Membership and Participation in our RCD Communities: What is it and how are we doing?

Robert M Freeman Jr
rfreeman.rcd@gmail.com
Harvard Business School
Boston, Massachusetts, USA

Kirk M. Anne
kirk.m.anne@gmail.com
Rochester Institute of Technology
Rochester, New York, USA

Gabriel King Smith
gabriel.smith@ubc.ca
University of British Columbia
Vancouver, British Columbia, Canada

Wei Yin
wy2288@columbia.edu
Columbia University
New York, New York, USA

Gil Speyer
speyer@asu.edu
Arizona State University
Tempe, Arizona, USA

Chris Reidy
chrisreidy@arizona.edu
University of Arizona
Tucson, Arizona, USA

ABSTRACT

The Research Computing and Data (RCD) community has coalesced over the past ten years to encompass hundreds of organizations that support both researchers and research support staff alike. While many of these organizations may rely on external funding, definitions of membership vary considerably, and their goals may include broadening participation, increasing diversity and inclusion, and performing outreach to encourage those besides "the usual suspects" to get involved. In addition, silent or absent audience members – ones who are minimally or not at all engaged – are easily overlooked. This preliminary work addresses a need for tools to help an organization know its membership, to characterize the depth of participation and engagement, and to identify and measure any untapped potential as part of its mission to maximize the capabilities of its community. We apply this approach to characterize and understand the Campus Research Computing Consortium (CaRCC) People Network community, both the membership and participation groups, including representation and diversity over time. We then further highlight those more deeply engaged via multiple approaches across various CaRCC activities. A "first draft" in developing a common tool set, we hope these methods will be adopted and improved upon by the larger RCD community.

CCS CONCEPTS

• **Social and professional topics** → **Computing organizations**;
• **General and reference** → **Metrics**.

KEYWORDS

RCD community, membership, participation, engagement, CaRCC, People Network

ACM Reference Format:
Robert M Freeman Jr, Kirk M. Anne, Gabriel King Smith, Wei Yin, Gil Speyer, and Chris Reidy. 2024. Membership and Participation in our RCD Communities: What is it and how are we doing? . In *Practice and Experience in Advanced Research Computing (PEARC '24), July 21–25, 2024, Providence, RI, USA.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3626203.3670630

PEARC '24, July 21–25, 2024, Providence, RI, USA

ACM ISBN 979-8-4007-0419-2/24/07
https://doi.org/10.1145/3626203.3670630

1 INTRODUCTION

The Campus Research Computing Consortium (CaRCC) is an organization of research computing and data (RCD) professionals that builds and fosters an inclusive community that shares knowledge and resources across the RCD community[2]. In existence since 2018, the flagship activity is the People Network: CaRCC's People Network "aims to foster, build and grow an inclusive community (termed the 'People Network') for campus professionals who support research computing and data resources and the researchers who use them. The Network includes synchronous and asynchronous opportunities to network with and leverage the collective expertise of the broader RCD+ community"[3]. It is composed of five different affinity groups, Tracks or "-Facings" that focus on specific interests and reflect the type of works one "faces" daily[5]. Through various activities within each Track and the People Network as a whole, CaRCC is able to work towards its mission of supporting the RCD community and fostering scientific research.

The RCD ecosystem has grown significantly in the past decade[4], and we have learned firsthand that understanding and supporting one's community is a significant challenge, not the least of which concerns membership. As each organization's membership differs, the challenge is in understanding who our participants are and knowing if we are meeting our objectives towards building community both in diversity and in breadth. Although we do not question the value of the work, we cannot help but wonder if the effort invested is well apportioned: how are individuals participating in People Network activities, if at all? And how might we quantify the different levels of participation?

Much work has been done in studying communities and fostering their development. The Community Engagement Model, for example, discusses four modes of community engagement[8]. Although ideal with particular metrics in place, we sought a first-run, simpler approach: In CaRCC's People Network, there are multiple modes of participation, and we'd like to further assess and quantify the depths. Participating via attending a monthly Track Call can be

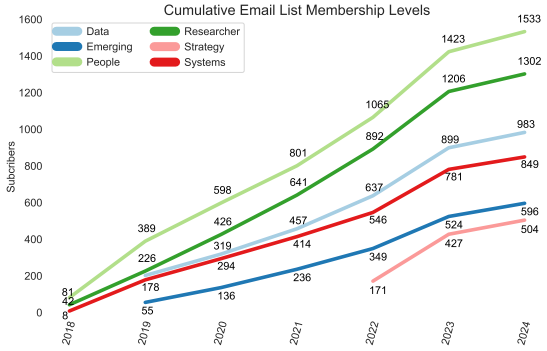

Figure 1: Cumulative People Network and Tracks Membership Counts

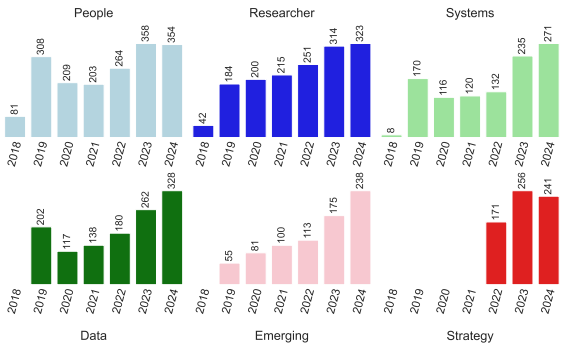

Figure 2: People Network and Tracks Membership Additions by Year

"low-effort" engagement; but this transforms into higher engagement if sustained over a period of time or leads to other (expanded) modes of participation.

Thus, we sought an initial set of measures: e.g., what is the frequency or longevity for attending Track Calls? Are members attending Calls for a single Track or multiple? And are members more fully engaged by becoming involved in other areas of CaRCC? Taken one step further, we can inquire on an organization's participation (the member's workplace institution), and assess and quantify this as well. Although beyond the rudimentary statistics communicated in annual reports to funding agencies, this work is preliminary: We fully expect to iterate on this with our stakeholders and our community, as well as with external RCD partners, taking care to ensure that we're reporting aggregated, not individual-level, data. With a solid, base set of methods to answer specific questions using various, activity-level data, it is our hope that we can inform CaRCC (and similar organizations) in its mission to increase the reach and effectiveness of our community, to maximize its potential, and to empower researchers at regional, national, and international scales.

2 MEMBERSHIP

2.1 Overview

For CaRCC's People Network, members can choose how they participate[3]. Reflecting the role-based "-Facings"[5], each Track has its own email list and leadership (coordinators and steering committee), with focused discussion happening in each Track's monthly community calls[4]. The current CaRCC Tracks are (in decreasing age) Researcher-Facing ("researcher"), Systems-Facing ("systems"), Data-Facing ("data"), Emerging Centers ("emerging"), and Strategy- & Policy-Facing ("strategy"). People Network and Track members are those who signed up for one or more mailing lists.

For each Track, we collected CSV downloads though 2024Q1 (3/2024) for the Google Group email distribution lists which contained name, email address, and date registered; bounced or deleted emails were noted though excluded from further work. For institutional designation, non-Gmail addresses were truncated to the top-level subdomain (e.g. g.college.edu ⟹ college.edu). For Gmail-based addresses, emails were cross-referenced against the membership sign-up list[3]; some members may have been added directly after attending a Track Call, though this practice was inconsistent across Track and over time. The membership data were joined with IPEDS[6] directory information based on the top-level subdomain names, providing institutional characteristics like state location, Carnegie ranking, and size.

2.2 Individual

Figure 1 is a running total of the membership for the People Network and its Tracks, with the People Network having 1533 members by the end of first quarter of 2024; across all five Tracks, there were 4234 membership entries, or 1762 unique members. The low count for the People Network as compared to the total unique members is likely the result of direct adds to Track-specific lists; we expected these counts to be higher, as it appears that Google may prune bounces and deletions. If we analyze this yearly per Track, shown in Figure 2, we see steady growth with strong growth typically in the first year or so. As expected, we saw some limited growth in 2020 due to the COVID pandemic, with some Tracks being more resilient. Please see our Zenodo post for additional membership sign-up patterns and other analyses[1].

2.3 Institutional

Using the institutional email domain information, we characterized the CaRCC's People Network through various lenses. Supplemental Table 1 shows per year the unique number of email domains that have been added to the membership ("Total" is a sum of the columns). Although not showing the accumulation of unique domains across all years, each Track and the People Network is growing in breadth. In Table 1 we see the most frequent institutions (email domains) across the People Network. This is likely expected due to sizes of the schools. Masking out R1s, we can see what other top research institutions are members in Table 2. This suggests prospective outreach opportunities. It would also be interesting to contrast these numbers against school size or some other characteristic to identify and then understand outliers.

Supplemental Tables 2 and 3, and Table 3 show further understanding of the characteristics of the People Network and Tracks. We still have much to learn from this, though it is encouraging to see increases in Established Program to Stimulate Competitive Research (EPSCoR) representation and a diverse set of institution types.

Moreover, as seen in Figure 3 and the supplemental IPEDS tables, regionally, CaRCC's People Network membership appears to cover most of the US for R1 institutions and "reasonable" coverage, including members outside the United States (e.g. Canada).

Domain	Data	Emerging	People	Researcher	Strategy	Systems	Total
harvard.edu	36	10	58	39	19	32	194
wisc.edu	25	6	34	24	7	20	116
indiana.edu	17	14	28	27	13	14	113
northwestern.edu	24	6	26	23	10	12	101
ucla.edu	15	10	29	28	6	12	100
illinois.edu	15	4	31	25	9	13	97
psu.edu	18	9	25	21	10	13	96
purdue.edu	14	8	28	24	9	11	94
ucsd.edu	16	7	23	18	5	17	86
rutgers.edu	15	6	26	21	4	10	82

Table 1: Top Institutions Overall via Unique Domains for Email Membership

Domain	Data	Emerging	People	Researcher	Strategy	Systems	Total
gmu.edu	10	7	12	14	3	10	56
clemson.edu	5	5	8	7	4	5	34
uvm.edu	4	5	7	6	3	6	31
chapman.edu	4	5	5	5	4	4	27
njit.edu	4	4	4	6	3	4	25
uw.edu	4	1	7	8	3	2	25
ohsu.edu	4	4	5	4	2	4	23
uri.edu	5	3	4	7	0	4	23
boisestate.edu	5	3	5	3	3	4	23
sdsu.edu	3	2	5	4	2	5	21

Table 2: Top non-R1 Institutions via Unique Domains for Email Membership

	2018	2019	2020	2021	2022	2023	2024
EPSCoR	8	49	34	37	27	43	50
non-EPSCoR	82	375	253	242	371	405	455
Total	90	424	287	279	398	448	505

Table 3: EPSCoR vs non-EPSCoR Members Added by Year

3 PARTICIPATION

3.1 Overview

Again, each of the People Network Tracks hold separate, monthly calls[3]. For each Track Call, attendees are asked to sign-in to a Call Document ("Call Doc") – a Google Doc that includes call information, notes, questions, and resource URLs – and voluntarily record their name, institution, and email. The maximum attendees number on the Call and attendee sign-in count is recorded on an internal "tracking" document.

From this document, we gathered and cached the 170 Call Docs (from 1/2018 through 3/31/24, excluding Joint, Party, or Plenary Calls), recorded the stats for each Call, and extracted attendee data, relying on the Call Doc structure to guide us. To reduce variations and fill in missing data from early Calls, we created more standardized identifiers via several transforms: a) extracted the domain names to transform via WHOIS lookup to get the top-level subdomain; b) a Wikipedia lookup for institution if missing an email address; c) a manual mapping table for inconclusive institution searches; and d) a WHOIS domain name search to identify the registering organization[7]. Finally, to minimize variations of email usernames, we also created a *first@subdomain.edu* identifier, relatively consistent and unique by institution at this small population size.

After the data were cleaned, we had 7156 fully-complete and 443 incomplete records (no email, 353; no institution, 65; no domain, 177). This value was greater than the sum of Call Tracking sheet sign-in counts – indicating faithful data recovery and likely failures to record Call attendee counts. The record set was merged with the IPEDS 2022 data[6], yielding 7019 annotated and 873 unannotated attendee records (e.g. supercomputer centers, non-US institutions, and non-profit organizations). Across all Calls, we've seen 1460 unique participants.

Due to voluntary sign-ins, we verified the *capture rate* of sign-ins to Zoom attendee counts per Facing and overall. Rates >= 66% to 82% were significant enough for general analysis, which all Tracks and years achieved, as shown in Supplemental Table 5, which also includes yearly Track Call participation counts.

3.2 Individual Participation

We deemed that quantifying participation can be a proxy for determining the level of engagement, and so we examined attendee Call longevity, frequency, and variety.

Longevity of participation: For how long have persons been attending Calls? Analyzing the duration of time between first and last appears of an attendee, Figure 4a shows that ~45% of persons have been attending calls for up to a year; and ~66% for up to two years (613 and 840 persons, respectively). The corollary is that ~33% of People Network attendees have participated for three or more years, a significant fraction considering CaRCC's five-year lifespan.

Frequency of participation: How many Calls have persons attended regardless of the Track? Figure 4b, a log plot, shows an asymptotic distribution of attendees for Call attendance count, with a drop below 30 persons (~ 1500 persons total) only after 14 Calls

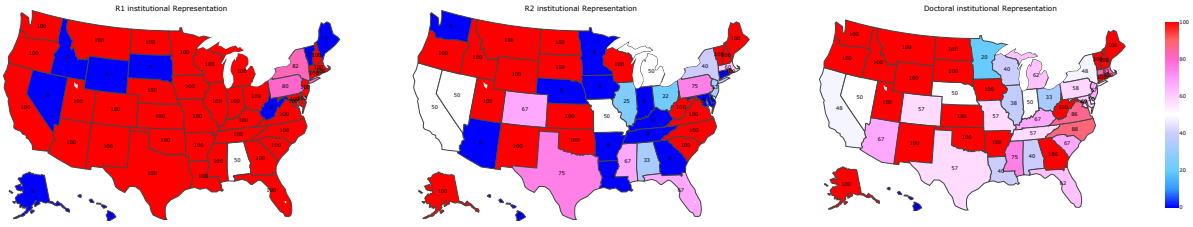

Figure 3: Percentage Membership Coverage of R1, R2, and All Doctoral Institutions

(~9% of all Calls). For reference, there are 170 total Calls, about 35 Calls per year, and about 7 Track Calls per year.

Frequency of participation across Tracks: Finally, what is the distribution of persons attending the binned number of Calls from one, two, three, or four (or more) different Tracks? In Figures 4c-f, the distributions show a significant fraction of attendees on one or two Tracks. But there is clearly a large population sampling multiple Tracks, and nearly 100 persons attending more than three Tracks (note that People Network leadership is ~50 persons). In future work we may reanalyze this with the leadership filtered out to determine any impact on engagement at these higher levels.

As CaRCC has a number of Working Groups[2] outside of the People Network, what can we tell about engagement beyond the monthly Calls? We found that 11.25% of the unique Call attendee list are currently volunteering for a CaRCC Working group; and 2.19% are currently serving in a Chair or Coordinator role. Overall, 13.44% of our unique call attendees are volunteering with CaRCC and contributing to the RCD Community.

And finally, if we compare the email membership against participation data, what does this reveal in terms of persons not attending Calls or not on the email lists? Table 4 shows that across Tracks and overall, a larger majority of persons who are members have not yet participated in a single Track Call; we see this as a huge, untapped potential of persons in the community for outreach. At this same time, after internal conversations, some Tracks may leverage other communication methods in lieu of their own Track list. Separately, we see a significant fraction of persons only attending Calls. We believe this may be a result of including other RCD organizations on our Monthly Call announcement, which may be supported by the Institutional participation data.

3.3 Institutional Participation

Although CaRCC characterizes itself on being a community of individuals, understanding institutional representation is also an important consideration. Using either the institution name or the domain name of the attendee email addresses, we looked at participation at the institutional level via analysis of the unique institutional domains. These results can be seen in Supplemental Table 6 (Calls by Unique Domains), Supplemental Table 7 (Call Participation by Institution Type), Supplemental Table 8 (Yearly Call Attendees by Region), Supplemental Table 9 (Yearly Calls by EPSCoR Designation), Supplemental Table 10 (Top Institutional Domains for Call Participation), and Supplemental Table 11 (Most Diverse Calls by Attendance). The positive signal here is that the People Network Tracks continue to attract more, new institutions with diverse characteristics to join in supporting the RCD community. We believe one of the primary drivers of this has been including other RCD community organizations on our People Network Monthly Calls announcements.

4 CHALLENGES AND FUTURE WORK

Collecting participation data was particularly challenging, primarily handling the different data sources and their formats. Reducing the various pieces of data (emails and names) into unique identifiers in order to extract information about members' institutional data was difficult. The WHOIS queries were successful (reducing host names to their base domain names), though some institutions have multiple domains requiring more coding to unify them. With hostnames of websites in IPEDS and emails reduced to base domain names, data joins were easier to perform. Additionally, when creating unique identifiers (firstname@college.edu), persons moving from one institution to another resulted in multiple entries.

As a preliminary foray, this has been our opportunity to test out methods and visualizations, and to decide as a team what works best. These we will refine after continued conversations with CaRCC leadership, Track Coordinators, and our community. And we have only "scratched the surface," as we have not yet begun to leverage other sources of data: our YouTube channel views, conversations on the email lists and in Slack posts, not to mention Zoom chat post content and frequencies and Call Doc content. We expect this preliminary work also will be helpful in assisting other CaRCC groups with surveying our community directly, one means of accessing that "untapped potential."

In summary, we feel that the new methods and analytical approaches presented here could be adopted more broadly in order to share common metrics on demographics, participation, and engagement with other organizations in the RCD community. We look to refine the analyses conducted here, investigate new ones, and share our findings once again as a means to increase our outreach efforts and more fully engage members and supporters of the RCD community.

ACKNOWLEDGMENTS

We would like to thank the members of the CaRCC Logistics and Engagement Groups, and the People Network Track Coordinators for their helpful conversations on membership, call participation,

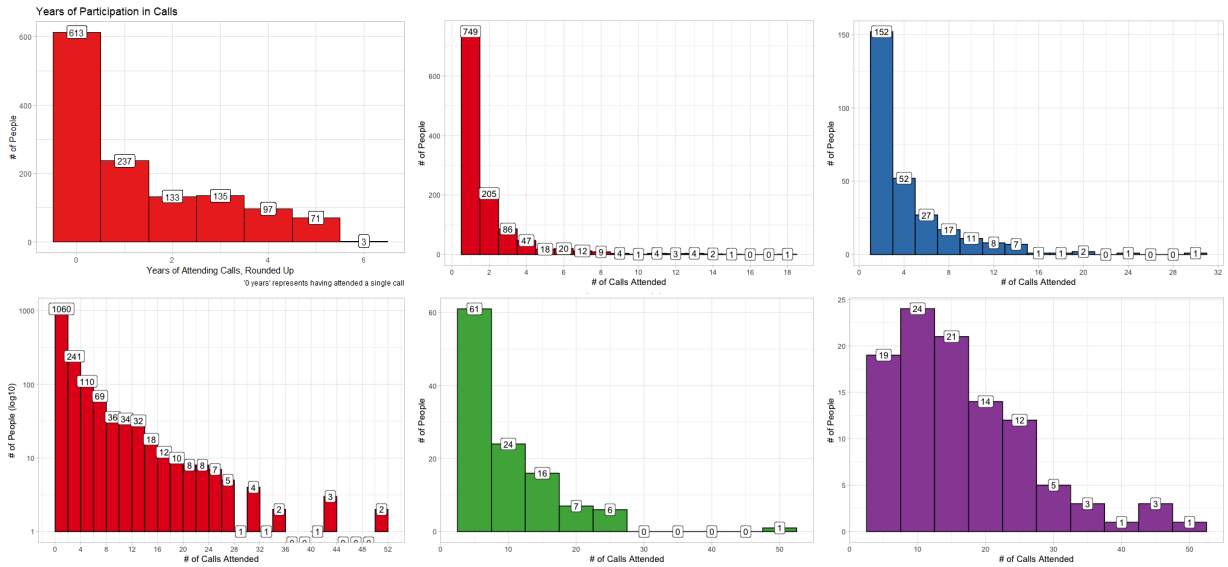

Figure 4: (a) Longevity and (b) Frequency of Call Attendees; and (c - f) Distribution of Multi-Track Attendance

Index	Any List	Data	Emerging	Researcher	Strategy	Systems
Calls Only	609 (23%)	246 (17%)	227 (23%)	249 (14%)	51 (7%)	320 (23%)
Both Calls & Lists	891 (34%)	256 (17%)	142 (14%)	598 (34%)	152 (21%)	244 (18%)
Membership Only	1133 (43%)	971 (66%)	620 (63%)	931 (52%)	516 (72%)	808 (59%)
Total	2633 (100%)	1473 (100%)	989 (100%)	1778 (100%)	719 (100%)	1372 (100%)

Table 4: Comparison of List Membership and Call Participation by Track

and engagement; and especially Dana Brunson, Sean Cleveland, Claire Mizumoto, Brian Haymore, and Patrick Schmitz for deep dives.

REFERENCES

[1] 2024. CaRCC Membership and Participation: Exploration of Analyses and Visualizations. https://doi.org/10.5281/zenodo.12341218
[2] Accessed April 25, 2024. Campus Research Computing Consortium. https://carcc.org/.
[3] Accessed April 25, 2024. People Network – CARCC. https://carcc.org/people-network/.
[4] Joel Cutcher-Gershenfeld, Torey Battelle, Dana Brunson, Thomas Cheatham, Jacob Fosso Tande, Douglas Jennewein, Julie Ma, Lauren A. Michael, Timothy Middelkoop, Henry Neeman, and Patrick Schmitz. 2023. Professionalization of Research Computing and Data: An Expanded Agenda. In *Practice and Experience in Advanced Research Computing* (Portland, OR, USA) *(PEARC '23)*. Association for Computing Machinery, New York, NY, USA, 129–136. https://doi.org/10.1145/3569951.3593610
[5] Christina Maimone, Scott Yockel, Timothy Middelkoop, Ashley Stauffer, and Chris Reidy. 2022. Characterizing the US Research Computing and Data (RCD) Workforce. In *Practice and Experience in Advanced Research Computing* (Boston, MA, USA) *(PEARC '22)*. Association for Computing Machinery, New York, NY, USA, Article 27, 7 pages. https://doi.org/10.1145/3491418.3530289
[6] U.S. Department of Education. 2022. *Integrated Postsecondary Education Data System (IPEDS)*. Washingnton, D.C. https://nces.ed.gov/ipeds/datacenter/data/HD2022.zip
[7]Richard Penman. 2024. *Whois module*.
[8] L Woodley and K Pratt. 2020. The CSCCE Community Participation Model–A framework for member engagement and information flow in STEM communities. https://zenodo.org/records/3997802.