NATIONAL ACADEMIES Sciences Engineering Medicine

U.S. National Committee for CODATA

2023 U.S. Research Data Summit
Focus Group Themes and Supporting Information

The National Academies of Sciences, Engineering, and Medicine appointed the U.S. Research Data Summit Planning Committee, under the auspices of the U.S. National Committee for CODATA, Board on International Scientific Organizations, to convene U.S. leaders of research data organizations across sectors who can shape and influence U.S. research data policies and practices. The objectives of the convening are to:

- Increase coherence of interests and activities among the cross-sector research data organizations
- Increase communication and collaboration across sectors on shared priorities
- Reduce duplication of effort
- Position the United States to be well-represented in international discussions on research data

Summit invitees include research data organization leaders in academia, government, industry, and related nonprofits who can effect change in the near future.

The Summit outcomes will reflect the experience, knowledge, discussion, recommendations, and commitment to the objectives of the participating leaders. They will be informed by the results from six focus groups convened by the US Research Data Planning Committee from April to July 2023 in preparation for the Summit. These focus groups are:

- AI: Organizational Approaches to AI Inputs & Outputs Focus Groups (1 session)
- Cross-Sector Access to Research Data Focus Groups (2 sessions)
- Data Needs for Decarbonization Focus Groups (2 sessions)
- Data for Disruption: Responding to Disasters and Shocks in Disadvantaged Communities Focus Group (1 session)
- Indigenous Data Governance Focus Group (1 session)
- Justice, Equity Diversity and Inclusion Research Data Integration Focus Group (1 session)

The topics were selected based on input from the 2022 survey by the Association of Research Libraries and the U.S. National Committee for CODATA, discussions with the U.S. National Committee for CODATA, and refinement by the U.S. Research Data Summit Planning Committee.

This document provides overarching themes from the focus groups, as well as summaries of the session notes for each of the focus groups. The overarching themes and the separate focus group

summaries are organized into the four categories that correspond with the objectives of the summit.

The Summit proceedings in brief will be shared with the public.

Planning Committee:

Co-Chairs:
Jennifer Hansen, Microsoft Corporation
Mary Lee Kennedy, Association of Research Libraries

Members:
Jason T. Black, Florida Agricultural and Mechanical University
Stephanie R. Carroll, University of Arizona
Bonnie C. Carroll, Information International Associates, Inc. (retired)
David McCollum, Oak Ridge National Laboratory
Cynthia R. Hudson Vitale, The Data Curation Network

Staff Officers:
Thomas S. Arrison, National Academies of Sciences, Engineering, and Medicine
Diamond de Guzman, National Academies of Sciences, Engineering, and Medicine

Facilitator:
Joel Cutcher-Gershenfeld, Brandeis University

Supporting Sponsors:
- Bill and Melinda Gates Foundation
- U.S. National Science Foundation
- Research Data Alliance-US
- Plus in-kind support from West Big Data Innovation Hub

NATIONAL ACADEMIES Sciences Engineering Medicine

Overarching Themes
Informing the Objectives of the U.S. Research Data Summit[1]

Objective 1: Increase coherence of interests and activities among the cross-sector research data organizations -- opportunities include:

- **Education and Training:** Education and Training in making data available in ways that are open and reusable.
- **Diverse Stakeholder Voices:** Mechanisms to lift up diverse stakeholder voices with respect to the governance of research data, particularly indigenous and underserved communities – nothing about us without us.
- **Trust in Science:** Increased societal understanding of the central role of research data in addressing global challenges.

Objective 2: Increase communication and collaboration across sectors on shared priorities -- opportunities include:

- **Natural and socio-economic data:** Communication and collaboration initiatives.
- **Public and private data:** Communication and collaboration initiatives.
- **Bridging across fields and disciplines:** Communication and collaboration initiatives.
- **Spatial and temporal data:** Communication and collaboration initiatives.
- **Researcher, policy maker, association, publisher, agency, business organization, citizen, and other collaborations:** Bridging across roles, organizations and institutions.

Objective 3: Reduce duplication of effort -- opportunities include:

- **Connecting Innovations:** There are dozens of illustrative innovations (see the listing by Focus Groups), but these "islands of innovation" need connection, communication, coordination, and, where appropriate, standardization.
- **Protocols for Connection, Communication, Coordination, and Standardization:** Provide guidance to innovative initiatives on how to identify relevant complementary initiatives and constructively engage with them.

Objective 4: Position the United States to be well-represented in international discussions on research data -- opportunities include:

[1] This document was prepared by Joel E. Cutcher-Gershenfeld to stimulate discussion at the October 10-11, 2023 U.S. Research Data Summit. The perspectives expressed are those of the author and do not necessarily reflect the official policies or positions of his employing organization. This document is not a report of the National Academies of Sciences, Engineering, and Medicine and has not been subjected to its review procedures.

- **Ecosystem Governance:** Research data functions in ecosystems of individual researchers, citizen scientists, data repositories, compute resources, software and models, public agencies, non-governmental organizations, commercial enterprises, indigenous populations, educational organizations, nation-states, and other stakeholders and rights holders with incomplete mechanisms for effective governance.

Illustrative Current State Innovations (see appendix for the listing of all 72 examples organized by Focus Group)

- **AI: Organizational Approaches to AI Inputs & Outputs:** 4 illustrative innovations from Focus Groups
  - *An Example:* Gates Foundation initial announcement on AI principles
- **Cross-Sector Access to Research Data:** 20 illustrative innovations from Focus Groups
  - *An Example:* WorldFAIR: CODATA coordinating with RDA through EU Horizon, surfacing cross-disciplinary issues to implement FAIR principles
- **Data Needs for Decarbonization:** 22 illustrative innovations from Focus Groups
  - *An Example:* Under the Bipartisan Infrastructure Law (BIL) / Infrastructure Investment and Jobs Act (IIJA), the Carbon Management EDX4CCS is addressing many needs for Carbon Capture, Utilization, and Storage (CCUS) data users
- **Data for Disruption:** 11 illustrative innovations from the Focus Group
  - *An Example:* The Belgian CRED EM-DAT database is a source of risk impact data.
- **Indigenous Data Governance:** 5 illustrative innovations from the Focus Group
  - *An Example:* ESIP worked with GIDA on operationalizing CARE principles for repositories
- **Justice, Equity Diversity and Inclusion:** 10 illustrative innovations from the Focus Group
  - *An Example:* The Minority Serving Cyberinfrastructure Consortium is advancing research and educational computing, data and software infrastructure at HBCUs, TCUs, HSIs, and other MSIs

Selected Overarching Questions From the Focus Groups

- **AI Focus Groups**
  - How can AI seamlessly integrate into research processes to accelerate discoveries and facilitate interdisciplinary collaborations?
  - How do we envision collaboration and knowledge sharing on AI among different stakeholders, such as research institutions, government bodies, and civil society, to enhance the utilization of AI for research breakthroughs?
- **Cross-Sector Focus Groups**
  - How to best advance cross sector collaboration among research data organizations?
  - Given the federal agencies implementation plans regarding the OSTP's Memo on Ensuring Free, Immediate, and Equitable Access to Federally Funded Research,

what types of cross-sector collaborations will advance adoption and equitable access to research data for pure and applied science?

- **Decarbonization Focus Groups**
  - What can be done to facilitate the use and integration of diverse public and private data sources to address the decarbonization challenge?
  - What new data sources are becoming available and what gaps still remain, to understand how U.S. federal, state, and local governments, as well as private industry, can most effectively achieve deep decarbonization by mid-century, particularly in the transport and electricity sectors?
- **Disruption Focus Group**
  - What is currently being done in encouraging cross-disciplinary data cooperation with respect to disasters, extreme events, shocks, and systems transitions (natural, economic, health, and security)?
  - How can disparate data-sets and experts from different disciplines gain a more holistic understanding of the many vulnerabilities that disadvantaged households and communities are simultaneously facing, particularly at the time when major disruptions occur?
- **Indigenous Data Governance Focus Group**
  - How to support indigenous communities having their own data repositories?
  - How to govern the use of AI to create synthetic indigenous genomic data sets?
- **Justice, Equity Diversity and Inclusion Focus Group**
  - The bigger picture that we are operating in with the DEI legislation and the SCOTUS ruling, which are resulting in a contraction in support of DEI -- how will that impact attention to DEI with research data?
  - How to ensure support, mentoring, and diverse leaders for the next generation of research data professionals, including those from underserved communities?

A Dozen Interconnected, Underlying Tensions (naming the tensions is the first step)

Research Data Tensions

- **Domain Expertise:** Domain expertise is essential for well curated data (avoiding data being "dumped" in generic repositories), but just focusing within domains risks data silos.
- **Trust in Research Data:** Trust in research data depends on data quality and improves with familiarity, but the data requirements for most societal challenges involves data of variable quality and requires researchers (and the public) to enter unfamiliar territory.
- **Data Provenance and Interoperability:** FAIR principles encourage wide reuse, but that increases interoperability challenges, multiplicative potential sources of errors, and raises provenance issues as data is combined and transformed.
- **Privacy, Risk, Bias,and Security:** Research data has inherent risk, embedded forms of bias, and threats to privacy and security, but the needed infrastructure, resources, and expertise to address these challenges are generally not available to individual researchers.

- **Common Pool Resources:** Those who benefit most from shared research data resources are generally not the same parties bearing the costs and risks.
- **Vulnerable populations:** Prioritizing research data involving vulnerable populations (CARE principles) creates additional vulnerabilities by making these data more visible.

Organizational and Institutional Tensions Associated with Research Data

- **Rates of Change:** New technologies associated with research data (AI and associated Machine Learning, Large Language Models, Narrow and Deep AI, etc.) are advancing with accelerating rates of change, while the associated organizations and institutions typically change incrementally.
- **Institutional Agility:** Institutions represent the stable foundations of society, yet today's challenges require institutional arrangements that are agile and adaptive.
- **Data Workers:** Individuals with needed expertise in the curation and reuse of research data too often lack the professional standing, resources, and career paths commensurate with the needed work.
- **Collective Impacts:** Stakeholders and rights holders can accomplish more together than they can separately with respect to research data, but no one party has the resources and legitimacy to lead.
- **Cross-Domain Governance:** Standards for privacy, data governance, transparency, and related matters vary across nations, regions, and domains, yet effective governance depends on reducing variation and increasing alignment.
- **Multi-Stakeholder Consortia:** Consortia that bring together relevant stakeholders associated with research data are relatively easy to launch, but hard to sustain.

NATIONAL ACADEMIES Sciences Engineering Medicine

Appendix

Summarized from U.S. Research Data Summit Focus Group Session Notes

AI Focus Groups: Organizational Approaches to AI Inputs & Outputs

Scope of Focus Group:

This group explored how organizations approach AI inputs and outputs in terms of policies, use, and ethics. Participants discussed the development of policies, improving organizational operations, and ethical considerations.

Elements of a Success Vision (Desired State):

- Fully reproducible results with Artificial Intelligence (AI)/ Machine Learning (ML)/Large Language (LL) models
- Transparency in data sources and funding associated with AI/ML/LL models
- AI helps to detect bias in AI
- Monotonous aspects of methods (protocol, model construction, etc.) are automated with AI
- Pattern recognition across fields, disciplines, and contexts accelerated with AI/ML/LL

Exemplars and Resources (Current State -- 4 illustrative innovations):

- Gates Foundation initial announcement on AI principles
- AGU Ethical and Responsible Use of AI/ML in the Earth, Space, and Environmental Sciences
- Gates Foundation Grant Challenges for AI technologies
- NAIRR Task Force – The National AI Research Resource Task Force ("Task Force") is a multi-stakeholder team across federal agencies charged with investigating the feasibility of a National Artificial Intelligence Research Resource (NAIRR), and proposing a roadmap detailing how to establish and sustain the NAIRR. The NAIRR is envisioned as a shared computing and data infrastructure that provides AI researchers with access to compute resources and high-quality data, along with appropriate educational tools and user support.

Open Questions (Delta State):

- How to resolve the IP issues around training data that is ingested in AI/ML/LL models?

- How to document and/or know what training data is in use?
- Do AI/ML/LL models perform differently when trained on scholarly data compared to all open data?
- Does there need to be a defined format for text files ingested in AL/ML/LL models?
- How (if all) to curate, store and share the output of AI/ML/LL models?

Cross-Sector Focus Groups: Cross Sector Access to Research Data (combination of two sessions)

Scope of Focus Group:

Natural and human-made challenges do not observe the organizational and institutional boundaries that we have established. This group explored the dynamics of finding, accessing, and reusing inter-operative data (FAIR approach) across fields, disciplines, domains, and sectors. This included public and private data.

Elements of a Success Vision (Desired State):

- Free, fair and equitable access to federally funded research data
- Advancing public trust in science and research data
- FAIR (Findable, Accessible, Interoperable, and Reusable) data standards widely adopted
- Field, discipline, domain, and sectors specific ecosystems to for data
- Environmental sustainability for compute, data, and software resources
- Institutional sustainability for data repositories
- Career paths for data professionals
- Data rights built into meta data
- Trusted data principles complementary to trusted data repositories
- A repository for data curation and reuse standards
- Increased use of research data in teaching at all levels
- Communities have voice in their own data – nothing about us without us
- Cross-community data findability
- Reciprocity across countries with data sharing and reuse
- Appropriate handling of sovereign data and data involving marginalized communities
- Agile and effective responses to non-traditional data
- Consortia connecting the many consortia associated with research data

Exemplars and Resources (Current State -- 20 illustrative innovations):

- *An Example:* WorldFAIR: CODATA coordinating with RDA through EU Horizon, surfacing cross-disciplinary issues to implement FAIR principles
- OSTP Memo on Ensuring Free, Immediate, and Equitable Access to Federally Funded Research

- OSTP Memo on National Institutes of Health Data Management and Sharing Policy
- *Science* on "Playing catch-up in building an open research commons"
- NIST Research Data Framework (RDaF)
- CERN/NASA Summit: "Accelerating the Adoption of Open Science"
- NASA Science Information Policy
- What Universities Owe Democracy JHU Book
- Industry Data for Society Partnership
- Shaping Europe's Digital Future
- Precisely Practicing Medicine with a Trillion Points of Data
- The Future of Science is Open
- Society of Research Administrators International (SRA)
- National Council of University Research Administrators (NCURA)
- Council on Governmental Relations (COGR): An Association of Research Institutions
- Carbon Call
- Trusted Cloud Principles
- MLCommons with People's Speech dataset and now Dollar Street
- Pfizer's Centers for Therapeutic Innovation
- Collaboration to accelerate data-driven discovery

Open Questions (Delta State):

- How best to ensure open questions accompany open data?
- How to foster data sharing and reuse with small grants?
- How best to foster public-private partnerships combining public and private data for open use?
- How best to carve out pre-competitive domains for the sharing and reuse of commercial data?
- How best to advance equitable access to research data across fields, disciplines, domains, and sectors?
- How to address the inconsistent schema and curation expertise in generalist repositories?
- How to address digital divides, such as internet connectivity in parts of rural and urban communities
- How best to reach researchers not currently aware of FAIR data and related matters?
- How to determine how long data should be preserved?
- How to foster succession planning for owners of research data with long-term value?
- How to ensure appropriate and effective licensing regimes for data – to assure data quality and to share curation/storage/distribution costs
- How to advance data for social good?
- How best to document impacts with data (not just use)?

NATIONAL ACADEMIES Sciences Engineering Medicine

Decarbonization Focus Groups: Data Needs for Decarbonization

Scope of Focus Group:

The U.S. has ambitious goals for decarbonizing its economy within the next two to three decades. Current landmark policies such as the Inflation Reduction Act and Bipartisan Infrastructure Law are changing the nature of corporate investments and household technology adoption decisions across multiple sectors (buildings, transport, manufacturing, electricity, land use and agriculture). Yet, pathways to net-zero emissions are characterized by numerous uncertainties at the federal, state, and local levels. This group explored many of these uncertainties, which result from gaps in data: socio-economic and demographic conditions, infrastructure quality and availability, technology readiness and market acceptance, ecosystem and environmental conditions, and so on.

Elements of a Success Vision (Desired State):

- Large amount of climate data that are openly accessible by the international community
- Integration of decarbonization data, which spans subsurface, surface and atmospheric systems, as well as contextual data and resources (human/infrastructure data, socio, economic date, etc.)
- Integration of decarbonization data across upper, middle, and lower income countries and regions
- There are widely utilized protocols for citations to complex, interconnected data sources
- Funders add language to grants and agreements to build a requirement that any data collected or used should be shared in a permanent location -- for research and practice of decarbonization

Exemplars and Resources (Current State -- 22 illustrative innovations):

- Under the Bipartisan Infrastructure Law (BIL) / Infrastructure Investment and Jobs Act (IIJA), the Carbon Management EDX4CCS is addressing many needs for Carbon Capture, Utilization, and Storage (CCUS) data users
- AGU Ethical Framework for Climate Intervention
- Humanitarian Data Exchange (HDX) on adaptation/response
- RDA Complex Citations Working Group
- ARM (Atmospheric Radiation Measurement) program put standards in place at the beginning of the program that supported design and development of data sources and models
- WorldFAIR: CODATA is coordinating this project with RDA through EU Horizon funding, intent to surface some cross-disciplinary issues to implement FAIR principles broadly and to get more accessible data into the pipeline

- NFDI: Large-scale national infrastructure project for research data sharing in Germany, across 30 different disciplines and use cases; outcomes open source to encourage global re-use
- Carbon Matchmaker
- H2 Matchmaker
- Interagency Working Group on Coal & Power Plant Communities & Economic Revitalization
- Open and shared standards like SDMX widely adopted.
- DOE's Public Access plan went live this spring 2023 that requires public curation of federally funded R&D products.
- EPRI non-profit energy research
- Organization pushing the open-source technology community and the open-science community together towards a common goal: https://opensource.science/open-source-science-white-paper-c4940a0
- Federal Register 2011, Request for Information: Public Access to Digital Data Resulting From Federally Funded Scientific Research (November 4, 2011) – This page contains the published document for the Request for Information offering the opportunity for interested individuals and organizations to provide recommendations on approaches for ensuring long-term stewardship and encouraging broad public access to unclassified digital data that result from federally funded scientific research.
- FAIR – The "FAIR Guiding Principles for scientific data management and stewardship," outlined here, were published in Scientific Data 2016. This set of principles provide guidelines to improve the Findability, Accessibility, Interoperability, and Reuse (FAIR) of digital assets.
- U.S. DOE Public Access Plan (June 2023) – The Department of Energy Public Access Plan (June 2023) describes how DOE-funded research and digital data will become more open and available to the public and how DOE will use persistent identifiers to help ensure scientific and research integrity. Building on the previous DOE Public Access Plan (July 2014).
- NITRD -- The Networking and Information Technology Research and Development (NITRD) Program is the Nation's primary source of federally funded research and development (R&D) in advanced information technologies (IT) in computing, networking, and software. NITRD is among the oldest and largest of formal Federal programs that coordinate the activities of multiple agencies to tackle multidisciplinary, multitechnology, and multisector R&D needs.
- World Bank effort (entirely closed-door)
- OASIS initiative

Open Questions (Delta State):

- How to connect data rich domains (e.g., climate) or organizations (e.g., federal) to less advanced data- domains or organizations (e.g., private sector, state/local, and developing countries?
- How to facilitate more open access to proprietary energy data?

- How to overcome the many data silos in this domain – by institution, sector, geography, etc.?
- How to handle complex citations to many interconnected data sources?
- How to advance shared or cross-walked ontologies – a semantic mapping is needed? Can automation be leveraged (e.g., ML)?

Disruption Focus Group:

Scope of Focus Group:

Disasters, extreme events, shocks, and systems transitions (natural, economic, health, and security) have large, complex, and long-lasting impacts. Disadvantaged communities that are already overburdened tend to be particularly affected, as they often suffer from acute and chronic stressors that compound each other. This group explored data on stressors that include, for example, energy and mobility burdens, safe drinking water access, employment opportunities, food deserts, public safety and healthcare, and school quality, among others.

Elements of a Success Vision (Desired State):

- Integrated, timely, and trusted disruption data, with predictive models and assessments of impacts
- Combined data on disruptions and socio-economic dynamics
- Energy risk data, with supporting dashboards, that encompass spatial and temporal scales
- Protocols guiding researchers, policy makers, institutional leaders, and citizen scientists in finding, accessing, combining data in interoperable ways
- University training prepares next generation scientists to work constructively and effectively with disruption data

Exemplars and Resources (Current State -- 11 illustrative innovations):

- The Belgian CRED EM-DAT data base is a source of risk impact data.
- Two data sources in disaster response:
  - CDC Social Vulnerability Index. Census data linked to other variables.
  - Global Health Security Index (GHSI) Hopkins risk data. Includes social vulnerability, economic status to identify vulnerability areas.
- Data Sharing for Disaster Response | LinkedIn
- MIT climate resiliency dashboard (for MIT campus) and another MIT example
- Example of data sharing project based on data standardization: https://www.linkedin.com/pulse/data-sharing-disaster-response-cameron-birge/?trackingId=BnQAx5fZTlCYiDN2nzVdfQ%3D%3D
- Operational readiness levels to help decide quality for use is being worked on. (Industry and government) Here is the example from Earth Science Information Partners (ESIP)

- Global population data set developers have been cooperating on what works for what purposes. Collaborate when it makes sense to do so. E.g., see https://www.popgrid.org
- ESIP and All Hazards info sources. All Hazards Consortium Have Geocollaborate that helps share. Precursor for broader ways to openly share. See: https://www.ahcusa.org/sise.html and https://frwg.geocollaborate.com/
- Consider Homomorphic encryption to preserve privacy
- Private sector has orders of magnitude better computing and access to high resolution Earth observations than academia – for example: Meta contributions to Humanitarian Data Exchange: https://dataforgood.facebook.com/dfg/docs/high-resolution-population-density-maps-demographic-estimates-documentation
- Water in UK is an example. It took a tax for an innovation fund and then the level of collaboration increased greatly

Open Questions (Delta State):

- How best to get good impact data across the many types of disruptive cases?
- How to deal with situations that are invariably fluid and complex - vulnerability can be temporary and multi-layered?
- How to address area with thinner data (e.g., socio-economic systems, environmental systems, infrastructure, etc.)?
- Even where relevant disruption data does exist, how to ensure standardization and equitable access?
- How to address privacy issues when dealing with disruption data that needs high levels of specificity?
- How best to address long range issue of displacement associated with displacement dynamics – people are moving, especially with refugees and climate change?
- How to get resources for data community collaboration? Getting cloud credits is relatively easy; how to overcome barriers such as data licensing, IP sharing, disciplinary silos, lack of trust, different needs, time frames, spatial scales? How to acquire the expertise to create useful public good outputs. How to bring in stakeholders? How to consider environmental justice and biases like English language? How to bring people together to cooperate?
- How to build disaster data curation and reuse into funding decisions?

Indigenous Data Governance Focus Group:

Scope of Focus Group:

Indigenous Data Sovereignty (IDSov) focuses on the protection of Indigenous rights and interests in the control and governance of Indigenous Peoples' data. Governance remains central to the realization of IDSov; Indigenous Peoples require data for the governance of their own nations and communities and also exercise their rights to govern their data. This group explored how US research data organizations implement the CARE Principles for Indigenous Data

Governance (Collective benefit, Authority to control, Responsibility, Ethics), if they support researchers and other institutions to do the same, and what they might need to operationalize CARE.

Elements of a Success Vision (Desired State):

- All research computing and data ecosystems understand and address the issues of data sovereignty with indigenous communities.
- Data Management Plans (DMPs) involving indigenous data appropriately anticipated data sovereignty.
- Indigenous communities have the resources to establish and sustain self-governed repositories for indigenous data.
- Full integration of FAIR and CARE principles.
- Connections among researchers working with indigenous data, combined with connections to indigenous communities.

Exemplars and Resources (Current State -- 5 illustrative innovations):

- ESIP worked with GIDA on operationalizing CARE principles for repositories
- Tribal data sovereignty is not a one-fits-all—data access and sharing is dependent on negotiated Data transfer, Use, and Ownership Agreements with the respective Tribal Nations (see current NIH DMP guidance)
- NIH-GREI initiative is focused on generalist repositories, with attention to specific community requirements
- Dataverse has a supported use case for ID community
- Jane Anderson, founding member of Local Context, and Steven McEarchern from Australia Social Science Data Archive started a working group, including members of GIDA

Open Questions (Delta State):

- How to support indigenous communities having their own data repositories?
- How to govern the use of AI to create synthetic indigenous genomic data sets?
- How to ensure that generalist repositories have appropriate policies and procedures for indigenous data?
- From the publishing perspective there is positive intent but we don't have comprehensive guidance. We emphasize care for heritage sites and impacts to human welfare and society. We also recognize data as a world heritage and encourage FAIR data but we don't integrate the two in our guidance and codes of conduct. That is work that needs to be done.
- How to address what are race-based protections of indigenous data with the recent SCOTUS ruling?
- Can the NIH Data Management Plan template be used as a potential model for researchers who would like to conduct research with indigenous communities responsibly?

- How to adapt the Global Code of Conduct for Research in Resource Poor Settings to also address the specific interests of indigenous peoples?
- How to expand what is in the Belmont Report to properly take into account indigenous data?

Justice, Equity Diversity and Inclusion Focus Group:

Scope of Focus Group:

This group explored how justice, equity, diversity and inclusion are being implemented, analyzed and evaluated in higher learning, organizations, government agencies, and the community at large. Participants discussed the development of policies improving and impacting JEDl considerations.

Elements of a Success Vision (Desired State):

- Research data work is increasingly attracting talent from HBCUs, TCUs, HSIs, and other MSIs.
- Bias in research data is consistently identified and addressed as part of standard research practices.
- Pathways into research data professional work begin in K-12 schooling and extend through undergraduate and graduate schools.

Exemplars and Resources (Current State -- 10 illustrative innovations):

- The Minority Serving Cyberinfrastructure Consortium is advancing research and educational computing, data and software infrastructure at HBCUs, TCUs, HSIs, and other MSIs
- Association for Women in Mathematics (AWM) wrote a statement on having the AWM Research Symposium in Georgia - at Clark Atlanta University
- DOE's website has detail on the DOE economic impact and Office of Economic Impact and Diversity | Department of Energy
- See the report to the NSF on the Missing Millions in research computing and data
- Ethical considerations are addressed in the EU AI Act, but the US doesn't have the same check point that EU does…we are creating technology and innovations that directly impact people
- Support and serve HBCUs to do data science research and education via the NSF-funded National Data Science Alliance (NDSA) through workshops, curriculum development workgroups, and research affinity cohorts
- An engagement plan for AI/ML researchers, practitioners, community partners, and other entities to collaborate with the NIH -funded AIM-AHEAD Consortium
- Dr. Justin Ballenger led a listening session with Atlanta University Center (AUC). From this listening session, faculty at Clark Atlanta are building a certification program in computer science/data science to help those who are already in the field.

- Desi Small-Rodriguez Sociologist. Demographer. Data Warrior. Relative.
- Identifying Assets and Collaborative Activities to Support Student Success in Environmental Data Science at Minority Serving Institutions

Open Questions (Delta State):

The bigger picture that we are operating in with the DEI legislation and the SCOTUS ruling, which are resulting in a contraction in support of DEI -- how will that impact attention to DEI with research data?

- Only 4% of data scientists are black – how to increase representation in the profession?
- How to ensure support, mentoring, and diverse leaders for the next generation of research data professionals, including those from underserved communities?
- Data science is collaborative work, but not every discipline is highly collaborative – what can be done about this broader culture change challenge?
- How to address disparities in venture capital funding for minority-led start-ups associated with research computing and data?
- How to address the talent shortages in K-12 education, particularly among individuals from underserved communities?