
# Minding the Gap:
## Leveraging Cyberinfrastructure to Transform EPSCoR Jurisdictions

Venice Bayrd
Montana State University
venice.bayrd@montana.edu

Gwen Jacobs
University of Hawai'i System
gweni@hawaii.edu

Patrick Schmitz
Semper Cogito Consulting
patrick@sempercogito.com

Scotty Strachan
Nevada System of Higher Education
sstrachan@nshe.nevada.edu

Patrick J. Clemins
University of Vermont
patrick.clemins@uvm.edu

Frederick C. Harris, Jr.
University of Nevada, Reno
fred.harris@unr.edu

## Acknowledgements

We would like to thank the early supporters who helped us propose this idea, and all of the members of the CaRCC EPSCoR CI Working Group for their input, support, and review of this report: Bastian Bentlage (Univ. of Guam), Dana Brunson (Internet2), Ragan Callaway (Montana NSF EPSCoR), Deborah Dent (Jackson State Univ.), Terry Donaldson (Guam EPSCoR), Brent Ewers (Wyoming EPSCoR), William Garrick (Portland State Univ.), Doug Jennewein (Arizona State Univ.), Todd Kipfer (Montana NSF EPSCoR) Sarah Mannheimer (Montana State Univ.), William "Bill" K. Michener (New Mexico EPSCoR), Claire Mizumoto (Univ. California San Diego), Shane Moeykens (Univ. Maine), Mandy Terrill (Oregon Health and Science Univ.), Emery Three Irons (Little Big Horn College), and Alice "Pips" Veazey (Univ. Maine). We are grateful to the many participants who shared their time and ideas in the workshops. We would also like to acknowledge the input and support of the CaRCC RCD Capabilities Model Working Group members. The CaRCC RCD Capabilities Model Working Group (RCD Capabilities Model Working Group) partnered closely with the EPSCoR CI Working Group to enable assessments and later analysis.

This work is supported in part by National Science Foundation (OIA) Award 2033483 (PI: Gwen Jacobs), Award 2033514 (PI: Venice Bayrd), and Award 2033519 (PI: Scotty Strachan) "Collaborative Research: Building Research Cyberinfrastructure in EPSCoR Jurisdictions: Assessment, Planning and Partnerships;" by an NSF RCN grant (OAC-1620695, PI: Alex Feltus, "RCN: Advancing Research and Education through a national network of campus research computing infrastructures – The CaRC Consortium"), and by an NSF Cyberinfrastructure Centers of Excellence (CI CoE) pilot award OAC-2100003 (PI: Dana Brunson), "Advancing Research Computing and Data: Strategic Tools, Practices, and Professional Development." Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

## Preferred Citation

Bayrd, V., Jacobs, G., Schmitz, P., Strachan, S., Clemins, P. & Harris, F. C. 2023. *Minding the Gap: Leveraging Cyberinfrastructure to Transform EPSCoR Jurisdictions*. NSF EPSCoR Workshop Report. https://doi.org/10.5281/zenodo.8097562

# Table of Contents

# Executive Summary

Catalyzing research competitiveness is central to the National Science Foundation Established Program to Stimulate Competitive Research (NSF EPSCoR) mission. Modern research and education rely on cyberinfrastructure (CI): networks, data, computers, software, and the people who support them. However, the rapid rate of technological change poses significant challenges to researchers and traditional institutional Information Technology (IT) organizations. Any lack of access to research-facing professional CI support results in researchers making difficult time management decisions between focusing on their science or solving the associated technology integration problems. In contrast, access to robust CI support can translate to improvements in areas including science scalability, reproducibility, interoperability, research impact, and security, thereby accelerating competitiveness.

Historical CI underinvestment in NSF EPSCoR jurisdictions (Hill, 2012) has created degrees of unevenness in the national distribution and availability of CI resources (Hill, 2012; Blatecky, *et al*. 2019), impacting competitiveness across all NSF funding areas. To address the gaps, Blatecky, *et al*. have suggested that the CI ecosystem in the United States will require increasing coordination, collaboration, and planning across multiple federal, state, institutional, and public/private boundaries for efficient, effective, and equitable access (2019); our findings, and in a broad sense, those of the Committee on the Future of NSF EPSCoR (2022) concur with that suggestion. New NSF EPSCoR E-CORE and E-RISE solicitations offer opportunity for improvement but also increased risk of perpetuating inequitable gaps in research technology access and capability. Cybersecurity, artificial intelligence, high performance computing, and data management are CI research areas of increasing national priority (see, for instance the "Key Technology Focus Areas" in the *Request for Information (RFI) on Developing a Roadmap for the Directorate for Technology, Innovation, and Partnerships at the National Science Foundation* ([*s.n.*], 2023)). As illustrated in this report, the gaps in EPSCoR cyberinfrastructure in data-centric areas alone suggest that EPSCoR institutions are unlikely to be competitive in the national research arena without purposeful companion efforts to improve already lagging CI.

The EPSCoR CI Working Group (EPSCoR CI Working Group) formed with funding support from NSF EPSCoR via a collaborative workshop award for the dual purpose of responding to the lack of data characterizing EPSCoR CI capabilities, and to a "Call to Action" to increase CI access across EPSCoR (Moore, 2019). The EPSCoR CI Working Group facilitated community generation of unprecedented baseline CI capabilities data characterizing CI distribution and availability across EPSCoR jurisdictions and institutions (Schmitz, *et al*., 2022). Those data then framed community workshop discussions focused on prioritizing identified gaps and generating solutions.

This report summarizes the results of community data generation activities and workshop discussions, ending with eleven recommendations that may be best addressed by collaboration across scales:

| Recommendations to NSF: |
| --- |
| • Re-establish CI as a required RII core component<br>• Establish an NSF EPSCoR CI Council<br>• Investigate models of CI human resources capacity-sharing for EPSCoR<br>• Enhance collaborative partnerships between EPSCOR, OAC, and TIP<br>• Incentivize proposal-stage participation by technical/CI staff |
| **Recommendations to Jurisdictions and Institutions:** |
| • Formalize CI assessment and planning<br>• Coordinate CI development across RII projects and jurisdictions<br>• Integrate regional network organizations<br>• Align foundational IT<br>• Measure CI impacts<br>• Communicate the role of CI |

**Table 1: Workshop recommendations to NSF and to EPSCoR Jurisdictions and Institutions**

The remainder of the report expands upon the context for this work, the nature of the workshop structure and summaries of discussions, and the rationale for each of the eleven recommendations.

# 1. Introduction

Cyberinfrastructure is central to national and international goals for Science, Technology, Engineering, and Mathematics (STEM) research and development. For instance, the vision sustained by the National Science Foundation (NSF), the European Open Science Cloud, and others for a federated virtual environment that facilitates international interdisciplinary open science collaborations (NSF, 2019; DGRI-EC, 2016) will require intensive reconceptualization of cyberinfrastructure (CI) resource, service, and provider frameworks. In the United States, the term cyberinfrastructure has grown to encompass a "complex adaptive system" or "ecosystem" of information technology, data tools and services, organizations, instruments, and related expertise (Borgman, 2015; Moore, 2019; Blatecky, *et al*, 2019), all of which are essential to realizing the above vision. Note that the cyberinfrastructure ecosystem outlined here extends far beyond traditional "high performance computing" (HPC) infrastructure. Research cyberinfrastructure as used in this report refers to the expertise, computation, networking, software, data, and related capabilities essential to supporting modern research, with a lesser focus on instrumentation and organizations that are also part of the NSF Cyberinfrastructure Ecosystem (*cf*., *e.g*., Moore, 2019). Research cyberinfrastructure, research technology, and research computing and data (RCD) are used interchangeably in this report.

To participate in modern NSF-supported research, with its focus on convergent topics, international interdisciplinary open science collaborations, broadening participation, and the NSF Big Ideas, researchers need to operate both on the computational desktop and beyond, on nationally networked cyberinfrastructure resources. Modern workflows are necessitating more interoperable interdisciplinary data, more collaboration, more security, more memory-, storage-, and computation-intensive processing, more automation, and more complex software. This means that researchers must migrate from small, self-service technology use models into the unfamiliar territory of shared computation and data environments, which are still evolving in terms of usability, access, and flexibility. If modern technology systems and their associated support professionals are unable to meet researchers "where they are" technically, then time-to-science (*i.e*., research velocity, hypothesis-to-publication) suffers dramatically and the research itself is exposed to risk (*e.g*., risk of being "scooped", or having minimized impact). Resources invested in building effective relationships between researchers and CI professionals can ease the opportunity cost of developing technological solutions; can create more impactful, scalable, reproducible, accessible, interoperable workflows; and can open up new lines of funding opportunity to further evolve both the science and the CI. EPSCoR jurisdictions find themselves particularly challenged in these areas.

In 2019, Dr. Loretta Moore, in her role as NSF OIA Section Head, gave a keynote address at the Internet2 National Research Platform (NRP) conference that concluded in part with a call to action to increase access to CI in EPSCoR jurisdictions. While the disparity in CI access was recognized, CI capabilities had not previously been characterized using a common framework across jurisdictions. The EPSCoR CI Working Group (EPSCoR CI Working Group) formed with funding support from NSF EPSCoR via a collaborative workshop award ((OIA) Award 2033483 (PI: Gwen Jacobs), Award 2033514 (PI: Venice Bayrd), and Award 2033519 (PI: Scotty Strachan) "Collaborative Research: Building Research Cyberinfrastructure in EPSCoR Jurisdictions: Assessment, Planning and Partnerships") for the dual purpose of responding to the lack of data characterizing EPSCoR CI capabilities and to Dr. Moore's "Call to Action" to increase CI access across EPSCoR (Moore, 2019). The EPSCoR CI Working Group facilitated community generation of original baseline CI capabilities data characterizing CI distribution and availability across EPSCoR jurisdictions and institutions, and those data have been presented and discussed in aggregate in a separate report by Schmitz, *et al*. (2022). Those data formed a pivot point from which the community then shifted to focus on workshop discussions prioritizing identified gaps and generating solutions. To address CI gaps, Blatecky, *et al*. have previously suggested that the CI system in the United States will require increasing coordination, collaboration, and planning across multiple federal, state, institutional, and public/private boundaries for efficient, effective, and equitable access (2019); our findings, and in a broad sense, those of the

Committee on the Future of NSF EPSCoR (2022) concur with that suggestion. We propose 11 multi-scale recommendations to address the gaps.

This report summarizes the results of the community assessment data generation activities and workshop discussions, ending with eleven multi-scale recommendations that address thematic community needs. The baseline assessment data generated as a result of the EPSCoR Cyberinfrastructure Workshop Series and related workshop activities demonstrate that the CI capabilities imperative to supporting modern research at EPSCoR institutions are often under-supplied, unsustainable, or in some cases, non-existent. The EPSCoR CI Working Group identified five cross-cutting themes that emerged from community discussion of needs during the workshop series. The five themes are as follows:

- **Theme 1: Foundational IT support:** Foundational IT must be in place to support modern research.
- **Theme 2: CI-Research mission alignment:** Administrative, operational, and resource models must align to support the research mission.
- **Theme 3: Engagement at multiple levels:** Institutional CI professionals must be supported to engage the broader community as well as the local research needs.
- **Theme 4: Workforce development:** Experiential learning in CI professional roles must be based on best practices to be an effective Workforce Development Pathway.
- **Theme 5: CI as human capital:** CI professionals with their technical and facilitation/liaison expertise form the key component in CI capital investment.

Despite a vision of CI as central to science and engineering research (*e.g.*, NSF, 2019), a common community observation that inflects the emergent themes from the EPSCoR CI workshop series is the tension and resource competition between "the science" and "the technology" in NSF EPSCoR Research Infrastructure Improvement (RII) proposals. Fundamentally, this tension needs to be addressed in part by strengthening trust and collaborative relationships between science and cyberinfrastructure teams, which positions both to be more successful in future non-EPSCoR proposals. Recent changes to the EPSCoR program hold potential to address this issue, and at the same time prepare jurisdictions for immediate and near-future federal priorities. The CHIPS and Science Act (Congress.gov, 2022) was signed into public law (Pub L. 117-167) to prioritize not only science and technology research and development, but *broad, geographically diverse* opportunities; more directly, it will fund CI-related initiatives sponsored through the new Technology, Innovation, and Partnerships (TIP) Directorate (*cf*. "Request for Information…" issued by NSF on April 28, 2023). The August 2022 "Nelson Memo" from the Office of Science and Technology Policy (OSTP) champions open, reproducible science and findable, accessible, interoperable, and reusable (FAIR) data *accessible to all*, as a cornerstone of economic development (Nelson, 2022). "The Missing Millions" report (Blatecky, *et al*., 2021) highlights the need for racial, ethnic, and gender representation equity in technology. May 2023 EPSCoR RII program revisions are well-positioned to take action on cyberinfrastructure issues of concern to EPSCoR jurisdictions, aligning the program with White House Executive Office priorities, national diversity, equity, and inclusion (DEI) priorities, and new CI-related NSF initiatives. The remainder of this report summarizes workshop activities and details concrete recommendations and action items that can be initiated across the multitude of interleaved scales inherent to these challenges.

## 2. Assessing EPSCoR Research CI Capabilities

The EPSCoR CI Working Group facilitated outreach and numerous EPSCoR community engagement events resulting in a current baseline assessment dataset documenting EPSCoR CI availability and distribution (*cf.* Schmitz, et al., 2022). Close partnership between the EPSCoR CI Working Group and the CaRCC RCD Capabilities Model Working Group enabled EPSCoR institutions to leverage the RCD Capabilities Model assessment tool to generate data. Using a common assessment framework has allowed for cross-comparisons, benchmarking, and common language development; all of these contribute to laying the foundation for future collaborative work both within and across jurisdictions.

**About the RCD Capabilities Model Assessment Tool:** The RCD Capabilities Model has been supported in part by NSF-funded efforts (OAC-1620695 and OAC-2100003), and was designed to allow institutions to assess their current capabilities and provide structured input for strategic decision-making using a shared community vocabulary (Schmitz, *et al.* 2020). The model presents Likert-scale rank-ordered questions that address roughly 150 capabilities. Each capability is further assessed along three axes: *Deployment at Institution*; *Multi-Institution Collaboration*; and *Service Operating / Support Level*. Questions are organized around five "Facings"[1] that are increasingly used in the CI community as a means of characterizing the roles of people who support CI/RCD:

- Researcher-Facing
- Data-Facing
- Software-Facing
- Systems-Facing
- Strategy and Policy-Facing

The Assessment Tool also allows institutions to mark specific capabilities as priorities. The resulting national assessment dataset provides important insights into the general state of support for CI, at both summary and granular levels (Schmitz, 2021). The companion EPSCoR CI dataset and analysis explores in-detail how EPSCoR and non-EPSCoR jurisdictions' capabilities compare, and highlights the gaps (Schmitz, *et al.*, 2022). A brief summary of the data and analysis is provided here.

**About the baseline dataset:** A total of 15 EPSCoR institutions completed RCD Capabilities Model assessments in 2020 and 2021, and a 16th contributed data in 2022. Non-EPSCoR assessments contributed in 2020-21 numbered 36, and four more contributed data in 2022. The EPSCoR institutions represent 12 states and U.S. territories, and include a broad mix of Carnegie Classifications and six (*i.e.*, 40%) minority-serving institutions. For more information on the breakdown of workshop statistics and participating institutions, please see Appendix 1.

Some highlights of what the data indicate include:

- EPSCoR institutions lag their non-EPSCoR peers in CI/RCD capabilities across the board, with the widest gaps in the Data-Facing and Researcher-Facing areas.
- EPSCoR institutions provide services at a less robust operational level than non-EPSCoR institutions, and are less able to provide services to all researchers across the institution.

[1] For further reading, See also: RCD Professionalization and Facings – carcc.org and Capabilities Model Introduction and Guide to Use.

- The top priorities of the contributing EPSCoR institutions are mostly in Data-Facing, Software-Facing, and Researcher-Facing topics, although the fourth-highest priority overall is the need for more strategic planning.

There is considerable variation (as evidenced by the standard deviation bars) in the areas of relative strength and weakness among the contributing institutions (*cf.* Fig. 1); this variation is also evident within institutions (Schmitz, *et al.*, 2022). An area for future research would be to investigate how inter- and intra-institutional variance is driven by intrinsic factors like institutional priorities or legacies of RCD support versus extrinsic factors such as funding, economic milieu, and/or political and regulatory context.

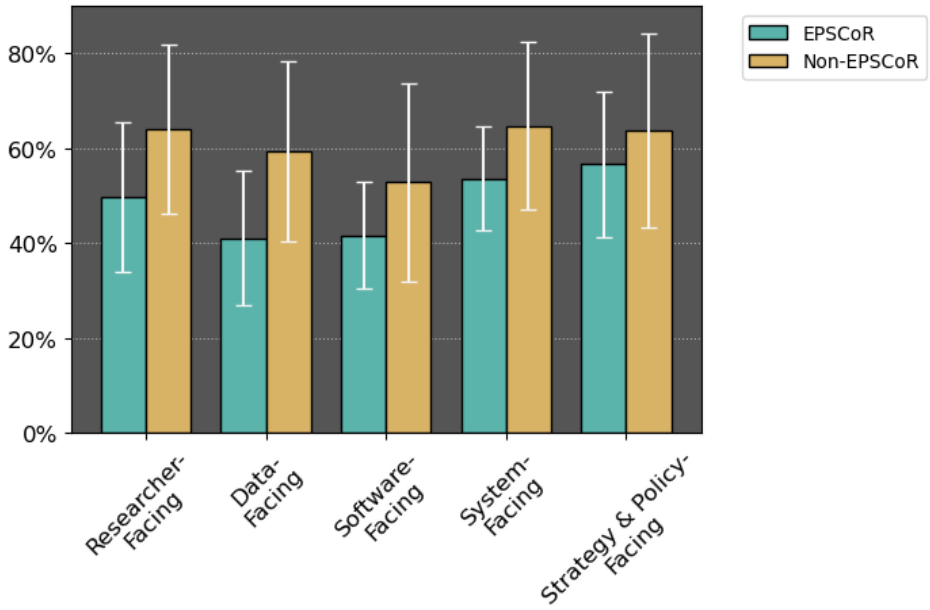


**Figure 1: Coverage by Facing for EPSCoR and non-EPSCoR Institutions**

Some of the widest gaps between EPSCoR and non-EPSCoR institutions are seen in the Data-Facing Areas (see Fig. 2). EPSCoR institutions reported starkly lower capabilities coverage in the areas of *Data Discovery and Collection*; *Data Analysis*; *Data Visualization*; and particularly, support for *Security/Sensitive Data*. Cybersecurity, artificial intelligence, high performance computing, and data management are of increasing national priority (see, for instance the "Key technology Focus Areas" in the *Request for Information (RFI) on Developing a Roadmap for the Directorate for Technology, Innovation, and Partnerships at the National Science Foundation* ([*s.n.*], 2023). The gaps in infrastructure support in data-centric areas suggest that EPSCoR institutions are unlikely to be competitive in the national research arena without companion efforts to improve already lagging cyberinfrastructure. Community-identified solutions discussed later in this report imply that any efforts toward CI improvement will be most successful using a comprehensive approach that considers the full context in which CI operates.

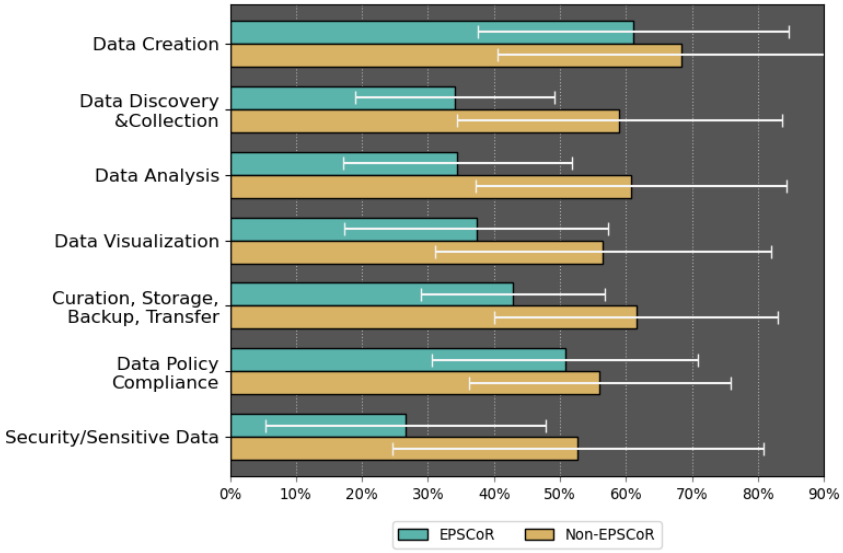


**Figure 2: Coverage of Data-Facing topics for EPSCoR and non-EPSCoR Institutions**

In addition to the gaps seen in the Data Facing Areas, another broad gap seen in the national dataset lies between EPSCoR institutions and non-EPSCoR institutions in the average *Service Operating Level*. The *Service Operating Level* is a facet of the capabilities' total coverage values that characterizes the robustness of service support for each area or topic. EPSCoR institutions reported an average level between "*Substantial Risk of Failure(s)*" and "*Lights on Only*" where non-EPSCoR institutions were on average well above "*Lights on Only*" towards a "*Basic/Economy*" level. Just as for the total coverage values, the gap is widest for Researcher-Facing and Data-Facing topics, indicating that these areas are the most challenging for EPSCoR institutions to operate in a sustainable manner[2]. These assessment values also add evidence to the assertion that more targeted efforts are needed to change existing CI-access dynamics.

The EPSCoR CI dataset shows clear gaps in coverage relative to peers in non-EPSCoR jurisdictions. It is worth noting that beyond the examples outlined above, **some institutions that began an assessment reported that they had no coverage across so many areas that it was too discouraging to continue, and dropped the RCD Capabilities Model assessment work. If these institutions were represented in the data, the gaps we describe between EPSCoR and non-EPSCoR institutions would be even more pronounced.** Nonetheless, the EPSCoR CI baseline dataset corroborates with the "lived experience" of the EPSCoR CI professionals who engaged in the workshops. The baseline dataset also provides a detailed framework for discussions on how to pivot from this point toward improved CI access in support of research. Readers are encouraged to read the full EPSCoR CI baseline dataset report (*cf.* Schmitz, *et al.*, 2022) documenting the gaps as a companion to this workshop report, which focuses on communicating community needs (see *Themes* (Section 3)) and generating potential solutions (see *Recommendations to NSF EPSCoR* (Section 4) and *Recommendations to Jurisdictions and Institutions* (Section 5)).

[2] We note that these values echo the 2003 blue-ribbon report assertion that "Research competitiveness through the foundations of facilities and technical supporting staff is more fully established and better funded in non-NSF EPSCoR jurisdictions" (Atkins), as well as that of Knepper and Börner (2016), who found a lack of participation in the XSEDE network in EPSCoR states, both of which underscore insights from *The Missing Millions* report that indicate (in Finding 2) that "Accessibility = Access + Ability" (Blatecky, *et al.*, 2021).

# 3. Workshop Attendance, Discussion, and Resulting Themes

**Purpose:** The EPSCoR CI Workshop Series comprised a set of targeted events that each had specific objectives. The Spring Workshop took place online on two separate days, and was designed to identify community priorities and analytical gaps to be addressed during the Fall Workshop in support of the broader objectives of increasing access to CI in EPSCoR jurisdictions. The Fall Workshop was designed to present and review the EPSCoR RCD Capabilities Model assessment data analysis; to ask participants to generate ideas around using RCD Capabilities Model results and analysis in campus/regional/national planning and partnerships; to discuss community needs (Themes) around CI access; and to help people connect to each other and to ongoing CI work through CaRCC, the RCD Nexus, and other avenues. What follows is an overview of workshop attendance, discussions, and the resulting community-identified needs, coalesced into five emergent Themes.

## Attendance and Jurisdictional Representation Statistics

The Spring and Fall Workshops were well attended, with representation from 20 jurisdictions at the online Spring Workshop, and 17 jurisdictions at the in-person Fall Workshop (see Appendix 1 for more detail).

Spring Workshop

- 47 Total attendees, 41 of whom were from EPSCoR jurisdictions
- 20 jurisdictions were represented: Alabama, Arkansas, Delaware, Hawaii, Idaho, Iowa, Kentucky, Louisiana, Maine, Montana, Nevada, North Dakota, Oklahoma, Rhode Island, South Carolina, South Dakota, U.S. Virgin Islands, Vermont, West Virginia, Wyoming
- One regional Research and Education Network (REN) was represented: Great Plains Network (serving seven EPSCoR jurisdictions)

Fall Workshop

- 45 Total attendees, 42 of whom were from EPSCoR jurisdictions
- 17 EPSCoR jurisdictions were represented: Alaska, Arkansas, Guam, Hawaii, Idaho, Kansas, Louisiana, Maine, Mississippi, Montana, Nevada, North Dakota, South Carolina, South Dakota, U.S. Virgin Islands, Vermont, Wyoming
- Three NSF program personnel were present

## Fall Workshop Panel Discussion: Collaborations at Regional, Jurisdictional, and Institutional Levels

During the Fall Workshop, workshop organizers convened a panel to discuss strategies, successes, and challenges for establishing CI partnerships at the regional, jurisdictional, and institutional levels. The discussion highlighted the importance of CI strategic plans across scales; the role of established relationships in building on emerging CI partnerships; examples of key systematic and historical challenges to coordinating and building lasting research technology infrastructure; success stories of CI intentionality in jurisdiction EPSCoR projects; and opportunities for transforming science and workforce development using CI. These stories of successes and challenges illustrate what can happen when we are intentional about building partnerships across scales, setting data-driven infrastructure priorities, and recognizing roadblocks to address. Brief overviews of each panelist's stories are presented in Table 2.

### Gwen Jacobs – Hawaii / regional collaborations

Prior to 2013, the University of Hawai'i System (UH) had no centralized support for cyberinfrastructure with the exception of long-standing investments in international, statewide and inter-island R&E networking infrastructure, critical for connecting Hawai'i to high-speed networks to support education and research. In 2013 the UH President made an intentional investment in CI by establishing a Director of CI and a CI team within UH Information Technology Services. Ten years later, the CI team has grown into an indispensable component of the research mission of UH, supporting researchers across all 10 campuses with HPC, data storage, management and dissemination, software engineering, networking and training. These efforts are supported by multiple NSF awards led by the CI team in CI professional development (ACI-REF #1341935), software frameworks (AGAVE #1450413, Tapis #1931575), HPC/storage (MRI #19120304, CC* Compute Koa #2201428, CC* KoaStore #2232862), academic cloud computing (Jetstream2, #2005506), and CI skills training (CITRACS, #2118222).

Four factors contribute to the Hawai'i model for a self-sustaining CI enterprise. 1) The UH President and UH CIO have a deep understanding of the importance of CI for supporting the research mission of the university and the Director of CI functions in an executive leadership role. 2) University investments prioritize funding for human capital over hardware, including granting PI status to CI professionals. 3) Recent EPSCoR RII Track-1 awards ('Ike Wai, #1557349 and ChangeHI #2149133) include CI as a core research and infrastructure component and invest funding for CI professionals for salary support, professional development and career advancement. 4) To build competitiveness, the CI team has leveraged multiple national collaborations with top-tier CI experts at the Texas Advanced Computing Center, with members of the NSF Jetstream consortium, and with national partners focused on CI Professionals workforce development.

### Scotty Strachan - Nevada / jurisdictional collaborations

Over the last 20 years, Nevada is a good representative of systemic lack of CI investment, awareness, and use across the jurisdiction. Nevada's higher education system experiences minimal use of NSF XSEDE resources, does not possess established institutional-scale research computing teams, was one of the last states to receive an NSF OAC-CC* infrastructure grant, and only has one institution with a written CI Plan. CI investments by past EPSCoR Track-1 projects (McMahon et al., 2011; Mensing et al., 2013; Le et al., 2015; Scully-Allison et al., 2018; Devitt et al., 2022) have not been sustained or supported by institutional leadership beyond NSF funding windows.

Recent developments in Nevada to assess and organize scattered RCD support efforts are emerging from the central System Computing Services (SCS) office in Nevada's System of Higher Education (NSHE). In 2022 NSHE-SCS established a new mission area with two professional positions to explicitly engage in research and education engineering at the State level. One of these positions is also the CI lead and co-PI for Nevada's NSF EPSCoR Track-1project, creating an opportunity to coordinate CI efforts and leveraging the state R&E network managed by SCS (NevadaNet). Early-stage objectives within the Track-1 project include: facilitation of campus-level CI surveys, benchmarking of current Nevada CI capabilities, RCD professional hires at two campuses, formation of a state-wide CI working group, creation of a prototype statewide Science DMZ network, expansion of the CI vision in the state Science and Technology Plan, and initialization of a data commons.

These efforts to break down socio-political barriers and coordinate technology remain very grass-roots and project-driven, however, as the research-intensive campuses and institutes have yet to formalize institutional scale CI planning and commitment (roles, resources, timelines). Nevada has experienced high "star faculty" and staff turnover rates, as expectations and reality frequently collide in the

| |
|---|
| technology space. The EPSCoR activities are intended to serve as a catalyst to spurring formal, high-level actions in these areas, and will benefit strongly from clearer blueprints, recommendations, and requirements for CI coordination, planning, and sustainability for the jurisdiction. |
| Venice Bayrd - Montana / institutional to jurisdictional collaborations |
| At Montana State University (MSU), Research Cyberinfrastructure (RCI) is a core facility supported in part by University Information Technology (UIT) and the Office of the Vice President for Research and Economic Development (VPRED). MSU RCI is primarily focused on HPC, high-volume low-latency storage and data transfer, virtual server hosting, and research-specific IT support. Other aspects of Data- and Researcher-Facing CI are supported by the Library, including data management and data curation. Data visualization capabilities are represented by multiple units on campus, including disciplinary ones, and like many MSU enterprise IT services, are currently decentralized. Despite historically being loosely coupled and partnering largely on a project-driven basis, as needed, good will exists amongst the various CI providers. In addition, a recent Research Alliance ([Montana State University], 2023) initiative is helping to bring research CI services more formally together with research administration and faculty professional development, facilitating boundary-spanning outreach.<br><br>The Library and RCI, with an already established and collegial relationship, agreed to undertake joint completion of the RCD Capabilities Model assessment in 2020. Despite some setbacks related to staff turnover, the process was notably beneficial beyond the initial goal of completing the assessment, and resulted in goals to generate a shared research CI service portfolio and to develop a collaborative CI strategic plan.<br><br>MSU's assessment work has also led to broader intra-jurisdictional conversations, facilitating inclusion of CI in the state S&T plan; proposing shared CI assessment, planning, and service provision (*e.g.*, an inter-institutional project-focused data commons); and including collaborative CI as an element of the upcoming 2023-2028 Track-1 project. Montana EPSCoR has supported CI capacity-building both with the current Track-1 project (CREWS; Award #1757351) which supported the hire of a project data manager; and with the subsequent Track-1 project (SMART FireS; NSF EPSCoR Award #2242802), which will support collaborative jurisdictional CI. |

**Table 2: Panel reviews of collaborations at Regional, Jurisdictional, and Institutional Levels**

The use cases presented in the workshop panel made clear that cyberinfrastructure capabilities and collaborations for entire EPSCoR jurisdictions remain very sensitive to multiple factors: 1) the overarching political, regulatory, and financial milieu; 2) funding agency solicitation opportunities and requirements; 3) basic institutional infrastructure, existing relationships, and decision-making processes (and lack thereof); and 4) individual personalities. **Considered in total, the workshop panel use cases and related participant discussions surfaced the need for multi-scale, integrated, and innovative approaches to resolving the historical lack of investment in EPSCoR cyberinfrastructure.** In the following *Themes* sub-section, we discuss a series of needs that emerged from workshop discussions. The themes serve as the basis for recommendations and action items to improve CI capabilities, access, and collaborations.

## Emergent Themes and Related Community Needs

Our community-driven workshop activities have exposed a set of cross-cutting themes that represent EPSCoR CI community needs. The themes form the basis of specific recommendations to NSF (Section 4) and jurisdictions and institutions (Section 5), as well as forming the basis of action items for the EPSCoR CI community (Section 6). Table 3 provides a matrix linking the five emergent themes to the respective recommendations that can most directly help shift the dynamic on the issues.

| | Themes | | | | |
|---|---|---|---|---|---|
| | **Foundational IT Support** | **CI-Research Mission Alignment** | **CI Engagement at multiple levels** | **Workforce Development** | **CI as Human Capital** |
| **Recommendations to the NSF:** | | | | | |
| Re-establish CI as a required RII core component | X | X | X | X | X |
| Establish a CI Council | | | X | X | X |
| Investigate models of CI human resources capacity-sharing for EPSCoR | X | | X | X | X |
| Expand EPSCoR-OAC-TIP collaboration | X | | X | | X |
| Incentivize proposal-stage participation by technical/CI staff | X | | X | X | X |
| | **Themes** | | | | |
| | **Foundational IT Support** | **CI-Research Mission Alignment** | **CI Engagement at multiple levels** | **Workforce Development** | **CI as Human Capital** |
| **Recommendations to EPSCoR Jurisdictions:** | | | | | |
| Formalize assessment & planning | X | X | X | X | X |
| Coordinate jurisdictional to regional CI development | | X | X | X | X |
| Integrate regional network organizations | X | | X | | |
| Align foundational IT | X | X | | | X |
| Measure CI impacts | | X | X | | |
| Communicate the role of CI | | X | X | | |

**Table 3: Recommendations to NSF and Jurisdictions referenced to emergent themes and actionable solutions from the workshop discussions**

**Themes:**

- **Theme 1: Foundational IT Support: Foundational IT must be in place to support modern research**: Community observations suggest that when EPSCoR institutions invest less in central academic IT services, it can result in a reduced focus on research, as researchers spend time addressing basic technological issues. EPSCoR institutions need access to sufficient, reliable, secure foundational IT, inclusive of both physical and human capital, as well as strategic planning capacity in order to develop research support and address higher-order CI priorities.

  Data security and sensitive data: This sub-theme emerged strongly in both the Spring and Fall Workshops. In Spring it was ranked as the top issue in a poll of current concerns. When paired with the assessment data from Figure 2 showing how much EPSCoR institutions are lagging in areas of "Security/Sensitive Data", workshop participants' emphasis on this topic demonstrates a significant area of concern and need. The increasing regulation of sensitive data assets will only amplify this challenge. Triple layers of compliance – at funding agency level, enterprise level, and research group level – make compliance even more challenging. Collaborations across institutions become increasingly difficult in this environment, particularly when some institutions have policies in place and others do not.

- **Theme 2: CI-Research Mission Alignment: Administrative, operational, and resource models must align to support the research mission:** There is a need in the EPSCoR research CI community for integrated CI and research mission alignment and agility that can help research CI survive personnel turnover. The community noted examples such as administrative posture, perceived mission, research technology awareness by different levels of leadership, resource priorities, documentation of partnerships and collaborations, (lack of) policy, turnover, and unestablished communications channels at institutional and jurisdiction levels as barriers to effective development of CI and CI collaborations for research and education. As one breakout group noted:

  > *"Administrators leave and often leave a big gap when they do. But when single-person shop or small shop grassroots RCD people leave, it also can have major ripple effects."*

- **Theme 3: Engagement at multiple levels: Institutional CI professionals must be supported to engage the broader CI community as well as local research needs:** CI professionals need time for jurisdictional, regional, and national community engagement as part of their professional development plan in order to stay apprised of evolving best practices and opportunities, and to support unfunded multi-scale collaborations that support service provision. CI best practices and approaches are constantly evolving in the national landscape, but awareness, implementation, and contribution by EPSCoR jurisdictions is often ad hoc. Beyond professional development benefits, serendipitous benefits may result from increased engagement. One workshop participant suggested:

  > *One possible reason that EPSCoR CI groups are more likely to have success with external funding for CI might be that they're more likely to have statewide CI collaboration groups, so they're more likely to get peer mentoring within their jurisdiction.*
  >
  > - An EPSCoR R1 participant

- **Theme 4: Workforce Development: Experiential learning in CI professional roles must be based on best practices to be an effective Workforce Development Pathway:** There is a need to add structure to often *ad hoc*, unstructured interdisciplinary CI training opportunities.

  Students and postdoctoral researchers are frequently leveraged as human capital in CI, developing their technical and service skills as well as relating their research areas to technology

infrastructure. However, training is often *ad hoc* and unstructured, and the community has expressed a need for models and best practices documentation. Some graduate students choose not to pursue a faculty appointment but still want to stay engaged with research; exposure to CI professional roles, national communities, and interdisciplinary cross-training enables this and contributes to modern STEM workforce development in multiple sectors. Providing structured opportunities for students and postdoctoral researchers to engage in CI and cross-sector partnerships is a need that also aligns with the Economic Development focus area Findings and Recommendations made by the Committee on the Future of NSF EPSCoR in their recent report (2022).

- **Theme 5: CI as Human Capital: Investment in CI as Human Capital is critical**: CI professionals with their technical and facilitation/liaison expertise form the key component in CI capital investment. The need to create new permanent positions for CI professionals in EPSCoR jurisdictions was heard throughout workshop engagements. As one participant said:

  > *One of our major challenges is staffing levels. Beyond the usual recruiting difficulties, EPSCOR institutions frequently have small staffs and difficulty covering all the needed expertise. I would be interested in discussions and examples of banding together to form composite teams.*
  >
  > - An EPSCoR R1 participant

  Funding for EPSCoR CI positions is often short-term and unsustainable. Nationally, CI career tracks are becoming common, but many EPSCoR institutions do not have critical support team mass or funding to recruit/retain through base funding channels. Additionally, some CI professionals cannot gain PI status because of their non-faculty roles, making sponsored funding out of reach. Even for those who have PI status, sponsored funding opportunities can be out of reach. For example, EPSCoR institutions are often not prepared to act and/or are lacking the robust internal communications pathways that enable PIs to propose or commit to institutionally aligned, base-funded position sustainability within the three-to-six-month response-time windows of extant solicitation opportunities (*e.g.*, the recent Strengthening the Cyberinfrastructure Professionals Ecosystem (SCIPE) Program Solicitation (NSF 23-521)). PIs operating within EPSCoR institutions may further lack the institutional infrastructure and communications frameworks to rapidly form teams, gather pricing quotes, or secure institutional support for more complex solicitation items.

The cross-cutting themes emerging from the EPSCoR Cyberinfrastructure Workshop Series provide compelling issues and topics for consideration at multiple scales, from local institutional scale to jurisdictional to national (including funding agency) scale; some may be best resolved at institutional scales, while others may be best addressed with integrated multi-scale approaches. The themes can serve as springboards for discussions focused on prioritizing actionable CI initiatives.

# 4. Recommendations to NSF EPSCoR

1. Re-establish CI as a required RII core component
2. Establish an NSF EPSCoR CI Council
3. Investigate models of CI Ecosystem human resources capacity-sharing for EPSCoR
4. Enhance collaborative partnerships between EPSCOR, OAC, and TIP
5. Incentivize proposal-stage participation by technical/CI staff

The five recommendations listed above represent solutions to community-identified needs. NSF EPSCoR's new 2023 vision of modular E-CORE and E-RISE projects is even more well-positioned than previous Track-based structures to incentivize and promote solutions to thematic challenges outlined in the previous section, and to contribute to multi-scale discussions and initiatives. Based on historical reports and our data, the recommendations that follow are key to success in overall NSF EPSCoR goals and objectives to expand capacity and increase research competitiveness. The individual Recommendation descriptions below are cross-referenced with the workshop-based Themes they most directly address, though a given recommendation may have connections to other Themes as well.

**Recommendation 1. Re-establish CI as a required RII core component:** To pilot this recommendation, include a *required* "Research Cyberinfrastructure and Technology Core" component amongst the E-CORE Research Infrastructure Improvement (RII) project cores. **We respectfully and intentionally recommend that CI be included as required, not optional.** Currently the only aspect of CI mentioned in the E-CORE solicitation is the mandated Data Management Plan. This has the unintended consequence of perpetuating the invisibility of cyberinfrastructure and CI professionals. It is salient to quote Borgman (2015) at length on this point:

> *People are often unaware how much they depend on an infrastructure, whether the electrical grid or the interoperability between two instruments, until that infrastructure ceases to function. … Those who benefit from using these infrastructures are often unaware of the background effort involved in keeping all the parts working smoothly together. … Invisible work is both glue and friction in collaborations, in the development of tools, in the sharing and reuse of data, and many other infrastructure components…*

In order to reverse the trends in technology inequities that the community data reveal, RII activities need to increase the visibility of CI, and include emphasis on both CI coordination and CI development.

The changes to the EPSCoR program announced during the May 15, 2023 NSF EPSCoR PI/PD meeting present a significant opportunity to rethink CI inclusion within the EPSCoR program portfolio. The authors of this report would like to underscore that the modular E-CORE and E-RISE approach that replaces the Track-1 awards has significant potential to address the previously described "science vs. technology" tension that many in the EPSCoR CI community remarked upon during workshop discussions by clearly separating potential funding streams for CI development initiatives into E-CORE solicitations, with topical scientific research incubation largely focused within E-RISE. However, there is also the risk that CI-related RII in jurisdictions could become even more fragmented and further widen technology gaps if NSF does not include a complementary requirement in the E-RISE solicitation that science incubators must coordinate not only with E-CORE Administrative efforts, but must also coordinate on E-CORE CI efforts. Additionally, baseline community best practices in CI coordination, assessment, and planning must be part of the RII foundation. A start to this would allocate specific funds, possibly by increasing the existing maximum award amount within E-CORE to enable inclusion of cyberinfrastructure coordination and development priorities; raising the maximum would explicitly avoid shifting existing tensions from "CI vs. Science" to "CI vs. Administrative and other cores."

Other functions of the proposed Research Cyberinfrastructure and Technology Core (which are distinct from the existing E-CORE Administrative Core and all other current Cores) include building coordinated foundations between institutional and jurisdictional-to-regional cyberinfrastructure and IT teams. In order for research, education, and training to be successful in the modern era, a foundation of technology must be in place (*e.g.*, network environments, identity management, security practices, data storage and transfer) that enables not only team science, but also administration, communications, and education at jurisdiction-scale (Theme 1).

To operationalize this recommendation, the Working Group further recommends to re-emphasize existing data management mandates, equitable CI accessibility (Theme 2), national CI engagement (Theme 3), and workforce development (Theme 4) as part of solicitations' CI requirements, while also being mindful that some institutions/jurisdictions may need significantly more investment in foundational IT (Theme 1) in order to support higher-level mandates such as end-to-end data management and structured workforce development. Require research computing and/or data professionals (as distinct from computer science faculty) to be funded as part of RII projects, encourage inclusion of such individuals on PI leadership teams, and require CI professional participation on the recommended NSF EPSCoR CI Council (Recommendation 2; Themes 3, 5). An important corollary is to include CI experts on proposal review panels and site visit teams (candidates might include, *e.g.,* OAC awardees). *Additional context for this recommendation is given in Appendix 2.* **Supports Themes 1, 2, 3, 4, 5.**

**Recommendation 2. Establish a CI Council:** Increase cyberinfrastructure's visibility and representation within the national EPSCoR community (Theme 3) by establishing and supporting as a core function of NSF EPSCoR an active CI Council akin to those already present for Communications (Comms), Education, Outreach, and Diversity (EOD), and Project Administrators and Project Directors (PA/PD). Include time during EPSCoR Annual Meetings for CI Council to gather and report back to the meeting as a whole. Functions the CI Council could serve include those outlined in Table 4.

| | |
|---|---|
| ❖ Orchestrating sustained conversation with NSF on data-driven status reports of CI in EPSCoR jurisdictions (Theme 3)<br>❖ Assisting in coordination of CI across jurisdictions and across cores (Theme 3)<br>❖ Providing students and early career researchers with exposure to multiple career pathways within research computing and data fields and a national network of colleagues (Theme 4)<br>❖ Providing NSF EPSCoR with perspective into national CI communities for best practices and policy developments (Theme 3)<br>❖ Collecting and providing access to jurisdictional stories of CI development (Theme 3) | ❖ Convening regular meetings with EPSCoR CI professionals and students around current topics of interest (Theme 3, 4)<br>❖ Facilitating EPSCoR contributions to and use of a national repository of campus CI Plans, mentorship plans, CI student and professional position descriptions, etc. (Theme 3, 4, 5)<br>❖ Coordinating CI workshops between EPSCoR jurisdictions and the national CI community (Theme 3, 4)<br>❖ Engaging with national CI communities to bootstrap awareness of EPSCoR CI needs and solutions – communities include CaRCC, RCD-Nexus CoE, ACCESS, Campus Champions, EDUCAUSE and others (Theme 3)<br>❖ Amplifying evolving community priorities, as identified. |

**Table 4: Proposed functions of an EPSCoR CI Council**

Ideally, Council members would lead or directly be involved in jurisdiction-level CI coordination and planning efforts as part of E-CORE teams (see Recommendation 1). Community infrastructure for this Council's activities exists today as part of CaRCC and the RCD-Nexus pilot Center of Excellence (OAC-2100003). The Working Group recognizes that national CI communities and organizations will need focused support (financial) to assist with EPSCoR-level needs; we suggest a potential mechanism for that support in Recommendation 3. **Supports Themes 3, 4, & 5.**

**Recommendation 3. Investigate models of CI Ecosystem human resources capacity-sharing for EPSCoR:** CI coordination and planning within EPSCoR jurisdictions is generally minimal, and CI capabilities suffer as a result. Recommendations 1 & 2 are intended to create and incentivize new human resources around research technology capabilities, which are intended to materially shift jurisdiction awareness and priority of research technology support. To further extend the first two recommendations, we recommend that NSF either be the engine or provide a mechanism (*i.e.,* an RFI) to identify and evaluate models for creating and organizing regional resource pools of CI professionals for EPSCoR.

Community input has shown that the available workforce of CI professionals is not enough to meet current and future demand across all EPSCoR institutions. This situation is very similar to that of the NSF's GRANTED program, where additional research administration support is needed across under-resourced institutions. To ensure equitability in research technology support, institutions will need to partner regionally and share subject matter experts (SMEs) for different technology elements of research workflows (*i.e.*, "Facings", with respect to the CaRCC RCD Capabilities Model terminology).

Any mechanisms to identify and evaluate resource-sharing models would be strengthened by partnering broadly across funding agencies and disciplinary communities to seek analogous examples that could be translated to apply to CI (for instance, in a 2019 Department of Energy (DOE) Biological and Environmental Research (BER) workshop report, the authors describe an "ICON-FAIR" model for resource-sharing in open watershed science (U.S. DOE, 2019; pp 16 – 22). The report presents many parallels to EPSCoR CI challenges alongside potential solutions. An effective human resource-sharing model would create an agile national CI human resource pool for EPSCoR researchers to draw upon when they need to assemble composite expertise beyond what is available on their campuses. NSF should engage multiple communities to identify successful human resource-sharing models that institutions can then use to develop their own solutions. Identification of successful human resource-sharing models would likely be useful not only for EPSCoR CI, but a much broader audience as well.

Existing NSF efforts to assemble and share access to expertise across the national CI community (*e.g.*, ACCESS) are centered around High Performance Computing (HPC) workflows, which do not represent the full spectrum of research activities within EPSCoR jurisdictions where commonly, smaller-scale but complex end-to-end workflows require attention to networks, automation, security, and data lifecycle. Addressing the CI needs of smaller-scale but still complex workflows would be a significant step toward releasing a bottleneck that currently impedes progress along the continuum of increasingly advanced CI resource use. Development and accessibility of CI ecosystem SMEs (a.k.a., Cyberinfrastructure Professionals (CIP), as seen in the Strengthening the Cyberinfrastructure Professionals Ecosystem (SCIPE) solicitation) are still emerging (*e.g.*, RCD-Nexus pilot, NSF SCIPE), and should be evaluated for their positive impact on current EPSCoR CI bottlenecks. EPSCoR support of the RCD-Nexus CoE pilot (OAC-2100003) for providing community assessment tools, SME mentoring, and community coordination is highly encouraged, as is EPSCoR input into and co-funding of SCIPE program initiatives. **Supports Themes 1, 3, 4, & 5.**

**Recommendation 4. Enhance collaborative partnerships between NSF EPSCoR, OAC, and TIP:** Echoing the spirit of the Committee on the Future of NSF EPSCoR's Recommendation Two (R2; Increased Integration of NSF EPSCoR), develop new opportunities and increase support and communication of existing pathways for jurisdictions to invest in cyberinfrastructure that supports research, especially for Foundational IT (Theme

1), and inclusive of the human effort required to implement and maintain the technology (Theme 3, 5). Continue existing co-funding efforts such as the EPSCoR co-funding budget increase for the Campus Compute (CC*; NSF 23-526) program. Build off the vision of successful efforts including GRANTED (NSF 23-221Y), SCIPE (NSF 23-521), CyberTraining (NSF 23-520), and CSSI (NSF 22-632) to develop new opportunities that strengthen diversity, equity, and inclusion (DEI) within the NSF funding portfolio; are compatible with the scale and scope of EPSCoR jurisdictions' existing resources; and that reflect the longer timeframes needed between solicitation announcements and due dates. Expanded collaboration between EPSCoR and TIP would help EPSCoR jurisdictions leverage existing but latent resources such as intellectual capital and industry-academia partnerships associated with TIP Directorate priorities. The *DCL: Towards an Equitable National Cyberinfrastructure* (NSF 21-108) is a good model mechanism to develop and communicate these opportunities. **Supports Themes 1, 3, & 5.**

**Recommendation 5. Incentivize proposal-stage participation by technical/CI staff:** Advance notice of key solicitation elements or themes, reduced friction in engagement and reporting, and professional staff co-PI inclusion would all contribute to institutions' capability to generate proposals and include CI/IT staff. Because institutional and jurisdictional/regional IT personnel and processes are by design not highly-reactive, it remains a significant challenge for RII proposal leadership to secure specific support commitments and participation of engineering staff without advance notice. An example of providing advance notice of key solicitation elements is the *Dear Colleague Letter: Announcement of Upcoming Topics for the 2023 NSF's Convergence Accelerator Solicitation* (NSF 23-066) that announced the themes for the 2023 Convergence Accelerator prior to releasing the final solicitation. Further incentivization of CI professionals' participation would be to ensure that EPSCoR solicitations are inclusive of non-traditional PIs such as CI/IT professionals who may not have faculty status.

Other incentives include a community-expressed interest that NSF work with jurisdictions to identify ways to decrease reporting burden and/or increase efficiency of available reporting frameworks. We received a number of comments that CI faculty/staff enthusiasm for participating in jurisdictional EPSCoR projects varies due to burdensome reporting and engagement requirements relative to other NSF programs. Both junior and senior faculty, as well as staff and students, already find that their available time for research activity is increasingly limited by administrative burden on campuses in EPSCoR jurisdictions. EPSCoR involvement is seen by many as "not worth the squeeze", limiting the pool of engaged personnel. If CI were a required E-CORE element, as suggested in Recommendation 1, many of these described inhibitors would be mitigated. For example, the national higher-education community has recently experienced a significant "brain drain" of IT expertise for both enterprise and research/education support, especially in "soft-money" support areas. Many hiring searches to replace personnel in these areas are failing repeatedly, greatly lengthening the time to finally fill necessary technical positions. This kind of time lag in CI/IT expectations and familiarity with EPSCoR will jeopardize any RII projects that are funded. Jurisdiction-level institutional knowledge and continuity would be greatly enhanced if CI were a required element of E-CORE just like basic Administration. **Supports Themes 1, 2, 3, 4, & 5.**

# 5. Recommendations to Jurisdictions and Institutions

Our workshop activities, survey results, and community feedback syntheses point to key areas where more intentional focus and structure within EPSCoR jurisdictions would reap significant benefits for research impact and associated education and workforce development. Specifically, the emergent themes point to how cyberinfrastructure functions as not just a technology core for research and associated education, but as a bridging mechanism between teachers and students, researchers and the public, administration and faculty, industry and academia. Evolution of national cyberinfrastructure development and practice appears to be outpacing NSF EPSCoR jurisdictions, therefore our recommendations are focused on encouraging jurisdictions to become more focused and intentional on developing the human and material aspects of cyberinfrastructure, as well as increasing interdisciplinary science and technology integration efforts. In particular, we strongly encourage EPSCoR jurisdictions and institutions to focus their RII efforts on building CI capability that generally supports "team science", which will in turn strengthen their overall competitiveness across all NSF directorates and other funding programs.

**Formalize CI assessment and planning:** Tie together a formal CI assessment process with the 2-3 year jurisdiction Science and Technology (S&T) Plan revision cycle. If the jurisdiction S&T Plan specifically includes a section on CI capabilities benchmarking, then coordinating EPSCoR projects (*e.g.*, E-CORE) should include the RCD Capabilities Model assessment process in their deliverables and function. This becomes a productive closed-loop process when S&T Plan revisions start to incorporate the gaps, opportunities, and priorities identified across the jurisdiction as part of the assessments. Additionally, the S&T Plan should address the following core CI elements in some fashion: Workforce Development Pathways; Survey and Assessment Tools; Data Management; Data Networks; Federated Identity and Access; Campus CI Plans; Jurisdictional/Regional CI Team. Some of these are excellent development targets for more focused E-CORE projects. Flexibility, responsiveness, continuous improvement, and "living document" practices are essential to planning processes, given dynamic campus landscapes and evolving research priorities. **Supports Themes 2 & 3, with impacts to Themes 1, 4 & 5 as well.**

**Coordinate CI development across RII projects and jurisdictions:** Create teams of CI personnel from across institutions that are charged with: facilitating organization- and jurisdiction-level periodic assessments and surveys toward developing clear CI priorities (Theme 2); contributing to S&T Plans, coordinating CI development within their home technology organizations and across the research enterprise, and engaging the national RCD community of practice (Theme 3); and creating structured workforce development plans and opportunities (Theme 4). Leverage EPSCoR projects both to pilot CI workforce development initiatives for STEM students and postdoctoral researchers (Theme 4), and for hiring permanent CI Professionals whose positions directly address jurisdictional CI priorities (Theme 5), assuming foundational IT is in place. Use EPSCoR RII projects to form and maintain jurisdictional-to-regional CI advisory and working groups, as well as widely-accessible infrastructure. Consider incentivizing Faculty Advisory Groups with buyout funding to ensure active stakeholders in the production of assessment, planning, and implementation. Identify staff workforce, administrative, and socio-political barriers from institutional to state level, and develop strategies to increase engagement in CI research and development funding activities. **Supports Themes 2, 3, 4, & 5.**

**Integrate regional network organizations:** Include the jurisdiction's Research and Education Network (REN) where possible in planning, assessment, and significant infrastructure/capacity building projects. In addition, the REN can be a source of guidance or assistance on key Team Science issues like federated identity, research networks, performance troubleshooting, and Cloud connections (Theme 1). As regional connectivity providers are often a politically-neutral and trusted party between campuses, integrating them into development of the regional CI ecosystem is likely to speed the process and reduce overall friction of pilot implementations (Theme 3). **Supports Themes 1 & 3.**

**Align foundational IT:** Foundational IT is understood here to mean both human and physical capital. Academic leadership within the jurisdiction should be working to break down socio-political barriers between administrative IT organizations, and aligning with the jurisdiction's objectives of education, research, service, and workforce development as outlined in the S&T Plan (Theme 2). Our workshops have illuminated the sometimes-significant friction that exists within EPSCoR institutions where basic technology capabilities related to research and education are concerned. EPSCoR project planning should take into account how individual institutions and their existing IT structures are prepared to engage and support Team Science and determine the gaps in their IT structures so investments in physical and human capital can be prioritized through institutional investment or targeted grant proposals (*e.g.*, RII, CC*, MRI) (Themes 1, 2, & 5). Neglecting to consider these foundational IT capabilities increases the risk of overall project failure or collapse of sustainability. Discovery of these capabilities and operating postures can be made possible through the RCD Capabilities Model assessment process. **Supports Themes 1, 2, & 5.**

**Measure CI Impacts:** EPSCoR projects often require significant goal-setting and success measurements, which are excellent opportunities to focus on measuring CI development impacts on research, education, and workforce development outcomes (Theme 2). While individual institutional metrics should be identified based on jurisdiction priorities and their CI/S&T Plans (Theme 3), a number of key variables are held in common: faculty recruitment and retention, student successes, publications and proposals supported, facilitation engagements, and researchers' perceived technology priorities. Notice that these are not "traditional" IT metrics, but instead are designed to measure success in the same way that the technology users measure success. The research IT metrics to collect and the tools to collect those metrics would be best determined in collaboration with the national CI community (*e.g.*, CaRCC, EDUCAUSE), the proposed NSF EPSCoR CI Council, and others so that metrics can be compared across institutions and the effort to establish the infrastructure to collect those metrics can be minimized (Theme 3). **Supports Themes 2 & 3.**

**Communicate the Role of CI:** Like most technology infrastructures, CI systems and professionals are generally overlooked by jurisdictions' administrative and political leadership unless there is a catastrophic problem. Communicating and lobbying the status and impacts of both human and physical CI to faculty members, institutional leaders, and jurisdictional political figures will broaden awareness and help with planning and sustainability (Theme 3). EPSCoR projects should tie intentional CI improvements with workforce development, educational outcomes, and overall jurisdiction strategy in their planning, reporting, and lobbying both internally as well as externally (Theme 2). Formalizing (*e.g.*, in the form of Memorandums of Understanding (MOUs)) and documenting existing and emergent CI partnerships with both internal campus and external partners should be prioritized as one strategy for future-proofing emergent initiatives (Bryant, 2020; 2023 (pers. comm.)). Consider policies and mechanisms in sponsored projects for proposal-stage and earlier consulting and involvement of institutional/jurisdiction CI personnel. **Supports Themes 2 & 3.**

# 6. Next Steps

As mentioned, the new NSF EPSCoR E-CORE and E-RISE solicitations offer promise with a more modular jurisdiction approach to Research Infrastructure Improvement (RII) projects. This structure offers opportunity for effective solutions as well as increased risk of perpetuating inequitable gaps in research technology access and capability. We anticipate that the E-CORE and E-RISE programs will evolve over the first cycle as the EPSCoR office and the community identify constructive changes and react to external opportunities. The community data, workshop feedback, and recommendations, alongside existing CI- and EPSCoR-focused reports are therefore timely inputs into EPSCoR programmatic planning.

We also propose to form a CaRCC EPSCoR CI Interest Group as a grassroots, self-organized community group that can start to address some of the many shared issues and solutions that are within jurisdictional control, as presented in this report. We are grateful for the opportunity to form the proposed Interest Group under the auspices of CaRCC and its existing infrastructure, and with CaRCC and RCD-Nexus support. The CaRCC EPSCoR CI Working Group is inspired by the degree of community interest and participation in the EPSCoR CI Workshop Series. The momentum generated by multiple engagements and lively discussions has led us to suggest this significant next step as a way to carry forward the community-inspired initiatives presented in this report. We call on the EPSCoR CI community to join us in founding the Interest Group, and what we hope will become the EPSCoR CI Council, founded with at least one cyberinfrastructure representative from every EPSCoR jurisdiction. We look forward to reaching out to the EPSCoR CI Community with more information in the near future.

# 7. References


Blatecky, A., Clarke, D., Cutcher-Gershenfeld, J., Dent, D., Hipp, R., Hunsinger, A., Kuslikis, A., & Michael, L. 2021. *The missing millions: Democratizing computation and data to bridge digital divides and increase access to science for underrepresented communities*. National Science Foundation. https://www.rti.org/publication/missing-millions

Blatecky, A., Brunson, D., Cheatham, T., Gershenfeld, J., Hedberg, A., Bottom, J., & Reed, D. 2019. *National Cyberinfrastructure Coordination Service Conference: Rethinking NSF's computational ecosystem for 21st century science and engineering*. https://www.rti.org/publication/national-cyberinfrastructure-coordination-service-conference

Borgman, Christine. 2015. Big Data, Little Data, No Data. Cambridge: MIT Press.

Bryant, Rebecca, Annette Dortmund, and Brian Lavoie. 2020. *Social Interoperability in Research Support: Cross-Campus Partnerships and the University Research Enterprise*. Dublin, OH: OCLC Research. https://doi.org/10.25333/wyrd-n586

Committee on the Future of NSF EPSCoR. 2022. *Envisioning the Future of NSF EPSCoR Report*. National Science Foundation. https://nsf-gov-resources.nsf.gov/2022-08/Envisioning-The-Future-of-EPSCoR-Report.pdf

Congress.gov. 2022. "H.R.4346 – 117th Congress (2021-2022): Chips and Science Act." August 9, 2022. https://www.congress.gov/bill/117th-congress/house-bill/4346

Devitt DA, Apodaca L, Bird B, Dawyot Jr JP, Fenstermaker L, Petrie MD. 2022. *Assessing the Impact of a Utility Scale Solar Photovoltaic Facility on a Down Gradient Mojave Desert Ecosystem*. Land. 2022 Aug 15; 11(8):1315. https://doi.org/10.3390/land11081315

Directorate-General for Research and Innovation (European Commission) (DGRI-EC). 2016. *Realising the European Open Science Cloud: First Report and Recommendations of the Commission High-Level Expert Group on the European Open Science Cloud*. Brussels, Belgium: EC. https://data.europa.eu/doi/10.2777/940154

Hill, Paul. 2012. *EPSCoR 2030: A Report to the National Science Foundation* (prepared by Paul Hill, Principal Investigator: Arlington, VA: National Science Foundation). https://wvresearch.org/epscor-2030-a-report-to-the-national-science-foundation

Knepper, R, and Börner, K. (2016) *Comparing the Consumption of CPU Hours with Scientific Output for the Extreme Science and Engineering Discovery Environment (XSEDE)*. PLOS ONE. https://doi.org/10.1371/journal.pone.0157628

Le VD, Neff MM, Stewart RV, Kelley R, Fritzinger E, Dascalu SM, Harris FC. 2015. *Microservice-based architecture for the NRDC*. In: 2015 IEEE 13th International Conference on Industrial Informatics (INDIN) 2015 Jul 22 (pp. 1659-1664). IEEE. https://doi.org/10.1109/INDIN.2015.7281983

McMahon MJ, Dascalu SM, Harris FC, Strachan S, Biondi F. 2011. *Architecting climate change data infrastructure for Nevada*. In: Salinesi, C., Pastor, O. (eds) Advanced Information Systems Engineering Workshops: CaiSE 2011 International Workshops, London, UK, June 20-24, 2011. Lecture Notes in Business Information Processing, vol 83. (pp. 354-365). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-22056-2_38

Mensing S, Strachan S, Arnone J, Fenstermaker L, Biondi F, Devitt D, Johnson B, Bird B, Fritzinger E. 2013. *A network for observing Great Basin climate change*. Eos, Transactions American Geophysical Union. 94(11):105-6. https://doi.org/10.1002/2013EO110001

Moore, Loretta. 2019. *Increasing access to cyberinfrastructure for EPSCoR and other underresourced institutions.* [Keynote address given at the Internet2 National Research Platform Conference.] https://internet2.edu/wp-content/uploads/2022/02/Loretta-Moore_Keynote-NRP-Presentation_rf.pdf

Nelson, Alondra (OSTP). 2022 August 25. *Ensuring Free, Immediate, and Equitable Access to Federally Funded Research.* https://www.whitehouse.gov/wp-content/uploads/2022/08/08-2022-OSTP-Public-Access-Memo.pdf

National Science Foundation (NSF). 2019. *Transforming Science Through Cyberinfrastructure: NSF's Blueprint for a National Cyberinfrastructure Ecosystem for Science and Engineering in the 21st Century*. Alexandria, VA. https://www.nsf.gov/cise/oac/vision/blueprint-2019/Coordination-Services.pdf

NSF Advisory Committee for Cyberinfrastructure. 2018. *CI2030: Future Advanced Cyberinfrastructure*. https://www.nsf.gov/cise/oac/ci2030/ACCI_CI2030Report_Approved_Pub.pdf

[*s.n.*]. *Request for Information (RFI) on Developing a Roadmap for the Directorate for Technology, Innovation, and Partnerships at the National Science Foundation*, v. 88 no. 82 Fed. Reg. 26345 (April 28, 2023). https://www.govinfo.gov/content/pkg/FR-2023-04-28/pdf/2023-08995.pdf

[Montana State University]. Research Alliance. 2023. https://www.montana.edu/research-alliance/

Schmitz, P., Bayrd, V., Strachan, S., & Jacobs, G. 2022. *A Baseline of EPSCoR Research CI Capabilities*. RCDNexus Technical Report. https://doi.org/10.5281/zenodo.6395204

Schmitz, P. 2021. *Assessing the Landscape of Research Computing and Data Support: The 2020 RCD Capabilities Model Community Dataset*. Practice and Experience in Advanced Research Computing, 1–8. https://doi.org/10.1145/3437359.3465580

Schmitz, P., Yockel, S., Mizumoto, C., Cheatham, T. and Brunson, D., (2021) *"Advancing the Workforce That Supports Computationally and Data Intensive Research*" in Computing in Science & Engineering, vol. 23, no. 05, pp. 19-27, 2021. https://doi.org/10.1109/MCSE.2021.3098421

Schmitz, P., Mizumoto, C., Hicks, J., Brunson, D., Krovitz, G., Bottum, J. R., Cutcher-Gershenfeld, J., Wetzel, K., & Cheatham, T. (2020). *A Research Computing and Data Capabilities Model for Strategic Decision-Making*. Practice and Experience in Advanced Research Computing (PEARC '20). PEARC '20: Practice and Experience in Advanced Research Computing, Portland, OR, USA. https://doi.org/10.1145/3311790.3396643

Scully-Allison C, Munoz H, Le V, Strachan S, Fritzinger E, Harris Jr FC, Dascalu SM. (2018) *Advancing Quality Assurance Through Metadata Management: Design and Development of a Mobile Application for the NRDC*. IJCA International Journal of Computers and Their Applications. 2018;25(1):20-9.

Stewart, C.A., Costa, C.M., Wernert, J.A., Snapp-Childs, W., Bland, M., Blood, P., Campbell, T., Couvares, P., Fischer, J., Hancock, D. Y., Hart, D. L., Jankowski, H., Knepper, R., McMullen, D. F., Mehringer, S., Pierce, M., Rogers, G., Sinkovits, R. S., and Towns, J. (2023) *Use of accounting concepts to study research: return on investment in XSEDE, a US cyberinfrastructure service*. Scientometrics. https://doi.org/10.1007/s11192-022-04539-8

U.S. DOE. 2019. *Open Watershed Science by Design: Leveraging Distributed Research Networks to Understand Watershed Systems Workshop Report*. DOE/SC-0200, U.S. Department of Energy Office of Science. https://doesbr.org/openwatersheds/

Van Eseltine, M, Nicholls, N., and Owen-Smith, J. (2020). *Report on the Scientific and Economic Effects of XSEDE*. Institute for Research on Innovation & Science (IRIS), University of Michigan.

Von Laszewski, G., Wang, F., Fox, G. C., Hart, D. L., Furlani, T. R., DeLeon, R. L., & Gallo, S. M. (2015). *Peer comparison of XSEDE and NCAR publication data*. 2015 IEEE international Conference on Cluster Computing (pp. 531–532). https://doi.org/10.1109/CLUSTER.2015.98

Von Laszewski, G., Wang, F., and Fox, G.C. (2021). *Comprehensive Evaluation of XSEDE's Scientific Impact using Semantic Scholar Data*. PEARC '21: Practice and Experience in Advanced Research Computing, July 19-22, 2021, Virtual. https://doi.org/10.1145/3437359.3465601

Wang, F., von Laszewski, G., Whitson, T., Fox, G. C., Furlani, T. R., Deleon, R. L., and Gallo, S. M. (2018). *Evaluating the Scientific Impact of XSEDE*. PEARC '18: Practice and Experience in Advanced Research Computing, July 22-26, 2018, Pittsburg, PA, USA. https://doi.org/10.1145/3219104.3219124

# Appendix 1: Breakdown of Workshop Statistics and Participating Institutions

While the EPSCoR CI assessment dataset includes only a portion of the many institutions in EPSCoR jurisdictions, the key findings from the aggregate dataset analysis resonated with many members of the community who have reviewed the results, including Spring 2022 and Fall 2022 Workshop participants who had and had not completed an assessment for their institution. Of the EPSCoR institutions represented in the Spring and Fall Workshops, only slightly less than half were affiliated with an institution that had completed an assessment.

- The Spring 2022 workshop included 41 participants from 25 EPSCoR institutions representing 20 jurisdictions, as well as six participants from non-EPSCoR institutions and organizations. Of the 25 participating EPSCoR institutions, 13 had not completed an assessment, but all reviewed and discussed the resulting data.
- The Fall 2022 workshop included 42 participants from 20 EPSCoR institutions representing 17 jurisdictions, as well as three NSF personnel. Of the 20 participating EPSCoR institutions, 12 had not completed an assessment, but all reviewed and discussed the resulting data.
- A total of 35 institutions representing 25 EPSCoR jurisdictions contributed assessments and/or participated in one or both of the workshops. Of the 28 designated 2022 EPSCoR jurisdictions, only Nebraska, New Hampshire, and Puerto Rico were not represented by either assessment contribution or workshop participation.

Participating institutions (**bold** indicates institution contributed an assessment):

- Boise State University
- **Chaminade University of Honolulu**
- Clemson University
- Dakota State University
- Kansas State University
- **Louisiana State University**
- Mayville State University
- Mississippi State University
- **Montana State University**
- North Dakota State University
- Shepherd University (West Virginia)
- University of Alabama at Birmingham
- University of Alaska
- **University of Arkansas**
- **University of Delaware**
- **University of Guam**
- **University of Hawai'i at Manoa**
- **University of Hawaii at Hilo**
- **University of Hawaii West Oahu**
- University of Idaho
- University of Iowa
- **University of Kentucky**
- **University of Louisville**
- University of Maine
- **University of Nevada, Reno**
- **University of New Mexico**

- University of North Dakota
- University of Oklahoma
- **University of Rhode Island**
- **University of South Carolina**
- **University of South Dakota**
- University of the Virgin Islands
- University of Vermont
- University of Wyoming
- West Virginia University

A common theme heard from participating institutions was that they were too overwhelmed with their regular work supporting researchers to complete the assessment and contribute their data, further underscoring our findings of insufficient CI/RCD staffing support[3]:

> *"Unfortunately, [our] ITS staff simply doesn't have the capacity to take this on right now."*
> - A Carnegie Research Doctoral workshop participant

[3] A total of 45 institutions in 28 EPSCoR jurisdictions have requested a copy of the RCD Capabilities Model assessment tool, however only 16 institutions completed and contributed data. Another eight made some progress but were not able to complete their assessment.

# Appendix 2: Additional Context on Recommendation #1 to NSF

Additional context and/or supporting actions for Recommendation #1 to NSF are provided below.

**Re-establish CI as a required RII core component:** Historically, CI has been sometimes a required element of EPSCoR programs, and sometimes not. However, technology capability has emerged as the common denominator in science and engineering velocity, impact, and success. For example, a recent survey of several jurisdictions at the 2023 Great Plains Network annual meeting in a CI planning session indicated that regional successes in CI were directly due to "EPSCoR CI requirements 15 years ago", but had "struggled with stagnation since" (PI Strachan, pers. comm.). Lagging awareness and investment in technology in EPSCoR jurisdictions has resulted in a risky patchwork of CI capabilities that perpetuates gaps and inequalities in science and workforce development. Recent efforts in making EPSCoR jurisdictions a priority in NSF OAC Campus Cyberinfrastructure funding solicitations have been successful in getting more institutions engaged in CI assessment and planning again, but in order to make immediate and lasting progress across the broad range of EPSCoR participants, complimentary changes within flagship EPSCoR funding programs are also needed. Our workshop series has resulted in high-priority community recommendations that can catalyze long-term improvements to institutional expertise and resulting competitiveness. We are recommending that NSF re-create CI requirements as part of an intentional effort to redefine and reemphasize the modern and broadly utilitarian "Infrastructure" piece of the Research Infrastructure Improvement program, which should include but is not limited to research CI, research communications, and research administration and reporting.

The new E-CORE and E-RISE modular approach should explicitly include CI elements, or risk their omission entirely with detrimental effects at the 5-10 year scale. For example, besides "Research Administration", the E-CORE should also require "Research CI and Technology" as a core required element. The reasoning for this is that regardless of jurisdictions' research and education goals and domains, a core foundation of technology and related facilitation is absolutely required. Key technology coordination functions that do not otherwise exist in most jurisdiction models include the following areas: regional identity and access, data sharing and transfer, data lifecycle management, jurisdictional CI team engagement, and research cybersecurity. Key activities that E-CORE CI personnel should be directed toward include: jurisdictional-to-regional CI planning and assessment, national CI community engagement, internal bridge-building between institutional technology centers and EPSCoR research groups, and CI workforce development. In these capacities, core CI personnel become the "technology glue" that enables effective and cutting-edge "team science" within jurisdictions, which will result in higher competitiveness for researchers across all areas of NSF.

Because these would be additional team functions, we recommend that technology-specific funding be added to the E-CORE, with explicit requirements on the relative amounts dedicated to Administration and CI/Technology, and separate from all of the "optional" E-CORE categories. Incentivizing jurisdictions to recognize and close the gaps in research technology capability and coordination will have rapid impact across a wide range of research and education efforts at both institutional and jurisdictional scale.

Feedback from the community, national assessment results, and our synthesis in this report all indicate that NSF would be better positioned to succeed in high-level objectives by becoming more intentional in emphasizing human along with material CI in science infrastructure-oriented funding programs. **Community recommendations** to address research CI needs center on socio-political incentivization solutions, and are directly quoted here:

- Promote inter- and intra- jurisdictional CI and interdisciplinary collaborations and communication;
- Embed CI positions within research projects;
- Enable CI pros to have PI status and write grants;
- Recognize science PI's CI efforts in Promotion & Tenure processes and change the institutional research culture to embrace CI efforts;
- Create jurisdiction-wide CI strategic plans as part of or complementary to the jurisdiction-wide S&T plan requirement;
- Create a separate EPSCoR funding track targeting data management (DM) gaps, toward developing a technical DM framework to support Track-1 and smaller research projects;
- Integrate data policy, practice, and frameworks into strategic planning;
- Increase partnership opportunities between EPSCoR and non-EPSCoR projects, incentivizing [projects] between universities, federal and state governments, and private industry, beyond track 4 (minimum viable consortium) (residencies, advisory boards);
- Provide a longer lead time to develop proposals and longer award periods to build sustainability;
- Change NSF reporting requirements to ease burden (administrative support staff are often not available to assist; this creates an opportunity cost for PIs).

The above quoted community recommendations make it clear that multi-scale solutions are essential to address ongoing EPSCoR CI inequities. Re-establishing CI as a core RII component would be a significant catalyst in this space.

Furthermore, we recommend that NSF require a jurisdiction-level CI plan specifying CI as an "R&D infrastructure improvement resource" (Theme 2), being mindful that human infrastructure is understood to be a key component of any infrastructure development efforts (Theme 5). Recent steps in making EPSCoR jurisdictions a priority in NSF OAC Campus Cyberinfrastructure (CC*) funding have been successful in getting more institutions engaged in CI assessment and planning, but in order to make lasting progress across the broad range of EPSCoR participants, complimentary investment in core infrastructure services is also needed. The NIH INBRE Bioinformatics Core and Data Science Core are representative examples of ongoing federal investments in core CI infrastructure.

As corollaries to this recommendation, include CI experts on proposal panels and site visit teams. Proposal review must include an intentional component to consider how EPSCoR proposals will ensure sufficient investment in CI to support the growing needs in proposed and emerging areas of research (*e.g.*, Artificial Intelligence (AI), Advanced Materials Science, Biotechnology, Quantum Information Science, Open Data). Reviewers and site panels are recommended to include members with CI expertise; potential sources to solicit reviewers include the Campus Research Computing Consortium (CaRCC), Campus Champions, EDUCAUSE Research Computing and Data Group, NSF ACCESS participant channels, The Research Data Alliance - USA branch (RDA-US), and the Research Data Access and Preservation (RDAP) Association.
**Supports Themes 1, 2, 3, 4, 5.**