources to researchers
  - Compute, Data, Models, plus training and outreach

- This is important in and of itself, but is not likely to change the make up of the field
  - Our "democratize access" mandate will make the technical resources available —but likely only to those *already* committed to working in AU

- Should we consider using NAIRR resources to provide incentives to entice a more diverse audience into working in AI?
  - Such incentives might include:
    - Summer salary for faculty at under-resourced institutions
    - Fellowships/scholarships for students who will incorporate AI into their research/career plans

Proposed Discussion Topics

1) Data hosting
2) Governance structure
3) "Day One" capabilities
4) NAIRR user capacity
5) Role for NAIRR in AI workforce development
6) Others?

Final Report Development: Key Dates

September 23: First draft sent to TF

September 30: TF feedback on first draft

October 12: Second draft sent to TF and interagency stakeholders

October 21: *TF meeting* – deliberate on report content

November 1: Third draft shared with OSTP/NSF leadership

November 22: Final draft sent to TF

December 7: *TF meeting* – vote on final report

Per its legislative mandate, the TF will exist for 90 days following submission of the final report. The TF will use December 2022-February 2023 to socialize the recommendations put forth in the final roadmap and implementation plan.