

A Baseline of EPSCoR Research CI Capabilities

Highlights from the 2020/2021 RCD CM Community Dataset

Technical Report
RCDNexus-TR-2022.2

Patrick Schmitz
Semper Cogito Consulting
patrick@sempercogito.com

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

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

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

Contents:

1. Introduction	1
2. Notes on the contributing institutions and analysis methodology	2
3. Summary and Facing level findings	4
4. Looking more closely at Researcher-Facing areas	7
5. Looking more closely at Data-Facing areas	7
6. Looking more closely at Software-Facing areas	8
7. Looking more closely at System-Facing areas	9
8. Looking more closely at Strategy and Policy-Facing areas	9
9. Looking at Priorities	11
10. Looking at outliers (low and high) among the data	14
11. Acknowledgements	15
12. References	15
Appendix I: EPSCoR Low Coverage topics relative to non-EPSCoR	16
Appendix II: EPSCoR Low Coverage topics relative to EPSCoR mean	17
Appendix III: EPSCoR High Coverage topics relative to non-EPSCoR	18
Appendix IV: 2020 Top Priorities for EPSCoR Institutions	19

1. Introduction

Research Cyberinfrastructure (CI, a.k.a. Research Computing and Data (RCD)) is changing at an accelerating rate, while the breadth of fields and disciplines that depend on these technologies is expanding and increasingly diverse. This poses significant challenges to academic institutions as they try to effectively assess and plan for the necessary support infrastructure to keep pace with the needs of researchers. EPSCoR jurisdictions have historically underinvested in the CI resources and services required to drive forward data and compute-intensive science (Hill, 2012). This has created an uneven distribution and availability of CI across the nation (Hill, 2012; Blatecky, et al., 2019). However, the degree of this inequality remains unclear because data is lacking on CI availability across EPSCoR jurisdictions and institutions. The 2022 EPSCoR CI workshop is working to address this by generating a unique baseline and compiling aggregate data concerning EPSCoR jurisdictions' current CI capabilities, gaps, and priorities in EPSCoR jurisdictions. These data will improve knowledge of CI distribution and access, serving to increase discussion at local, regional, and national levels

This report presents an analysis of the 2020/21 aggregated data from 15 EPSCoR institutions that completed Research Computing and Data Capabilities Model (Schmitz, et al. 2020) assessments in 2020 and 2021. These institutions represent 11 states and U.S. territories, and include R1, R2, and other Carnegie Classifications. The RCD Capabilities Model allows institutions to assess their current capabilities and provides structured input for strategic decision-making using a shared community vocabulary. The model presents roughly 150 capabilities (in the form of questions) structured around five "Facings" that are increasingly used as a means of characterizing the roles of people who support CI/RCD: Researcher-Facing, Data-Facing, Software-Facing, System-Facing, and

Strategy and Policy-Facing. The Assessment Tool also allows institutions to mark specific capabilities as priorities. The resulting dataset provides important insights into the state of support for RCD, at both summary and granular levels (Schmitz, 2021).

For the community of institutions in EPSCoR jurisdictions, the dataset also shows clear gaps in coverage relative to peers in non-EPSCoR jurisdictions. While patterns revealed in the data may be unsurprising to some, it is important to have this baseline – both to corroborate the experience of leaders at these institutions, as well as to provide a detailed framework for discussion among EPSCoR institutions on how best to support research in their jurisdictions. It is our hope that the report provides a basis for fruitful discussions in the March 2022 preliminary workshop, as well as the Fall 2022 workshop to be held in conjunction with the national EPSCoR conference¹.

While the relatively small number of participating institutions and the wide variance among them means that most of our conclusions are not statistically significant, the main described resonate with many members of the community who have reviewed the results, and are compelling enough to merit consideration. 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.
- The top priorities of the contributing EPSCoR institutions are mostly in Data-Facing, Software-Facing, and Researcher-Facing topics, although the fourth-highest priority is the need for more strategic planning.
- An interesting pattern emerged between the 2020 dataset and the combined 2021 dataset in which half the 2020 contributors repeated their assessment. For the repeating institutions, average Strategy and Policy-Facing coverage increased from 44% to 60% – a remarkable 35% in one year.

The rest of this paper is structured as follows: Section 2 presents the demographics of the contributing institutions and details of the analytic methodology; Sections 3 through 8 present summary and more detailed analyses of the assessment data, comparing EPSCoR and non-EPSCoR institutions; Section 9 discusses the priorities data; and Section 10 presents some notable outlier topics in the data, with associated data in 4 appendices.

2. Notes on contributing institutions and analysis methodology

In 2020, we had a total of 10 EPSCoR institutions among the 41 contributors. Of these 10, nine were public institutions and one was private; seven were R1 institutions, and the others were an R3, an M1, and a baccalaureate college. The data are not sufficient in sample size or diversity to fully represent the EPSCoR community, however the results did resonate with the working group members and others with experience at EPSCoR institutions.

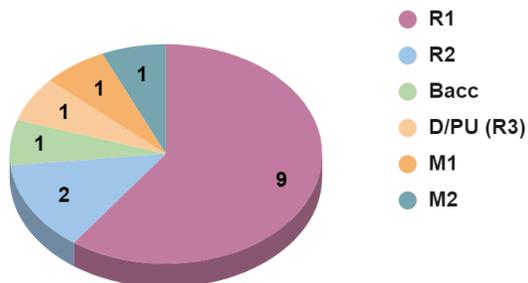


Fig. 1 Contributing institutions by Carnegie Classification

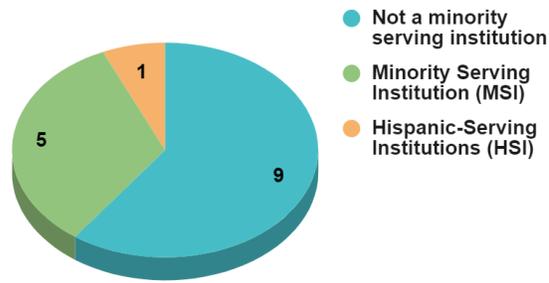


Fig. 2 Contributing institutions Minority-serving status

In 2021, five of the 2020 institutions repeated their assessments, and five additional (new) institutions contributed assessments. All new institutions were public, but only two were R1, somewhat diversifying the pool (see Fig 1).

¹ <https://umaine.edu/portland/event/27th-nsf-epscor-national-conference/>

The EPSCoR dataset includes more minority-serving institutions than in the full 2020/21 dataset which is about 75% not minority serving (see Fig 2), however there are no Historically Black Colleges and Universities (HBCUs). This is an area the RCD Capabilities Model working group continues to focus on improving.

This yields 15 institutions in the combined 2020/2021 dataset, all but one of which are public, including a range of Carnegie classifications, and representing 11 jurisdictions: Delaware, Guam, Hawaii, Kentucky, Louisiana, Montana, Nevada, New Mexico, Rhode Island, South Carolina, and South Dakota (see Fig 3).

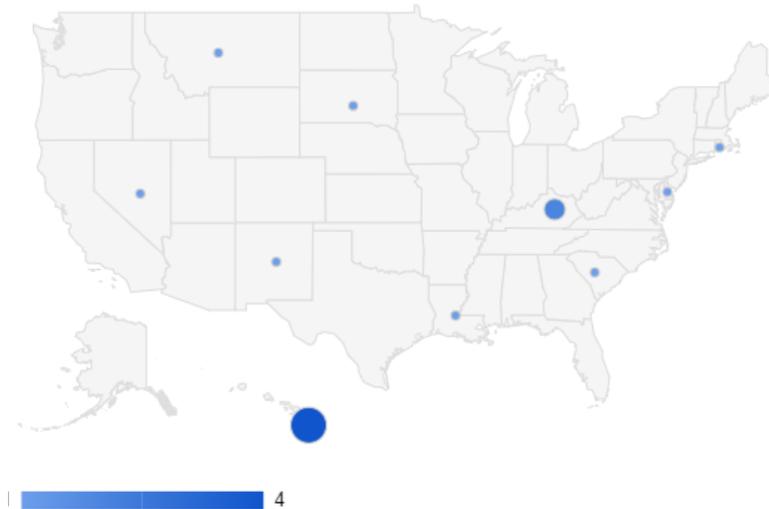


Fig 3: EPSCoR jurisdictions contributing to the 2020/21 dataset (plus Guam, not shown)

In response to considerable community feedback, version 1.1 of the tool used by contributors in 2021 included two changes to the calculation of the coverage value for each row/topic in the model

1. We adjusted weighting of the three main columns in the assessment tool (*Deployment at Institution*, *Multi-Institutional Collaboration*, and *Service Operating Level*) to reduce the impact of the collaboration column, not only to reduce the computed coverage when there was no collaboration at all, but also to give a slight boost for *Leading multi-institutional collaboration*.
2. The version 1.0 “*Local Relevance*” column was removed, and this was incorporated into the *Deployment at Institution* choices. In the version 1.1 version of the assessment tool, when a row/topic is marked as “*Not relevant or applicable*” this row was simply ignored in the summary values (e.g., for that theme in the facing, and for the overall facing coverage). In the version 1.0 used in 2020, a *Local Relevance* of zero yielded a 100% coverage for that row/topic, which had the effect of slightly boosting the summary computed values (not a reasonable result).

The first change affected nearly all assessments, where the second change affected only the relatively few assessments that selected more than a very few topics as “*Not relevant or applicable*.” As such, the net impact of these changes was a slight increase in the summary computed coverage values for most institutions. However, in a few isolated cases, this reduced the summary computed coverage values for an institution (i.e., where the second change had a greater impact).

For the combined 2020/2021 dataset, we retrofitted the revised weighting to all 2020 assessments, and disregarded all 2020 assessments for which there was a 2021 assessment from the same institution (i.e., we used the newer data for repeating institutions). This resulted in a total of 51 unique contributing institutions, so that EPSCoR institutions represent just over 29% of total institutions and just over 34% of total represented states and territories.

Note that some caution must be exercised in comparing 2021 graphs to those in the 2020 report, as the 2021 graphs use the new weighting scheme. Where we mention relative percentages comparing 2020 and 2021, we are using average coverage values computed *with the new weights for both years*.

3. Summary and Facing level findings

There is considerable variation in the areas of relative strength and weakness among the contributing institutions, as illustrated in Fig. 4. However, the range of variation is somewhat smaller among contributing EPSCoR institutions as compared to non-EPSCoR institutions.

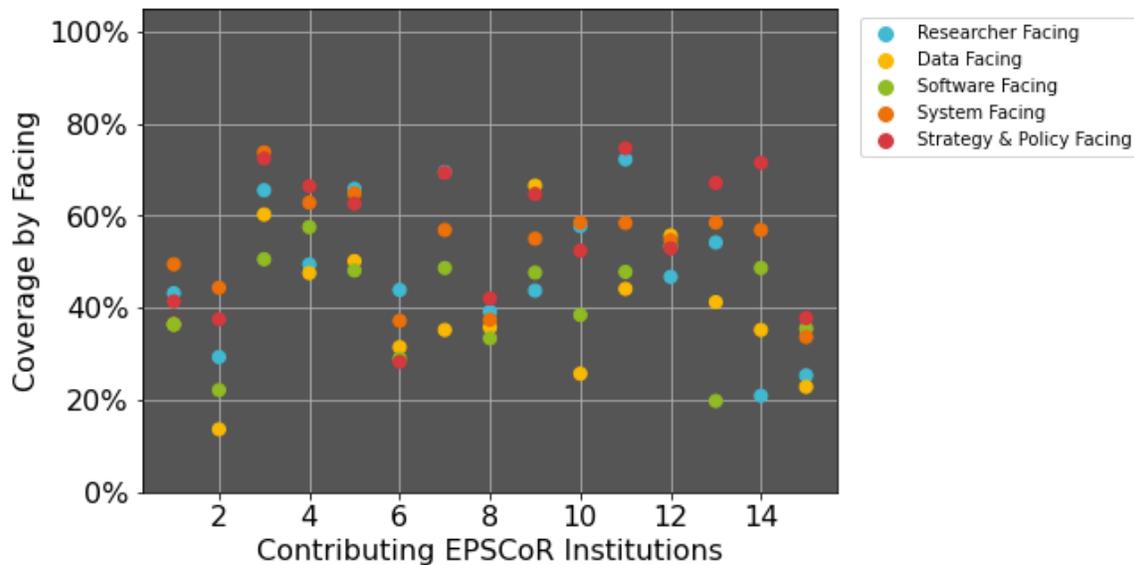


Figure 5: Coverage by Facing for all contributing EPSCoR Institutions

At both the summary level and in the individual facings areas, EPSCoR institutions lag behind their non-EPSCoR peers. Just as last year, significant gaps exist across all the facing areas. However, gaps are greatest in the Data-Facing, Researcher-Facing, and Software-Facing topics (see Fig 4).

It is interesting that from 2020 to 2021, the gap in Strategy and Policy-Facing topics has narrowed. While the non-EPSCoR average rose slightly (1.1 percentage points) from 2020 to 2021, the EPSCoR average rose more sharply, from 50.2% to 57.4%. This is due, in part, to the addition of new EPSCoR institutions (that averaged 52% in Strategy and Policy-Facing coverage). It should be noted, however, for the five EPSCoR institutions that repeated assessments in 2021, the average Strategy and Policy-Facing coverage rose from 44% to 60% – a remarkable 35% increase. Could this year-over-year progress in topics related to strategy and policy be correlated to engagement with the RCD CM and using it in their strategic planning practices? Perhaps these institutions had resolved to make improvements in this area and so were motivated to use the RCD CM as part of that, and perhaps the exercise of conducting the assessment helped them to structure their Strategy and Policy-Facing work. In any case, it is a notable data point².

Also consistent with last year is a significant range of results across the community, as indicated by the error bars (which show one standard deviation up and down). In both Data-Facing and System-Facing, there is somewhat less variation among EPSCoR institutions than non-EPSCoR institutions.

² As a point of comparison, the average Strategy and Policy-Facing coverage for non-EPSCoR institutions that repeated an assessment in 2021 also rose, from 64% to 72%. For all institutions, the average Strategy and Policy-Facing coverage for repeating institutions is far above that of the new 2021 institutions. This is an interesting trend to watch over time.

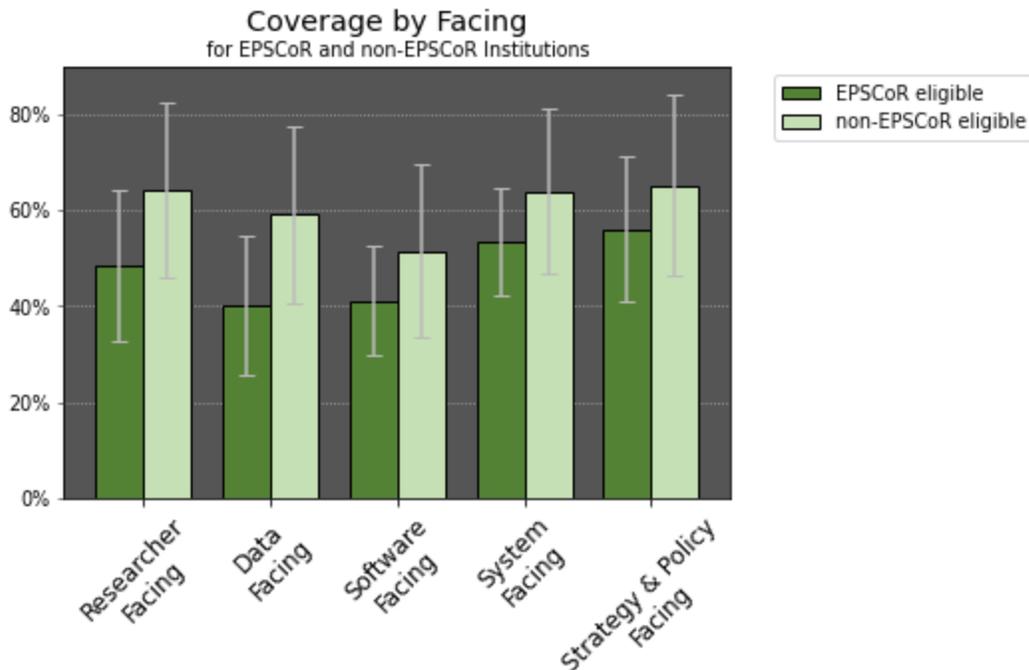


Figure 4: Summary Coverage by Facing

Considering the individual columns that make up the computed coverage

An additional factor to be considered overall is the broad gap between EPSCoR institutions and non-EPSCoR institutions in the average *Service Operating Level*, which characterizes the robustness of service support for each area or topic (see Fig. 6). The average is just 2.75 for EPSCoR institutions (i.e., between “*Substantial Risk of Failure(s)*” and “*Lights on Only*”) and is 3.31 for non-EPSCoR institutions (between “*Lights on Only*” and “*Basic/Economy*”). Just as for the total coverage values, the gap is widest for Research-Facing and Data-Facing topics, indicating that these areas are the most challenging for EPSCoR institutions to operate in a sustainable manner.

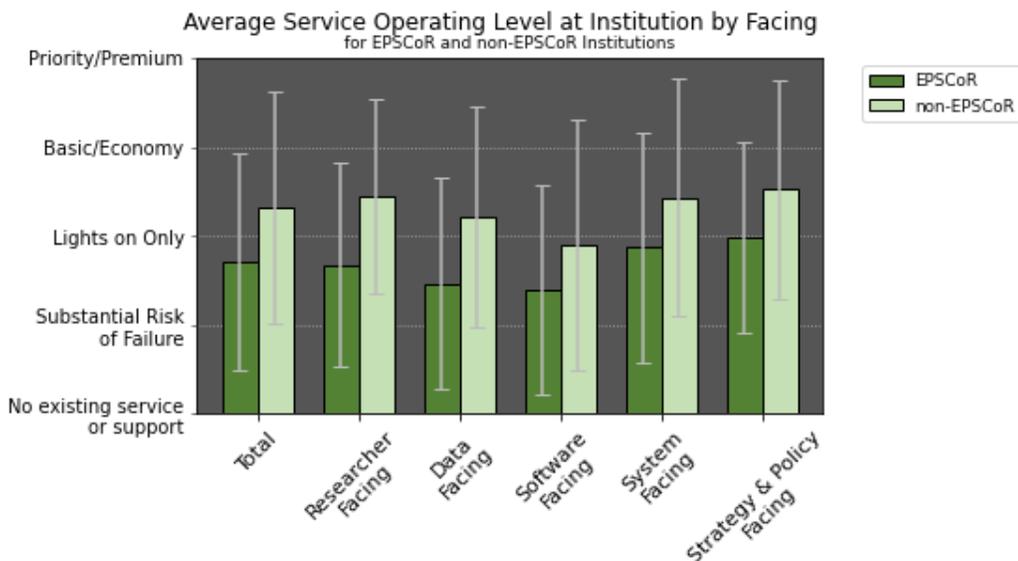


Figure 6: Service Operating Levels column

The average *Service Operating Level* for 2020/2021 is higher than it was for 2020 EPSCoR institutions (2.56), and the average *Service Operating Level* for EPSCoR institutions that repeated their assessments rose from 2.31 to 2.61, a nearly 13% increase year-over-year.

The *Deployment at Institution* column asks how widely available each service area or topic is on campus. In general, these values are somewhat higher than the *Service Operating Level* values, but gaps between EPSCoR and non-EPSCoR institutions follow the patterns seen above. From 2020 to 2021, average values for *Deployment at Institution* rose only slightly (approximately 3%, from 3.11 to 3.22) for EPSCoR institutions, and the average values for *Multi-Institutional Collaboration* were nearly unchanged (2.17 for 2020/2021 vs. 2.19 for 2020). See Fig. 7 for details.

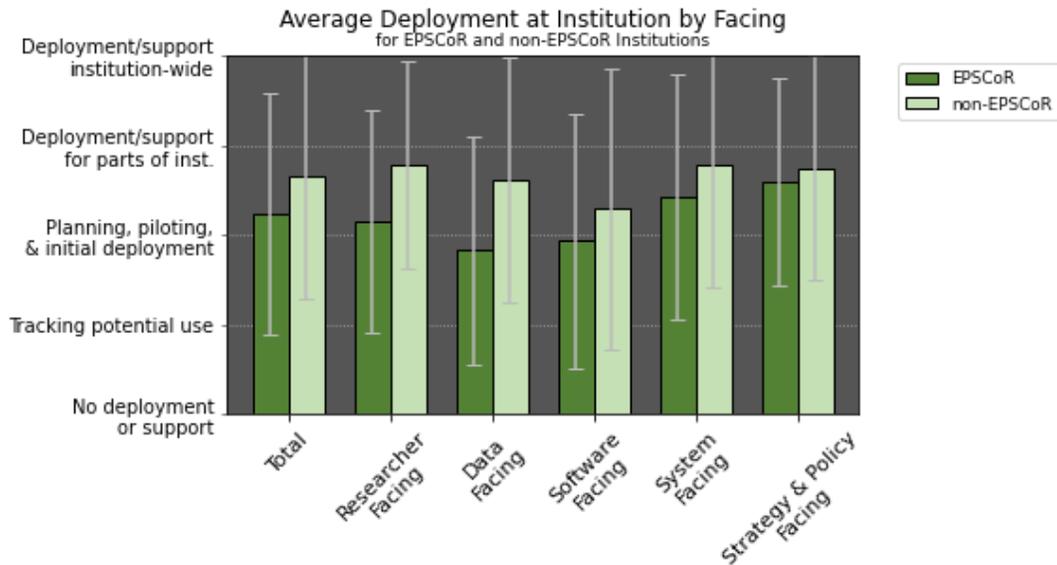


Figure 7: *Deployment at Institution* column

4. Looking more closely at Researcher-Facing areas

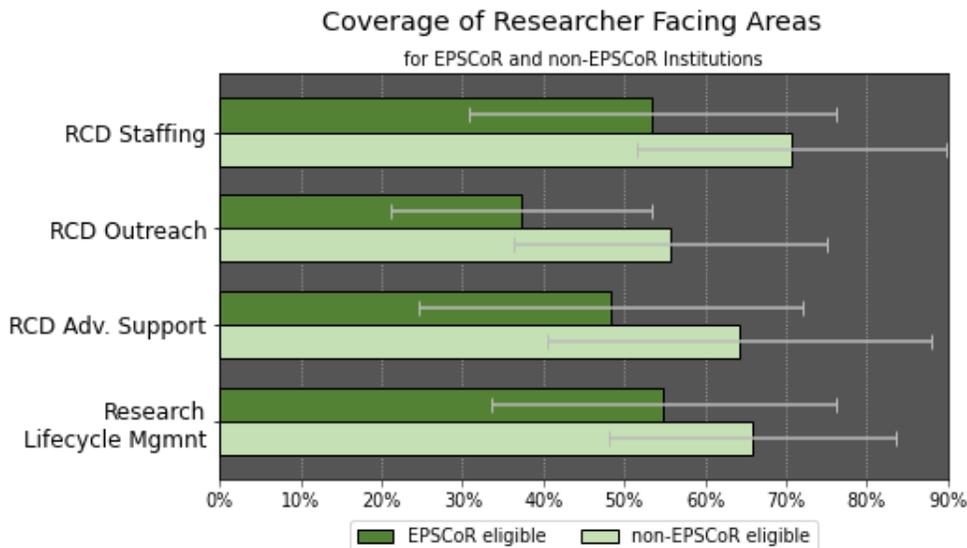


Figure 8: Researcher-Facing Areas

The general patterns in Researcher-Facing results (see Fig. 8) are roughly consistent with what we saw in 2020. Most coverage values trended slightly upward with the 2021 weighting, and the value for RCD Staffing increased even more. However, there is one exception; in 2020, EPSCoR institutions were stronger on Research Lifecycle Management than the non-EPSCoR institutions in the dataset, whereas the 2021 dataset shows EPSCoR institutions trailing somewhat. There are two factors contributing to this relative change:

1. Reweighting the model significantly reduced the EPSCoR average (roughly 5%, perhaps due to the change in handling “not relevant”).
2. The average coverage in this area for non-EPSCoR institutions rose 2.3%.

Nevertheless, gaps remain in Staffing, Outreach, and Advanced Support topics, likely indicating that EPSCoR institutions have a lower level of resources available to support these functions (even though *many* institutions report challenges securing a sustaining budget for these Researcher-Facing roles).

5. Looking more closely at Data-Facing areas

As in 2020, some of the widest gaps between EPSCoR and non-EPSCoR institutions are seen in the Data-Facing Areas (see Fig. 9). While gaps are more modest in the areas of *Data Creation* and *Data Policy Compliance*, the following areas show a much lower average coverage in EPSCoR institutions: *Data Discovery and Collection*; *Data Analysis*; *Data Visualization*; and particularly, support for *Security/Sensitive Data*.

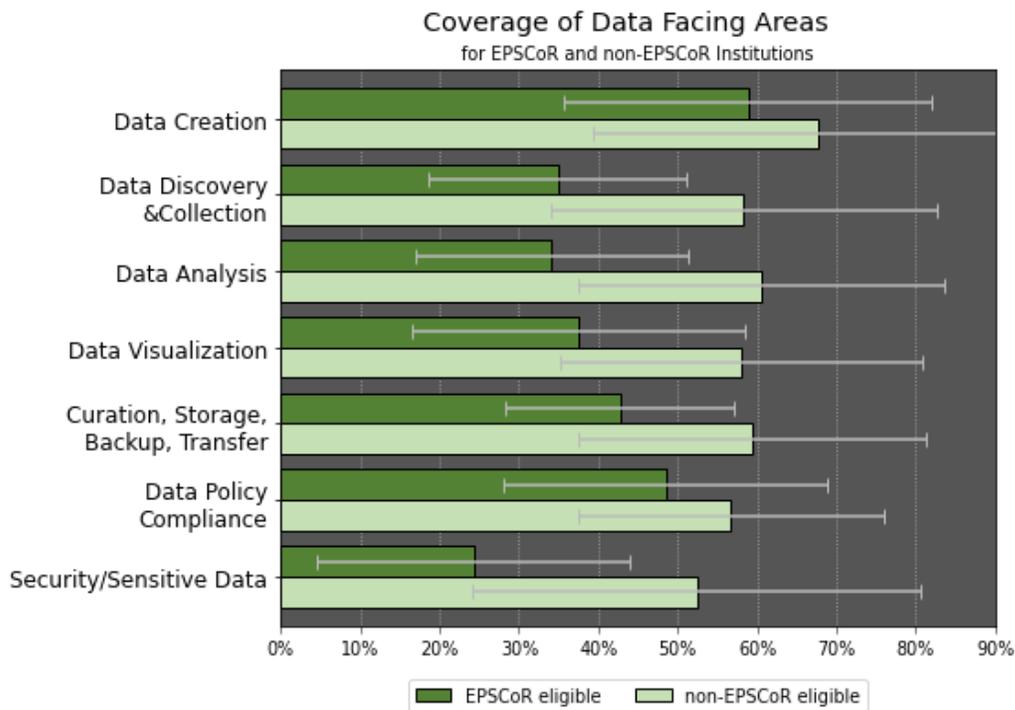


Figure 9: Data-Facing Areas

Although these patterns largely echo those seen in the 2020 dataset, the combined 2020/2021 averages have moved somewhat. The gap for *Data Analysis* widened somewhat as the average coverage in non-EPSCoR institutions rose, while the gap in *Data Visualization* narrowed with a much higher average coverage value among EPSCoR institutions (due both to the new weighting model as well as to increased coverage). While the widest gap was seen in *Security/Sensitive Data*, it actually narrowed considerably relative to 2020 (again, the EPSCoR average rose due to a combination of the new model and to increased coverage).

6. Looking more closely at Software-Facing areas

In the 2020 data, average coverage for EPSCoR institutions was somewhat higher than that of their non-EPSCoR peers in the Software-Facing areas of *Software Package Management* and *Research Software Development*, while coverage was nearly the same for *Software Optimization and Troubleshooting*. However, in the combined 2020/2021 data, EPSCoR institutions lag behind in all areas, with the widest gaps in *Workflow Engineering*; *Software Portability, Containers, and Cloud*; and *Securing Access to Software* (see Fig. 10).

Note that there was an error in the 2020 report for the area: *Software for Physical Specimen Management*, for which the report described a median of zero for EPSCoR institutions; this was the result of a data processing error revealed in the course of using the new workflow developed for 2021.

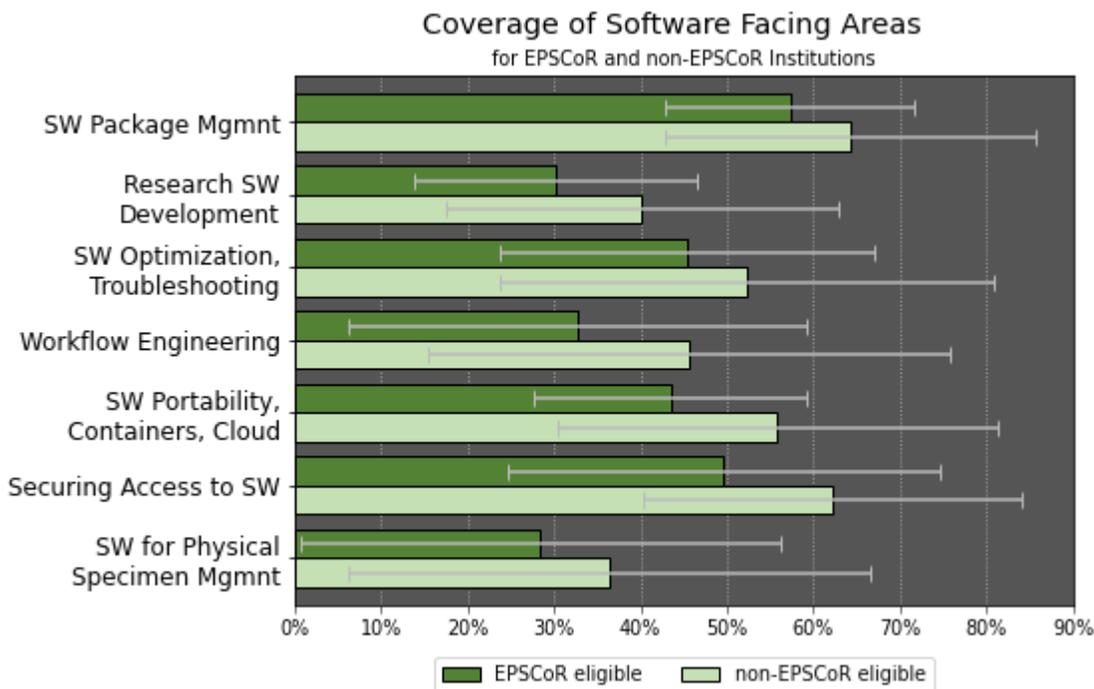


Figure 10: Software-Facing Areas

7. Looking more closely at System-Facing areas

Just as for the Software-Facing areas, average coverage in the 2020 data for EPSCoR institutions was somewhat higher than their non-EPSCoR peers in several System-Facing areas: *Infrastructure Software*, *Monitoring and Measurement*, and *Documentation*. However, in the combined 2020/2021 data, EPSCoR institutions show average coverage values well below those of their non-EPSCoR peers in all System-Facing areas, with significant gaps in *Compute Infrastructure*, *Storage Infrastructure*, *Documentation*, and *Planning* (see Fig. 11).

8. Looking more closely at Strategy and Policy-Facing areas

Average coverage values for Strategy and Policy-Facing areas do not diverge as much between EPSCoR and non-EPSCoR institutions as they do among other facings, although EPSCoR schools still lag somewhat behind (see Fig. 12). The gaps narrowed in most areas between 2020 and the 2020/2021 datasets, with the exception of *Funding*, which was nearly the same in 2020 but remains close in the 2020/2021 data. The biggest gap is in the area of *Institutional Culture for Research Support*, which may be associated with the pattern of lower research funding in EPSCoR jurisdictions.

Coverage of System Facing Areas
for EPSCoR and non-EPSCoR Institutions

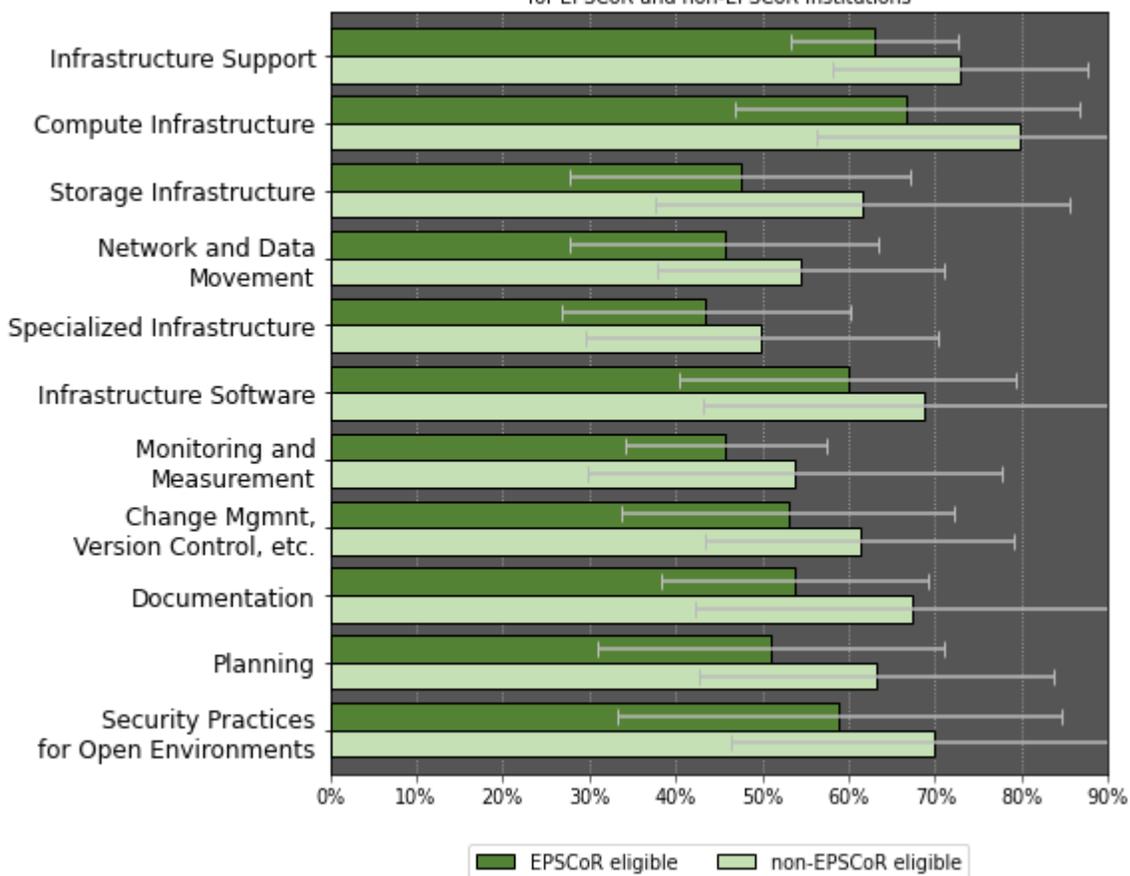


Figure 11: System-Facing Areas

Coverage of Strategy and Policy Facing Areas
for EPSCoR and non-EPSCoR Institutions

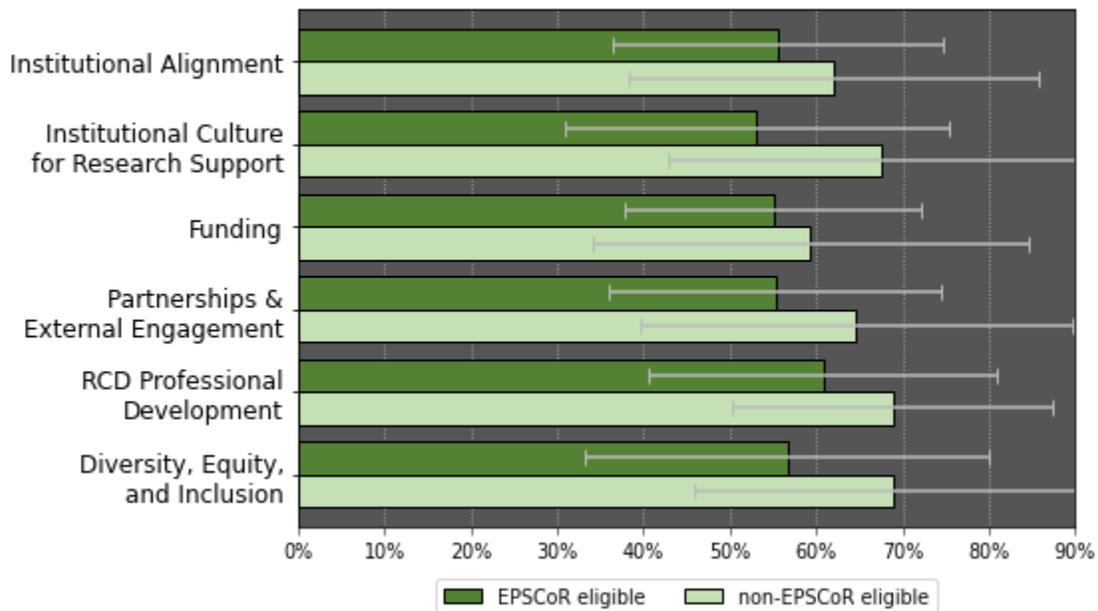


Figure 12: Strategy and Policy-Facing Areas

Interestingly, the narrowest gap is in the area of *Funding*. This area included questions about whether services are funded sustainably, as well as how funding opportunities are identified, pursued, and coordinated. For EPSCoR Institutions, averages for individual *Funding* questions range from five to eight percentage points below those of their non-EPSCoR peers, with one exception:

“Do Research Computing and Data (RCD) services groups/teams submit (extramural) grant proposals for RCD investments and innovations?”

Here, the EPSCoR average is 64% compared with the average of 60% for non-EPSCoR schools³. Could it be that EPSCoR funding opportunities provide an incentive to RCD groups to engage directly in grant proposals relatively more than their non-EPSCoR peers? Or could this indicate that RCD groups at EPSCoR institutions are taking greater initiative in pursuing funding opportunities? Or is it just that among the contributing institutions, RCD staff are more likely to have Principal Investigator (PI) status, which (at least anecdotally) is associated with more funding proposals for RCD infrastructure? In any case, it does seem to indicate that RCD groups at EPSCoR institutions are taking initiative to address the funding challenges they face. While the difference is quite small (and not statistically significant), it *is* among the few outliers in which EPSCoR average coverage is higher, and thus merits consideration.

9. Looking at Priorities

We updated the methodology for aggregating priorities this year, to account for the fact that some institutions marked only a few priorities and others marked most or all rows as a priority. Each topic could be marked as either a *High Priority* or *Medium Priority*, and at the outset we assigned two priority points for each *High Priority* and one priority point for each *Medium Priority*. However, we allowed up to 10 topics across the assessment to be marked as a priority, and if more than 10 were marked, we scaled the priority points, multiplying them by 10 divided by the number of total topics marked as priorities. We then summed the priority points for each question, and sorted to get the top priorities. We present both the overall top list, as well as the top five in each Facing in Tables 1-6 below.

Priority Points	Topic/Question	In Facing
1.946	Do researchers have access to consulting on data lifecycle requirements during data creation (e.g., anticipating metadata, storage, reuse, publisher requirements, funder requirements)?	Data-Facing
1.915	Do researchers have access to resources (e.g., staff) who can develop software for wide usage? (including websites, portals, etc.)	Software-Facing
1.700	Do researchers have access to support for research software package compilation and installation?	Software-Facing
1.621	Does your Research Computing and Data (RCD) team/group have a strategic plan? i. Is this strategic plan updated on a regular basis (e.g., annually, semi-annually)?	Strat/Policy-Facing
1.562	Do researchers have access to consulting and expertise on data wrangling/manipulation and data analysis?	Data-Facing
1.558	Do researchers have access to support, facilitation or training on how to compile, install, and deploy research software (e.g. The Carpentries, documentation on how to install and deploy anaconda environment, etc.)?	Software-Facing
1.518	Do researchers have access to dedicated resources (e.g., staff) who can perform data wrangling/manipulation and data analysis?	Data-Facing
1.518	To what extent is there a clear vision, effective guidance, and strategy for the allocation and prioritization of support resources/personnel?	Researcher-Facing
1.496	Do researchers have access to introductory user support and training related to the use of research computing and data resources available at local, regional, and national level? I.e., are there researcher-facing engagement and support staff who provide this?	Researcher-Facing

³ Remember that this is *not* the percentage of institutions for which the answer is yes, but rather the computed coverage value for this topic (considering the breadth of support/adoption across campus, the robustness of support for this activity, etc.).

Priority Points	Topic/Question	In Facing
1.470	Do researchers have access to resources (e.g., staff) who can develop research software? E.g., staff to be written into grants to architect or develop specific research applications or workflow components.	Software-Facing
1.463	Can researcher-facing staff effectively serve as advocates for the research community to leadership and IT governance?	Researcher-Facing

Table 1 - Overall top Priorities for contributing EPSCoR Institutions

Among contributing EPSCoR institutions, 14 of the 15 marked priorities – a much higher proportion relative to non-EPSCoR institutions in which only 20 of the 36 marked priorities. A total of 1,931 topics were marked as priorities among the 51 contributing institutions for an average of about 38 per institution (but a median of only seven, reflecting the impact of some institutions marking all or nearly all topics as priorities). The median number of topics marked by the 15 contributing EPSCoR institutions was quite high, at 73.

Table 1 includes the top 11 priorities, as numbers 10 and 11 were so close in priority points (after which there was a marked drop to number 12). It is worth noting that there are no System-Facing priorities in this top set, and as Table 5 shows, even the top five topics had relatively few priority points. Overall, only four of the top priorities for the 2020 contributing institutions were still in the top in the combined 2020/2021 list, and while those were numbers one through four in 2020, they have all slipped down the list somewhat in the current dataset (cf. [Appendix IV](#) for the 2020 priorities). Four of the new entrants in the top list are Software-Facing topics, two are Researcher-Facing topics, and the top priority is a new Data-Facing topic. The number two priority for EPSCoR institutions in 2020 was “*Are Research Computing and Data services funded in a sustainable manner?*”, which is not even in the top 11 this year.

While three of the top five Researcher-Facing priorities are central to work with researchers, the top one is closely connected to strategic planning and resource allocation (see Table 2). The fifth will resonate with many RCD leaders who are concerned about recruiting and (especially) retaining top talent. All five of these are in the general area of Research Staffing, clearly reflecting the challenges in this area.

Priority Points	Researcher-Facing Topic/Question
1.518	To what extent is there a clear vision, effective guidance, and strategy for the allocation and prioritization of support resources/personnel?
1.496	Do researchers have access to introductory user support and training related to the use of research computing and data resources available at local, regional, and national level? I.e., are there researcher-facing engagement and support staff who provide this?
1.463	Can researcher-facing staff effectively serve as advocates for the research community to leadership and IT governance?
1.312	Do researcher-facing staff have the skills and capacity to broadly support researchers across levels (graduate students to PIs) and across domains with information about the use and effectiveness of new technologies?
1.097	Are researcher-facing staff provided with professional development and networking opportunities?

Table 2 - Top Five Researcher-Facing Priorities for contributing EPSCoR Institutions

It is interesting that two of the top five Data-Facing priorities are focused on data wrangling (see Table 3), perhaps reflecting an increase in the number of researchers working with large amounts of data, with which they need professional staff support. The fourth Data-Facing priority has a connection to the overall #2 as a software capability.

Priority Points	Data-Facing Topic/Question
-----------------	----------------------------

1.946	Do researchers have access to consulting on data lifecycle requirements during data creation (e.g., anticipating metadata, storage, reuse, publisher requirements, funder requirements)?
1.562	Do researchers have access to consulting and expertise on data wrangling/manipulation and data analysis?
1.518	Do researchers have access to dedicated resources (e.g., staff) who can perform data wrangling/manipulation and data analysis?
1.339	Do researchers have access to resources (e.g., staff) to develop software supporting data discovery and collection? E.g.,: i. resources to develop software for collection (crawling/scraping/etc.). ii. resources to develop user interfaces or do web development to collect and interact with data with appropriate security protocols and policies.
1.194	Do researchers have access to consulting and expertise on data visualization?

Table 3 - Top five Data-Facing Priorities for contributing EPSCoR Institutions

The top Software-Facing topic by a wide margin is support for general software development; it might surprise some HPC leaders that this is a much higher priority than traditional Software-Facing issues related to cluster use (numbers two and three)⁴. Although quite a bit lower in priority points, the fifth priority reflects an increasing interest in cloud computing (see Table 4; this was also the seventh Software-Facing topic for all institutions).

Priority Points	Software-Facing Topic/Question
1.915	Do researchers have access to resources (e.g., staff) who can develop software for wide usage? (including websites, portals, etc.)
1.700	Do researchers have access to support for research software package compilation and installation?
1.558	Do researchers have access to support, facilitation or training on how to compile, install, and deploy research software (e.g. The Carpentries, documentation on how to install and deploy anaconda environment, etc.)?
1.470	Do researchers have access to resources (e.g., staff) who can develop research software? E.g., staff to be written into grants to architect or develop specific research applications or workflow components.
0.806	Do researchers have access to guidance or training for cloud computing? This can include: i. Local private campus cloud infrastructure ii. National (e.g., XSEDE-supported) cloud infrastructure iii. Commercial cloud platforms (e.g., AWS, Azure, GCP, etc.)

Table 4 - Top Five Software-Facing Priorities for contributing EPSCoR Institutions

The most striking thing about System-Facing priorities is how low they ranked overall, indicating that this area is relatively well covered at most institutions. Nevertheless, the highest priority is again related to cloud computing support, with three of the four remaining topics related to data movement and management (see Table 5). As noted above, this may reflect a growing need to support researchers with large amounts of data, and sensitive/controlled data, in particular.

Priority Points	System-Facing Topic/Question
0.784	Are there institutional resources for leveraging commercial cloud services for research computing and researchers?

⁴ It is worth noting that for the full 2020/21 dataset (all institutions) this was the third highest priority, behind the two more cluster-related capabilities.

Priority Points	System-Facing Topic/Question
0.729	Do researchers have access to mechanisms for isolated and secure support for movement of sensitive/secure data?
0.693	Do researchers have access to support for high performance data movement with dedicated data transfer nodes (DTN) and associated data movement software such as Globus, FDT, BSCP, or rclone, among others?
0.660	Is performance data (on resources that support research) analyzed and used for operational decision making?
0.660	Do researchers have access to mechanisms for isolated and secure support for storage of sensitive/secure data?

Table 5 - Top Five System-Facing Priorities for contributing EPSCoR Institutions

The top Strategy and Policy-Facing priority by a wide margin is strategic planning, with strategic alignment to campus a distant second (see Table 6). The topic of funding that was second overall for EPSCoR institutions in the 2020 report, would have been fifth in 2020 with the revised weighting method, and is therefore a *higher* priority this year. In addition, the topics four and five are also closely related to funding.

Priority Points	Strategy and Policy-Facing Topic/Question
1.621	Does your Research Computing and Data (RCD) team/group have a strategic plan? i. Is this strategic plan updated on a regular basis (e.g., annually, semi-annually)?
0.918	Is your Research Computing and Data (RCD) strategic plan aligned to campus plans? i. Does the RCD plan connect/relate to an institutional (campus-wide) strategic IT plan? ii. Does the RCD plan connect/relate to an overall institutional (campus-wide) strategic plan?
0.783	Are Research Computing and Data services funded in a sustainable manner? E.g., i. Is there recurring program budget for the staff and services operations (i.e., not primarily dependent upon grants or other non-recurring funding)? ii. Are campus funding partnerships formalized with an MOU or equivalent agreement? iii. For activities funded from contracts and grants, is there a strong track-record of renewed funding?
0.704	Does institution-level management and planning recognize and value the impact of Research Computing and Data (including return/value on investment)? E.g., are research computing and data services valued at the same level as or higher than other enterprise services when discussing of prioritization of campus (budget) resources?
0.704	Are new funding opportunities proactively identified and assessed at an institutional level, for relevance to institutional mission and alignment to Research Computing and Data needs and priorities?

Table 6 - Top 5 Strategy and Policy-Facing Priorities for EPSCoR Institutions

10. Looking at outliers (low and high) among the data

Appendices I to III present some notable contrasts between EPSCoR and non-EPSCoR institutions, and within the EPSCoR average coverage values.

Appendix I presents those Capabilities Model topics for which the average coverage for contributing EPSCoR institutions is far below (i.e., < 50% of) the average for contributing non-EPSCoR institutions. Notably, the gaps all occur in the Data-Facing and System-Facing areas. That the majority of gaps are in the Data-Facing topics aligns with the overview statistics presented in Figure 5 above, and yet only a few of the widest gaps are closely related to the Data-Facing priorities. Similarly, while there are a number of these wide gaps in the System-Facing topics, only

the question “*Do researchers have access to mechanisms for isolated and secure support for movement of sensitive/secure data?*” is associated with a high priority.

Appendix II presents Capabilities Model topics for which the average coverage for contributing EPSCoR institutions is more than one standard deviation below the overall average coverage (for contributing EPSCoR institutions). Also of note, none of these outlying low-coverage values are associated with Strategy and Policy-Facing topics, and while patterns otherwise align to the other visualizations above (in Figures 1-8) a great majority are not associated with top priorities. This would seem to indicate that institutions are not simply identifying areas of low coverage for attention, but rather are marking priorities according to other criteria (presumably in response to the strategic needs of their researchers). Taken together with the changes in average coverage values for Strategy and Policy-Facing topics discussed in Section 8, this seems to indicate that the RCD Capabilities Model is being effectively used in strategic planning practices.

Finally, **Appendix III** presents Capabilities Model topics for which the average coverage for contributing EPSCoR institutions is higher than the average for contributing non-EPSCoR institutions. Aside from the quite low number of these cases, it is notable that the vast majority occur in System-Facing topics. In particular, the topic with the highest relative coverage is “*Is there a practice in place utilizing active network measurement tools (i.e., perfSONAR) for the research-supporting network, DMZ, DTN etc.*” Does this reflect a particular emphasis or initiative to deploy these tools in support of data movement, or are there other factors likely contributing here? Another notable outlier topic is “*Do Research Computing and Data (RCD) services groups/teams submit (extramural) grant proposals for RCD investments and innovations?*” which is discussed in Section 8 above.

11. Acknowledgements

We would like to thank the members of the CaRCC EPSCoR CI Working Group for their input and review of this report: Dana Brunson (Internet2), Deborah Dent (Jackson State), Doug Jennewein (ASU), Fred Harris (U NV Reno), Gwen Jacobs (U Hawaii), Patrick Schmitz (Semper Cogito), Pips Veazey (U Maine), Scotty Strachan (Nevada System of Higher Education), Venice Bayrd (Montana State), and Zach Byerly (Louisiana State University). We would also like to acknowledge the input and support of the RCD Capabilities Model Working Group members.

This work is supported in part by National Science Foundation (OIA) [Award 2033483](#), [Award 2033519](#), and [Award 2033514](#), 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 CaRCC 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.

This paper is provided under CC-BY Attribution 4.0 International License. See <https://creativecommons.org/licenses/by/4.0/> for what this means.

12. References

- 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>
- Hill, P. (2012). 2030 Report.pdf. <https://www.nsf.gov/od/oia/programs/epscor/2030%20Report.pdf>
- 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>
- 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>

Appendix I: EPSCoR Low Coverage topics relative to non-EPSCoR

Topics/Questions for which the EPSCoR average coverage is less than 50% of the average non-EPSCoR value. The first two columns have the average coverage values for the topic, for contributing EPSCoR and non-EPSCoR institutions, respectively. The “Relative %” column is the relative percentage of EPSCoR average coverage compared to the non-EPSCoR average coverage.

EPSCoR	non-EPSCoR	Relative %	Capability Topic
Data-Facing			
20.3%	53.3%	38.0%	Do researchers have access to compute and data environments to manage and use “notice triggering” data (e.g., PHI, HIPAA, Export Control, licensed data)? This can include: i. tools, systems, and environments that can scale from small to large data sets (i.e., up to high performance computing). ii. data security protocols in use, and monitored.
24.1%	57.1%	42.2%	Do researchers have access to dedicated resources (e.g., staff) who can perform data wrangling/manipulation and data analysis?
19.3%	43.2%	44.8%	Do researchers have access to resources (e.g., staff) for software development of tools that support data visualization?
18.3%	40.6%	44.9%	Do researchers have access to compute and data environments to manage and use extremely sensitive data (e.g., requiring cold room/air-gapped storage and computing, closely monitored access, dedicated data stewardship etc.)?
25.8%	56.1%	45.9%	Do researchers have access to dedicated resources (e.g., staff) who can perform data wrangling/manipulation and data analysis?
20.7%	44.2%	46.8%	Do researchers have access to resources (e.g., staff) for software development of tools that support data wrangling/manipulation and data analysis?
28.7%	60.1%	47.7%	Do researchers have access to expertise about common Terms of Service for frequently crawled websites/data repositories and best practices guidance? E.g.,: i. library or other staff with knowledge about common Terms of Service for frequently crawled websites/data repositories and best practices guidance? ii. library or other staff with skills and capacity to inform policies and educate researchers on data use agreements (DUAs)?
System-Facing			
6.2%	29.7%	20.7%	Is there a practice in place for whole system testing (e.g./chaos monkey) on resources that support research?
20.2%	52.1%	38.7%	Do researchers have access to specialized compute capability (e.g. bare metal hardware, reconfigurable BIOS, OS, and network, or experimental cloud testbeds)?
16.1%	36.7%	43.8%	Do researchers have access to support for edge computing and data resources?
21.8%	46.4%	46.9%	Do researchers have access to mechanisms for isolated and secure support for movement of sensitive/secure data?
15.6%	31.4%	49.8%	Do researchers have access to infrastructure for data buffering between high I/O lab instruments and the data center, and/or external resources (“data capacitors” or “burst buffers”)?

Appendix II: EPSCoR Low Coverage topics relative to EPSCoR mean

Topics/Questions for which the EPSCoR average coverage is less than 1 standard deviation below the overall EPSCoR mean coverage (i.e., the mean for all topics, 47.9%; for comparison, the non-EPSCoR mean is 60.4%).

EPSCoR Avg	Question or Topic
Researcher-Facing	
22.9%	Does your institution have a process to assess researcher awareness, satisfaction, and engagement related to Research Computing and Data services and support?
Data-Facing	
18.8%	Do researchers have access to resources (e.g., staff) who will develop tools/software that supports data backup, storage, and integrity checking?
19.3%	Do researchers have access to resources (e.g., staff) for software development of tools that support data visualization?
20.7%	Do researchers have access to resources (e.g., staff) for software development of tools that support data wrangling/manipulation and data analysis?
23.7%	Has your institution defined and deployed a process for identifying which research data to archive, preserve, or discard?
18.3%	Do researchers have access to compute and data environments to manage and use extremely sensitive data (e.g., requiring cold room/air-gapped storage and computing, closely monitored access, dedicated data stewardship etc.)?
20.3%	Do researchers have access to compute and data environments to manage and use “notice triggering” data (e.g., PHI, HIPAA, Export Control, licensed data)? This can include: i. tools, systems, and environments that can scale from small to large data sets (i.e., up to high performance computing). ii. data security protocols in use, and monitored.
Software-Facing	
13.7%	Do researchers have access to usability testing for research software developed on campus?
20.0%	Do researchers have access to support for research workflow packages (e.g., Toil, Pegasus, NextFlow)?
20.4%	Do researchers have access to resources (e.g., staff) for software development for discovery and research use of physical collections?
22.3%	Do researchers have access to security validation for research software? (e.g., analysis for vulnerabilities that can be exploited by hackers, especially for locally developed software)?
System-Facing	
6.2%	Is there a practice in place for whole system testing (e.g./chaos monkey) on resources that support research?
15.6%	Do researchers have access to infrastructure for data buffering between high I/O lab instruments and the data center, and/or external resources (“data capacitors” or “burst buffers”)?
16.1%	Do researchers have access to support for edge computing and data resources?
18.8%	Do researchers have access to virtualized networking techniques such as Software Defined Networks, overlays, etc.?
20.2%	Do researchers have access to specialized compute capability (e.g. bare metal hardware, reconfigurable BIOS, OS, and network, or experimental cloud testbeds)?
21.8%	Do researchers have access to mechanisms for isolated and secure support for movement of sensitive/secure data?

Appendix III: EPSCoR High Coverage topics relative to non-EPSCoR

Topics/Questions for which the EPSCoR average coverage is higher than that for non-EPSCoR institutions.

EPSCoR Avg Cover.	non-EPSCoR Avg Cover.	Relative	Question or Topic
System-Facing			
68.3%	54.8%	125%	Is there a practice in place utilizing active network measurement tools (i.e. perfSONAR) for the research-supporting network, DMZ, DTN etc.?
73.1%	63.7%	115%	Do researchers have access to support for researcher workstations or laptops?
70.7%	63.4%	111%	Do researchers have access to support for special science instruments (e.g. cryo EM, DNA sequencer, telescope, etc)?
40.0%	36.7%	109%	Have Research Computing and Data staff established a workflow environment to support end-to-end network performance troubleshooting?
87.1%	81.5%	107%	Do researchers have access to a high-performance network that supports research within campus?
73.0%	68.8%	106%	Do researchers have access to a Science DMZ (a means to securely enable high performance inter-campus data flows that bypass campus firewalls)?
Strategy and Policy-Facing			
64.1%	59.8%	107%	Do Research Computing and Data (RCD) services groups/teams submit (extramural) grant proposals for RCD investments and innovations?
52.3%	51.0%	103%	Are Research Computing and Data staff provided with opportunities for career advancement? i. Are there clear career paths defined for each role? ii. Are staff encouraged/supported in pursuing career advancement opportunities?

Appendix IV: 2020 Top Priorities for EPSCoR Institutions

Priorities as marked by EPSCoR in the 2020 dataset, but (re)ranked using the new weighting method that accounts for the number of priorities marked.

Priority Points	Topic/Question	In Facing
2.230	Does your Research Computing and Data (RCD) team/group have a strategic plan? i. Is this strategic plan updated on a regular basis (e.g., annually, semi-annually)?	Strat/Policy-Facing
1.663	Do researchers have access to consulting and expertise on data wrangling/manipulation and data analysis?	Data-Facing
1.661	Do researchers have access to dedicated resources (e.g., staff) who can perform data wrangling/manipulation and data analysis?	Data-Facing
1.278	To what extent is there a clear vision, effective guidance, and strategy for the allocation and prioritization of support resources/personnel?	Researcher-Facing
1.230	Are Research Computing and Data services funded in a sustainable manner?	Strat/Policy-Facing
1.196	Do researchers have access to guidance or training for cloud computing? This can include: i. Local private campus cloud infrastructure ii. National (e.g., XSEDE-supported) cloud infrastructure iii. Commercial cloud platforms (e.g., AWS, Azure, GCP, etc.)	Software-Facing
1.111	Are researchers made aware of research computing and data related resources? E.g.: i. intra-campus resources (e.g., support, training, engineering, central IT services, library services, related centers or institutes) ii. cross-institution, regional, national, and/or international entities that comprise the larger ecosystem of Research Computing and Data (e.g., ACI-REF, Campus Champions, Research Software Engineers, CASC, CaRCC, CI Engineers, PEARC)?	Researcher-Facing
0.999	Is your Research Computing and Data (RCD) strategic plan aligned to campus plans? i. Does the RCD plan connect/relate to an institutional (campus-wide) strategic IT plan? ii. Does the RCD plan connect/relate to an overall institutional (campus-wide) strategic plan?	Strat/Policy-Facing
0.992	Do researchers have access to consulting and expertise to help them identify appropriate data repositories (on campus, in domains, and more generally) to place their data?	Data-Facing
0.959	Do researchers have access to consulting and expertise on data visualization?	Data-Facing
0.959	Do researchers have access to software that supports data visualization?	Data-Facing