

Characterizing the US Research Computing and Data (RCD) Workforce

Christina Maimone*

christina.maimone@northwestern.edu
Northwestern University
Evanston, IL, USA

Scott Yockel

syockel@g.harvard.edu
Harvard University
Cambridge, MA, USA

Timothy Middelkoop

tmiddelkoop@internet2.edu
Internet2
Columbia, MO, USA

Ashley Stauffer

als81@psu.edu
Penn State University
State College, PA, USA

Chris Reidy

chrisreidy@arizona.edu
University of Arizona
Tucson, AZ, USA

ABSTRACT

A growing share of computationally and data-intensive research, both inside and outside of academia, requires the involvement and support of computing and data professionals. Yet little is known about the composition of the research computing and data (RCD) workforce. This paper presents the results of a survey (N=563) of RCD professionals' demographic and educational backgrounds, work experience, current positions, job responsibilities, and views of working in the RCD field. We estimate the size of the RCD workforce and discuss how the demographic diversity and distribution of backgrounds of those in the RCD workforce fail to match that of the larger academic and technical workforces. These survey results additionally support the insights of those working in the field concerning the need to recruit a wider variety of professionals into the RCD profession, better define job descriptions and career pathways, and improve institutional recognition for the value of RCD work.

KEYWORDS

RCD, Research Computing, Cyberinfrastructure, Workforce Development, Population Survey, Professionalization

ACM Reference Format:

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 (PEARC '22), July 10–14, 2022, Boston, MA, USA*. ACM, New York, NY, USA, 7 pages. <https://doi.org/10.1145/3491418.3530289>

1 INTRODUCTION AND BACKGROUND

The practice and process of science has changed due to the diversification of research domains leveraging high performance computing (HPC) resources, an increased need to manage and move larger

*Primary Author

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

PEARC '22, July 10–14, 2022, Boston, MA, USA

© 2022 Copyright held by the owner/author(s). Publication rights licensed to ACM. ACM ISBN 978-1-4503-9161-0/22/07.
<https://doi.org/10.1145/3491418.3530289>

amounts of data, the prevalence of cybersecurity threats, and growing governmental regulations on data. Researchers, both within academia and outside of it, can no longer do it alone; they now rely on research computing and data (RCD) professionals who work at the intersection of cyberinfrastructure (CI), research, and data and provide skills built on years of experience. The demand for RCD professionals is projected to grow [16, 32, 33]. Pathways to RCD careers, however, can be “invisible” due to a lack of standardization of job titles, poorly defined job descriptions, the dispersed nature of the work across multiple units within research organizations, and inconsistent terminology to describe RCD activities [16, 33].

Despite growing need for CI expertise and acknowledgement that the RCD workforce is essential to US competitiveness in global science and engineering initiatives, a number of challenges exist that may deter professionals from entering, and staying in, the RCD field [41]. Arafune et al. [2020] discuss many of the challenges that affect the research computing innovation workforce [16]. They group these challenges, representing the perspectives of 100 RCD professionals, into several thematic areas, including: diversity and inclusivity; career paths and normative role descriptions; recruitment and onboarding into the workforce; compensation; workforce retention; the identity of the field as a discipline; and continued training and education for professionals.

There is an active community of research about RCD professionals and the ecosystem of overlapping organizations, research collaborations, and institutions that enable work in the RCD field. However, this growing field has not yet generated a descriptive characterization of the RCD workforce as a whole, instead focusing research on specific subgroups. For example, Katz et al. [2019] document the professional duties, challenges, and aspirations of 28 staff data scientists in academia [28]. Lauer [2020] surveyed the HPC workforce to examine the relationship between occupational climate factors, race, gender, and career withdrawal cognitions of 425 professionals in the field [29]. Frantzana [2019] examines women’s experiences and representation in the HPC community [25]. Sholler [2019] reports findings from forty-five interviews with researchers, research support staff, and student researchers engaged in cloud vendor enabled research [42]. Neeman et al. [2020] provide a deep understanding of those in RCD facilitator roles and the programs available to support their professional development [32]. Moving beyond individuals, Broude Geva et al. [2020] detail the organizations active in the RCD field and the connections between them,

introducing the ecosystem metaphor to characterize the current RCD environment and its interdependent and interacting components [18].

This paper fills part of the notable gap in characterizing the full US RCD workforce by reporting the results of a survey of RCD professionals conducted in 2021. We provide preliminary estimates of the workforce size, demographics, distribution across institutions, and role characteristics. This is a necessary endeavor if research institutions are to succeed in recruiting these crucial professionals who make computational and data-intensive research possible.

2 ESTIMATING THE WORKFORCE SIZE

A literature review did not discover any existing estimates of the size of the RCD workforce. Because the RCD profession is not yet well-defined, many people whose work aligns with RCD activities do not yet identify as RCD professionals or identify their work as research computing and data work [41]. This makes such individuals difficult to identify and count, and it means that those who can be identified are likely only a fraction of the total workforce.

The RCD workforce spans all economic sectors (academia, national laboratories, for-profit industry, and nonprofits), however the best sources of information we have about the extent of the workforce come from academic organizations. Academic institutions perform about half of all US basic research and 10-15% of total US research and development [44]. While a majority of research happens outside of academia, information about non-academic RCD activities, and the RCD professionals working for such organizations, is rarely shared publicly. In contrast, staff lists for large academic computing centers and university research computing groups are frequently included on their websites, universities promote information on the RCD services available to their researchers, and multiple academic RCD community email lists have evolved to spread information within the community.

Looking first at such email lists, approximate membership as of fall 2021 for the email list for the Extreme Science and Engineering Discovery Environment (XSEDE) Campus Champions was 825, Campus Research Computing Consortium (CaRCC) People Network was 1,100, Coalition for Academic Scientific Computation (CASC) was 250, EDUCAUSE Research Computing and Data Community Group was 525, and the Virtual Residency Program was 1,100.¹ The overlap in membership between these lists ranges from 7-30%; there are at least 2500 research computing professionals on at least one of these five lists.

Beyond email lists, the number of academic institutions with research computing and data services also provides an indication of the size of the workforce. There are 131 universities with a Carnegie Classification of R1 (Doctoral universities with very high research activity) and 135 with a classification of R2 (Doctoral universities with high research activity) [8]. Nearly all R1 universities, and approximately 60% of R2 universities, have research computing resources [32]. There are a few estimates of the number of RCD professionals in specific roles at these institutions. Neeman et al. [2020] estimate 7 RCD professionals in facilitation roles alone at each R1 institution, and 3 RCD facilitators at each of the R2 institutions that offers RCD services [32]. In addition, approximately

75% of R1 institutions have librarians with research data roles, with an average of 3 such positions at institutions with these roles present [43]. Yet facilitators and data librarians are only two of many RCD roles. Conservatively estimating that there are also at least as many people in other types of RCD roles (e.g. data scientists, research software engineers, system administrators, networking specialists, managers and directors) at all R1 universities (20 people per university total) and the 60% of R2 universities that publicize RCD resources (6 people per university total), that is approximately 3,100 RCD professionals at R1 and R2 universities alone.

In addition to universities, many RCD professionals also work at large academic computing centers that support regional or national research programs. As of January 2022, San Diego Supercomputer Center (SDSC) [14], Texas Advanced Computing Center (TACC) [15], Ohio Supercomputer Center [11], Renaissance Computing Institute (RENCI) [13], National Center for Supercomputing Applications (NCSA) [9], and Pittsburgh Supercomputing Center (PSC) [12] combined list approximately 1,000 staff on their websites. This is only a sample of the full set of such centers.

These sources together suggest there are at least 5,000 RCD professionals at academic institutions.²

3 SURVEYING THE WORKFORCE

A working group of the Campus Research Computing Consortium (CaRCC) developed and conducted a survey of the demographic characteristics, backgrounds, experience, positions, and views of the RCD workforce in 2021 [31].³ The nine members of the working group drafted the survey and then circulated it to approximately a dozen additional individuals in the RCD field at different academic institutions with varying backgrounds and positions to help ensure that the questions and answer choices reflected the diversity of the field.

The survey was open from June 2, 2021 through September 30, 2021. Invitations to participate in the study were sent to the email lists of RCD professionals noted above, as well as the US-RSE and National Center for Women & Information Technology; those receiving the survey announcement were encouraged to forward it to their professional networks and additional email lists. Emails were also sent to individuals at various institutions mid-way through the survey period to encourage broader participation. The survey was conducted with Qualtrics. The survey results were analyzed with R, RStudio, and the tidyverse and janitor R packages [24, 38, 40, 45].

The survey received 595 responses, 32 of which were found to be duplicates based on a comparison of job title, institution, and demographics.⁴ There are 563 valid, unique responses for analysis; 539 respondents are currently employed in a position with RCD responsibilities.

²This assumes that there are additional RCD professionals not accounted for here and/or limited overlap between the estimated 3,100 individuals at R1 and R2 universities, the 1,000 or more individuals at academic computing centers, and the 2,500 individuals on RCD email lists.

³This research has been reviewed and approved by an Institutional Review Board (IRB) at Northwestern University, IRB STU00215053.

⁴Most duplicate responses occurred months apart, likely due to the survey being open for four months. When duplicate responses were identified, the first response was kept, and any subsequent responses were excluded from analysis. This choice was made so that answers would not be influenced by having seen the questions previously.

¹This data was acquired through direct correspondence with the list managers.

Respondents were affiliated with 175 different institutions, including 8 companies, 5 government organizations, 10 medical centers or schools, and 144 academic institutions. The median number of respondents per institution was two; the maximum was 26. Few RCD professionals outside of academia responded to the survey: 86% of respondents with a current RCD position and known institution are employed within an academic environment – a college, university, or academic computing center. Over half of the reported academic institutions (63%), accounting for 82% of academic respondents, were R1 universities. 69% of R1 universities and 22% of R2 universities had at least one respondent to the survey.

Given the above estimate of at least 5,000 people in the academic RCD workforce, the 439 respondents with a current RCD position at an academic institution reflects at most 9% of the academic RCD workforce. While we believe this to be the largest survey of the RCD field to date, there are still significant portions of the RCD workforce, and even just the academic RCD workforce, not represented among the survey respondents. 31% of R1 institutions and the majority of R2 institutions with RCD services are not represented among survey respondents. Relatively few respondents from academic computing centers participated, given that over 1000 RCD professionals work at such centers. In terms of roles, 30% of survey respondents in a staff RCD role are managers, directors, or other senior leaders; these roles are likely overrepresented.

The survey also captured information about people's job responsibilities and their compensation not reported here. These data will be further explored and shared in future papers. De-identified data from the survey is available [30].

4 SURVEY RESULTS

4.1 Demographics

Age: The majority (60%) of survey respondents are 35-54 years old, which is higher than in the US labor force as a whole (41%) [2]. There were fewer younger (<35 years old) respondents than in the US labor force: 19% vs. 35%; the number of respondents older than 54 was comparable to the broader labor force: 21% vs. 24%.

Race/ethnicity: 84% of respondents who provided a racial identification identified as white alone; 11% identified as Asian/Asian-American alone; 2% identified as Black/African-American alone; 3% identified with more than one listed choice, a non-listed choice, or Native American/Hawaiian/Alaskan/Pacific Islander; 2% of respondents identified as Hispanic (across all racial groups). Comparing these numbers to the US 2020 Census: 76% white alone; 6% Asian/Asian-American alone; 13% Black/African-American alone; 5% multiple/another answer; 19% Hispanic overall (across all racial groups) [6]. Statistics from the US Bureau of Labor on those in computer and mathematical occupations: 65% white; 23% Asian/Asian-American; 9% Black; 8% Hispanic (across all racial groups) [7]. The National Center for Education Statistics reports the following numbers for US college and university faculty: 75% white alone; 12% Asian/Asian-American alone; 6% Black/African-American alone; 1% multiple/another answer; 6% Hispanic (exclusive from other choices) [10].

Gender: 73% of respondents who provided a gender identity identified as male, 25% as female, and 2% as non-binary, nonconforming, genderqueer, or another gender. US workers in computer

and mathematical occupations are 75% male, 25% female [7]. US college and university faculty are 50% male and 50% female [10]. The Practice and Experience in Research Computing 2021 (PEARC21) post-conference survey report indicated 67% of attendees were male, 29% female, and 4% non-binary or another gender. An EDUCAUSE survey of the higher education IT workforce found 62% of respondents were male and 38% female [26].

US Citizenship/Residency: 92% of respondents are US citizens or permanent residents; 5% are not; 3% no answer. Nationally, approximately 1% of the US labor force are temporary foreign workers with nonimmigrant visas [21].

Sexual Orientation: 87% of respondents who provided a sexual orientation identify as heterosexual; 4% identified as gay or lesbian. 3% or less of respondents identified with each of the other sexual orientation choices provided (asexual, bisexual, pansexual, queer) or a write-in response. Nationally, 6% of US adults identify as lesbian, gay, bisexual, or transgender (LGBT) [27].

Disability Status: 7% of respondents report having a disability. Approximately 4% of the US civilian workforce has a disability [5].

4.2 Education and Experience

Educational Attainment and Field: 37% of respondents have a PhD or Doctorate; 33% have a Master's degree; 25% have a Bachelor's degree. 94% of respondents reported one (67%) or two (26%) fields for their educational experience. For those without a doctorate, the proportion of respondents reporting education in each field is: computer and information sciences (57%); engineering (19%); arts and humanities (13%); math and statistics (13%); physical sciences (12%); social sciences (9%); life sciences (9%); other fields (8%). For respondents with a doctorate, the proportion reporting education in each field is: physical sciences (39%); computer and information sciences (28%); life sciences (24%); engineering (18%); math and statistics (14%); social sciences (6%); arts and humanities (4%); other fields (2%).

48% of respondents overall say their educational field is "very related" or "extremely related" to their current work, but this varies by educational field. For example, 70% of respondents who reported their education as in computer and information sciences, but only 37% of those whose education was in physical sciences, say their educational field is "very related" or "extremely related" to their current work.

Years and Type of Work Experience: The median number of years of total work experience is 20, and the median years of RCD work experience is 9. For 23% of respondents, all of their work experience is in the RCD field; the median is 67% of work experience being in the RCD field. Only 12% of RCD professionals currently working for an academic institution have RCD work experience at a non-academic institution. 49% of respondents have some corporate work experience, but most of this experience is outside of the RCD field.

4.3 Current Positions

The results in this section and those that follow are for the 96% of respondents who were employed in an RCD position at the time they responded to the survey (N=539).

Table 1: Respondents with current RCD positions by employer type

Employer Type	% Respondents
R1 University, Public	47%
R1 University, Private	20%
R2 University	9%
Other University or College	4%
Medical School or Center	5%
Academic Computing Center	2%
Government	1%
Company	1%
Other or Unknown	10%

Employer Types and Units: Table 1 shows the percentage of respondents employed by each type of institution. 26% of respondents work for a group with "research computing" in the name. "Research" and "computing" are the most common words in group names.

Tenure and Promotion: Median tenure working for the same group or unit is five years, but 27% of respondents have worked for the same group or unit for ten or more years. Median tenure in the same position is 3 years. These numbers correspond to statistics for the higher education IT workforce, where median tenure in a position is also 3 years overall, about a quarter of staff have been in their position for 10 or more years, and median tenure at the same institution is 10 years [36, 37]. Nationally, median job tenure is 4 years; it is 5 years for those in educational, training, and library occupations [3].

47% of respondents have not had an RCD role with a group or unit other than their current one. 45% of respondents have been promoted within their unit or group; respondents have been promoted at similar rates across genders. 41% of those in their position for 5-9 years have not been promoted; 27% of those in their position for ten or more years have not been promoted.

Position Types and Levels: Table 2 shows the percentage of respondents in each type of position. Managers, directors, and other senior leaders are likely overrepresented among survey respondents. The work experience column provides the median years of RCD work experience. Male respondents are more likely than female respondents to be in positions supervising staff (39% of male respondents vs. 30% of female respondents) and in the job categories of manager, director, or other senior leadership positions (30% of male respondents vs. 24% of female respondents). White respondents and respondents of other racial and ethnic identities are in supervisory and management roles at similar rates.

Job titles: The most common word (excluding those indicating the job's level) in job titles is *research*, appearing in 43% of respondents' official or alternate job titles. This is followed by variations on *compute* (26%), *system* (20%), and *engineer* (17%). 55% of respondents reported an alternate job title that they use in addition to or in place of their official job title.

Funding: For respondents in staff roles, 98% are in full time positions, and 98% are in salaried positions. 63% are in positions funded by "hard" money or central institutional funds; 10% are

Table 2: Position types and experience for respondents currently employed in RCD positions. Experience is median years of RCD experience for the position type.

Primary Position Type	Respondents	Experience	Supervise Staff
AVP, VP, CIO, or other senior leader	2%	20	89%
Director	13%	15	93%
Manager	12%	14	90%
Lead/principal individual contributor	13%	10	34%
Senior individual contributor	21%	10	8%
Individual contributor	24%	4	2%
Faculty	10%	10	41%
Student position/internship	2%	1	-
Other/Unknown	4%	-	-

funded by "soft" money or grant funds; 23% are in positions funded from a mixture of sources; 4% do not know the source of their position funding or did not answer.

Worksite: 56% of respondents report normally working on-site outside of pandemic restrictions; 34% work a combination of on-site and remotely; 10% work remotely.

Responsibilities: 53% of respondents provide RCD services for their entire institution; 22% for multiple institutions; 12% nationally; and 13% for a college, school, center, institute, or research group. 55% of respondents have a position fully focused on RCD work, while 21% spend 76-99% of their time on RCD work. Overall, 89% of respondents spend at least half of their work time on RCD responsibilities. 37% of respondents supervise staff; 36% supervise student workers.

4.4 Position Domains

CaRCC has developed a model, referred to as Facings, for characterizing the primary types of work that RCD professionals engage in [1]. These facings are Researcher, Systems, Data, Software, and Strategy & Policy. The survey asked respondents if they were familiar with the Facings model and how they would characterize their position. 39% of respondents were familiar with the model, and an additional 19% were "sort of" familiar with the model; 41% were not familiar with the model.

Only 3% of respondents allocated their time to a single facing (researcher, systems, data, software, or strategy & policy); 66% of respondents allocated at least some proportion of their time to four or more facings. 56% of respondents allocated at least half of their time to one of the five named facings: 20% of respondents spend at least half of their time in a researcher facing role; 13% systems facing; 10% software facing; 9% strategy & policy facing; 5% data facing. The median proportion of time allocated to each facing is: 25% for researcher and 10% for all other areas.

4.5 Position and RCD Field Satisfaction

We asked respondents whether their RCD work was "acknowledged and valued by" different groups, on a five-point scale from "Always" to "Never." The proportion of respondents who reported that their work was valued "Most of the time" or "Always" by their RCD peers and colleagues was 80%; by their supervisor or manager, 85%; by the researchers they work with, 77%; by the organization they work for, 57%. This recognition matters for respondents' view of the RCD field more broadly. Those whose work is valued by their

peers are more likely to be "somewhat" or "extremely satisfied" with working in the RCD field (92%) than those who do not agree that their work is valued by their peers (77%). This pattern holds for acknowledgement by the other groups as well: by managers (92% vs. 71%), researchers (90% vs. 82%), and their institution (96% vs. 80%).

While overall 88% of survey takers are satisfied with working in the RCD field, only 67% responded that they "somewhat" (39%) or "strongly agree" (27%) that there are many opportunities in their chosen career path; the proportion is similar across those who have been promoted in their group or unit and those who have not (69% vs. 64%). 77% of all respondents answered that they "somewhat" (37%) or "strongly agree" (40%) that they have a future in the RCD field. When asked how likely they were to recommend a position like their current one to others, respondents on average answered a 7.7 on a scale from 0-10 with 63% of respondents reporting an 8, 9, or 10.

Finally, 80% of respondents indicated that they "somewhat" (34%) or "strongly agree" (46%) that they feel included and welcome in the field. Male (83%), white (83%), and heterosexual respondents (83%) are more likely to agree that they feel included and welcomed than those of other genders (73%), racial or ethnic identities (71%), or sexual orientations (77%). Only 61% of Hispanic respondents agree that they feel included. Respondents without a disability (80%) are more likely to agree they feel included than those with a disability (71%). Feelings of inclusion are fairly consistent across position levels (range 77-83% for categories with more than 10 respondents), citizenship/permanent resident status, and the degree to which respondents' educational backgrounds are relevant to their current work.

5 DISCUSSION

As Arafune et al. [2020] note, there is an acute need to improve recruitment and retention efforts with the RCD field [16]. Identifying opportunities to further expand participation in the RCD profession, both in terms of the number of people and the diversity of their backgrounds and skills, is necessary to meet the continued demands for a workforce with the requisite skills and knowledge to advance science and engineering initiatives.

5.1 Demographics

The RCD workforce has fewer individuals under 35 than the US labor force as a whole. This is consistent with an aging overall higher education IT workforce [26]. The high educational attainment levels of those within the RCD profession contributes to a relatively low proportion of younger workers. Programs to make students, especially undergraduates, aware of opportunities in the RCD field and changes in hiring practices to incorporate workers without graduate degrees in the field can help attract younger workers and individuals from a wider variety of demographic backgrounds. The imbalance in the age distribution also limits mentorship opportunities for more senior staff and reduces opportunities for them to develop the leadership and management skills needed to progress in their careers.

A higher proportion of RCD professionals identify as white (84%) than in the broader US computational and mathematical occupations workforce (65%). The racial and ethnic distribution of RCD

professionals is more similar to that of college and university faculty, although the proportion of faculty who identify as white (75%) is smaller than that for RCD professionals. In terms of gender, the distribution of the RCD workforce matches that of the US computational and mathematical occupations workforce. A smaller proportion of female RCD professionals are in management and supervisory positions than are their male counterparts.

Increasing the diversity of the RCD workforce will be challenging and requires intentional action and engagement from institutional leadership. Even top tech companies, with higher average compensation rates than the RCD field, struggle with cultivating a diverse workforce [22]. Yet work to increase diversity and inclusion is imperative even beyond considerations of representation and the need to recruit new people. Research has demonstrated that inclusion and diversity in the workplace yield an increase in productivity and profitability, especially when racial diversity exists in upper management and companies go beyond representational diversity to ensure inclusive practices [17, 23, 39].

This suggests that an important step in increasing diversity in the RCD profession is to ensure that all members of the workforce are respected and feel welcome. While the majority of respondents in all demographic groups agree that they feel included and welcomed in the RCD field, respondents with historically dominant identities (white, male, heterosexual, not disabled) agreed at higher rates than respondents with identities that have been historically excluded from computational and technical fields. Efforts to make the field more inclusive for such individuals must be part of the effort to broaden participation and recruit new and diverse talent to the profession.

5.2 Education and Experience

Historically, academic institutions have placed a high value on academic credentials for staff roles that enable research, therefore it is not surprising that a large percent (70%) of respondents have a graduate degree. However, the RCD profession, like the technology industry, is largely skills based, and traditional university degrees may have limited applicability for many RCD roles. Degree requirements that do not align with positions' responsibilities hinder efforts to recruit a diverse workforce. Bachelor's degree requirements screen out approximately 70% of African-Americans, 80% of Latinx workers, and 80% of all workers living in rural areas [4]; graduate degree requirements further limit the pool of eligible workers.

Further work is needed to understand how current RCD professionals gained their skills and experience; fortunately, the collection of case studies about RCD career paths is currently underway [20]. A more detailed understanding of the types of work and educational experience that support successful RCD careers will be a first step in identifying alternative pathways for people to gain that experience.

The need to diversify the educational backgrounds of those in the RCD workforce extends to educational domains as well. RCD professionals are distinguished from IT professionals by their knowledge of both research processes and specialized technical skills. The RCD workforce as a whole needs people with experience in all research domains to support researchers from the growing set of

fields conducting computational and data-intensive research. Those with backgrounds in the social sciences are particularly underrepresented among RCD professionals at all education levels compared to the proportion of researchers in the social sciences [34].

5.3 Current Positions

Many institutions employ a small number of RCD professionals, meaning that those in the RCD workforce routinely perform a wide range of tasks. This is evident in the data, where nearly all respondents report that they spend significant time on work that falls under multiple CaRCC Facings (researcher, software, data, systems, strategy & policy). The complexity of the work of RCD professionals does not always match well to the standardized job descriptions and categories developed by human resources departments. This potential mismatch is seen in over half of respondents using a job title other than their official one.

Additionally, institutions should consider whether traditional employment arrangements are required for RCD roles. Prior to the pandemic, less than half of RCD professionals were working remotely even part of the time, and only 10% exclusively. Except for the management of physical hardware, most RCD work can be performed successfully remotely, and virtual consultations can reduce barriers to direct one-on-one support, especially where RCD professionals are supporting geographically dispersed researcher populations. To meet the growing need for RCD talent and compete with other employers, institutions will need to consider flexible work arrangements, especially when they are not located in large labor markets with sufficient technical talent [19].

5.4 Perspectives on the RCD Field

The survey results show that a large proportion of those working within the RCD field feel their work is valued by those they work with directly and would encourage others to work in the profession. However, the relatively low percentage of respondents who said that their work was valued "most of the time" or "always" by the institution they work for indicates that there is still progress to be made to help research institutions understand the value that RCD professionals contribute to their institutions. These data are consistent with the existence at most academic institutions of a two-tiered culture between faculty and staff, which can lead to imbalance in status and recognition of RCD professionals as part of the research mission [35].

Institutions also need to do a better job defining career paths and progression for RCD professionals. While 88% of RCD professionals indicate general satisfaction about working within the RCD field, only 67% agree that there are many opportunities within their chosen career path. Those who don't agree there are many opportunities or don't agree they have a future in the RCD field are twice as likely to have applied for a new job outside of the RCD field within the last six months (15% vs. 7%).

Nearly half of respondents have only ever been in an RCD role with their current group or unit. While they may have changed jobs or position grade levels, they have only experienced the RCD profession through a single institution. This has significant implications for the RCD field. For RCD work to develop as a profession with common roles, practices, expectations, norms, and

skills, information and experience must flow between institutions. Professional networks, conferences, and online collaborations are key for providing a common, shared experience and supporting professionalization.

6 CONCLUSIONS

This is the first survey to quantitatively measure the broad RCD workforce and the unique characteristics, skills, and backgrounds of its members. The demographic diversity of the RCD workforce falls short of that of university faculty and of the broader US mathematical and computational workforce; there are, however, many pathways by which the diversity of the profession could be improved. The RCD workforce is highly educated but would benefit from recruiting individuals from a wider range of educational domains, especially social science. A large majority of respondents feel included and welcome in the field and feel their work is valued by those they work with, but there is work to be done conveying the value of RCD work to institutional leaders and ensuring that those from all demographic backgrounds feel equally welcome in the field. Overall, the survey results support prior qualitative assessments of the challenges the RCD field and those in it face to recruit and retain talent, develop career pathways, and grow the profession.

Additional papers on parts of the survey not covered here are underway, including analysis of job responsibilities and compensation (which were not covered here) and further breakdowns of survey responses for different groups of RCD professionals. The intention is to repeat the survey every few years to track the progression of the workforce and profession as they evolve. While this survey captured information about a significant proportion of the academic RCD workforce, future surveys should specifically target R2 universities and academic computing centers and institutions that are not currently reached by community email lists to broaden participation and build an increasingly representative sample of the academic RCD workforce. Beyond academia, few RCD professionals from government and corporate institutions participated. Identifying such individuals and connecting academic and non-academic RCD communities is likely to be challenging as the majority of those working in academic RCD positions do not have RCD experiences outside of academic institutions. Developing such connections will be part of the evolution of the RCD profession.

This data provides a foundation against which to measure the progress of community efforts in areas such as recruiting professionals who reflect the diversity of the US research community and population, developing common job descriptions and titles, supporting professional development, and improving institutional recognition for the value of RCD work. The RCD community has many highly engaged members, and we have the opportunity to shape the development of this new profession to support those in it and welcome new people to it. Data on the workforce will allow us to better assess this work.

ACKNOWLEDGMENTS

In addition to four of the authors, Jay Alameda, Galen Collier, Kimberly Grash, and Amy Neeser were members of the CaRCC professionalization working group that developed the survey reported in this paper. Douglas Dodson, Kimberly Grash, Susan Ivey, Jim Leous,

Jackie Milhans, Claire Mizumoto, Amy Neeser, Patrick Schmitz, and Colby Witherup-Wood provided helpful feedback on a draft of this paper. Thanks to all of them for their contributions and support. This work is part of the RCD Nexus, which is supported as an NSF-funded Cyberinfrastructure Center of Excellence (CI CoE) demonstration pilot (award NSF-2100003).

REFERENCES

- [1] 2018. Research Computing and Data Professionals Job Elements and Career Guide. <https://carcc.org/wp-content/uploads/2019/01/CI-Professionalization-Job-Families-and-Career-Guide.pdf>
- [2] 2019. Labor Force Share, by Age Group, 1998, 2008, 2018, and Projected 2028 : U.S. Bureau of Labor Statistics. <https://www.bls.gov/emp/graphics/2019/labor-force-share-by-age-group.htm>
- [3] 2020. *Employee Tenure in 2020*. Technical Report USDL-20-1791. Bureau of Labor Statistics, U.S. Department of Labor. <https://www.bls.gov/news.release/pdf/tenure.pdf>
- [4] 2021. No College, No Problem. Some Employers Drop Degree Requirements to Diversify Staffs. <https://www.npr.org/2021/04/29/990274681/no-college-no-problem-some-employers-drop-degree-requirements-to-diversify-staff>
- [5] 2021. *Persons with a Disability: Labor Force Characteristics - 2020*. News Release USDL-21-0316. US Department of Labor: Bureau of Labor Statistics. <https://www.bls.gov/news.release/pdf/disabl.pdf>
- [6] 2021. U.S. Census Bureau QuickFacts: United States. <https://www.census.gov/quickfacts/fact/table/US/PST045219>
- [7] 2022. 2020 Annual Averages - Employed persons by detailed occupation, sex, race, and Hispanic or Latino ethnicity. <https://www.bls.gov/cps/cpsaat11.htm>
- [8] 2022. The Carnegie Classification of Institutions of Higher Education. <https://carnegieclassifications.iu.edu/>
- [9] 2022. National Center for Supercomputing Applications (NCSA). <http://www.ncsa.illinois.edu/people/staff-directory/>
- [10] 2022. The NCES Fast Facts Tool provides quick answers to many education questions (National Center for Education Statistics). <https://nces.ed.gov/fastfacts/display.asp?id=61> Publisher: National Center for Education Statistics.
- [11] 2022. Ohio Supercomputer Center. <https://www.osc.edu/directory>
- [12] 2022. Pittsburgh Supercomputing Center (PSC). <https://www.psc.edu/staff-directory/>
- [13] 2022. Renaissance Computing Institute (RENCI). <https://renci.org/mission-and-vision/staff-directory/>
- [14] 2022. San Diego Supercomputer Center (SDSC). https://www.sdsc.edu/cgi-bin/staff_dir.cgi
- [15] 2022. Texas Advanced Computing Center (TACC). <https://www.tacc.utexas.edu/about/directory>
- [16] Lisa Arafuno, Dana Brunson, Thomas Hacker, and Preston Smith. 2020. *Building the Research Innovation Workforce: A workshop to identify new insights and directions to advance the research computing community*. Workshop Report. <https://www.rcac.purdue.edu/ciworkforce2020/report>
- [17] Ruth Sessler Bernstein, Bulger Morgan, Paul Salipante, and Judith Y. Weisinger. 2020. From Diversity to Inclusion to Equity: A Theory of Generative Interactions. *Journal of Business Ethics: JBE* 167, 3 (Dec. 2020), 395–410. <https://doi.org/10.1007/s10551-019-04180-1> Num Pages: 395-410 Place: Dordrecht, Netherlands Publisher: Springer Nature B.V.
- [18] Sharon Broude Geva, Dana Brunson, Thomas Cheatham III, James Deaton, James Griffioen, Curtis W. Hillegas, Douglas M. Jennewein, Gail Krovitz, Tobin Magle, Patrick Schmitz, Karen Tomko, and James C. Wilgenbusch. 2020. Fostering Collaboration Among Organizations in the Research Computing and Data Ecosystem. In *Practice and Experience in Advanced Research Computing*. Association for Computing Machinery, New York, NY, USA, 393–401. <https://doi.org/10.1145/3311790.3396645>
- [19] Bhaskar Chakravorti. 2020. To Increase Diversity, U.S. Tech Companies Need to Follow the Talent. <https://hbr.org/2020/12/to-increase-diversity-u-s-tech-companies-need-to-follow-the-talent>
- [20] Shafaq Chaudhry, Arman Pazouki, Patrick Schmitz, Betsy Hillery, and Kerk Key. 2022. Understanding Factors that Influence Research Computing and Data Careers. In *Practice and Experience in Advanced Research Computing (PEARC '22)*. ACM. <https://doi.org/10.1145/3491418.3530292>
- [21] Daniel Costa and Jennifer Rosenbaum. 2017. *Temporary foreign workers by the numbers: New estimates by visa classification*. Technical Report. <http://epi.org/120773>
- [22] Megan Dickey. 2019. The future of diversity and inclusion in tech. <https://techcrunch.com/2019/06/17/the-future-of-diversity-and-inclusion-in-tech/>
- [23] Sundiati Dixon-Fyle, Kevin Dolan, Vivian Hunt, and Sara Prince. 2020. *Diversity wins: How inclusion matters*. Technical Report. McKinsey & Company. 56 pages. <https://www.mckinsey.com/featured-insights/diversity-and-inclusion/diversity-wins-how-inclusion-matters>
- [24] Sam Firke. 2021. *janitor: Simple Tools for Examining and Cleaning Dirty Data*. <https://CRAN.R-project.org/package=janitor>
- [25] Athina Frantzana. 2019. *Women's Representation and Experiences in the High Performance Computing Community*. Ph.D. Dissertation. The University of Edinburgh.
- [26] Joseph Galanek, Dana C. Gierdowski, and D. Christopher Brooks. 2019. *The Makeup of Today's IT Workforce*. Technical Report. EDUCAUSE. <https://www.educause.edu/ecar/research-publications/the-higher-education-it-workforce-landscape-2019/foreword-introduction-and-key-findings>
- [27] Jeffrey Jones. 2021. LGBT Identification Rises to 5.6% in Latest U.S. Estimate. <https://news.gallup.com/poll/329708/lgbt-identification-rises-latest-estimate.aspx> Section: Politics.
- [28] Luba Katz. 2019. *Careers of Data Scientists: Report from 13 Academic Institutions*. Technical Report. Zenodo. <https://doi.org/10.5281/zenodo.4404824>
- [29] Angellina Lauer. 2020. *Cyberinfrastructure Professionals' Withdrawal Cognitions: An Exploration of Race, Gender, and Occupational Climate in High Performance Computing*. Ph.D. Dissertation. <https://dspace2.creighton.edu/xmlui/handle/10504/129478> Accepted: 2021-03-12T15:13:20Z Publisher: Creighton University.
- [30] Christina Maimone, Scott Yockel, Jay Alameda, Timothy Middelkoop, Amy Neeser, and Ashley Stauffer. 2022. CaRCC Research Computing and Data (RCD) Workforce Survey Data 2021 - Part 1. <https://doi.org/10.5281/zenodo.6574465> Type: dataset.
- [31] Christina Maimone, Scott Yockel, Timothy Middelkoop, Jay Alameda, Ashley Stauffer, and Amy Neeser. 2021. CaRCC Research Computing and Data (RCD) Workforce 2021 Census Questions. <https://doi.org/10.5281/zenodo.5914430>
- [32] Henry Neeman, David Akin, Hussein Al-Azzawi, Kevin L. Brandt, Jamene Brooks Kieffer, Dana Brunson, Dirk Colbry, Sandra Gesing, Anna Klimaszewski-Patterson, Claire Mizumoto, Joy A. Pine-Thomas, Anita Z. Schwartz, Horst Severini, Mohammed Tanash, and Daniel Voss. 2020. Cyberinfrastructure Facilitation Skills Training via the Virtual Residency Program. In *Practice and Experience in Advanced Research Computing*. ACM, Portland OR USA, 421–428. <https://doi.org/10.1145/3311790.3396629>
- [33] OECD. 2020. *Building digital workforce capacity and skills for data-intensive science*. OECD Science, Technology and Industry Policy Papers 90. OECD, Paris. <https://doi.org/10.1787/e08aa3bb-en>
- [34] Abigail Okrent and Amy Burke. 2021. *The STEM Labor Force of Today: Scientists, Engineers, and Skilled Technical Workers*. Technical Report. National Science Foundation. <https://ncses.nsf.gov/pubs/nsb20212>
- [35] David Payne. 2021. Calls for culture change as "them versus us" mindset drives rift between academic and non-academic staff. <https://www.natureindex.com/news-blog/calls-culture-change-them-versus-us-drives-rift-between-academics-administrators-research-science>
- [36] Jeffrey Pomerantz and Christopher D. Brooks. 2016. *The Higher Education IT Workforce Landscape, 2016*. Technical Report. EDUCAUSE Center for Analysis and Research, Louisville, CO. <https://library.educause.edu/-/media/files/library/2016/4/ers1603.pdf>
- [37] Adam Pritchard, Keith McIntosh, and Jasper McChesney. 2019. Aging, Retirement, and the Pipeline. *EDUCAUSE Review* Fall 2019 (Oct. 2019). <https://er.educause.edu/articles/2019/10/aging-retirement-and-the-pipeline>
- [38] R Core Team. 2021. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>
- [39] Orlando C. Richard, María del Carmen Triana, and Mingxiang Li. 2021. The Effects of Racial Diversity Congruence between Upper Management and Lower Management on Firm Productivity. *Academy of Management Journal* 64, 5 (Oct. 2021), 1355–1382. <https://doi.org/10.5465/amj.2019.0468>
- [40] RStudio Team. 2021. *RStudio: Integrated Development Environment for R*. RStudio, PBC, Boston, MA. <http://www.rstudio.com/>
- [41] Patrick Schmitz, Scott Yockel, Claire Mizumoto, Thomas Cheatham, and Dana Brunson. 2021. Advancing the Workforce That Supports Computationally and Data Intensive Research. *Computing in Science Engineering* 23, 5 (Sept. 2021), 19–27. <https://doi.org/10.1109/MCSE.2021.3098421> Conference Name: Computing in Science Engineering.
- [42] Dan Sholler. 2019. "Invisible Work" as a Lens for Understanding Humanware's Role in Research Cloud Computing: Evidence from an interview-based study. In *Proceedings of the Humans in the Loop: Enabling and Facilitating Research on Cloud Computing (HARC '19)*. Association for Computing Machinery, New York, NY, USA, 1–5. <https://doi.org/10.1145/3355738.3355754>
- [43] Rebecca Springer. 2019. Counting Data Librarians. <https://sr.ithaka.org/blog/counting-data-librarians/>
- [44] Josh Trapani. 2021. Academic Research and Development. <https://ncses.nsf.gov/pubs/nsb20213>
- [45] Hadley Wickham, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, Alex Hayes, Lionel Henry, Jim Hester, Max Kuhn, Thomas Lin Pedersen, Evan Miller, Stephan Milton Bache, Kirill Müller, Jeroen Ooms, David Robinson, Dana Paige Seidel, Vitalie Spinu, Kohske Takahashi, Davis Vaughan, Claus Wilke, Kara Woo, and Hiroaki Yutani. 2019. Welcome to the tidyverse. *Journal of Open Source Software* 4, 43 (2019), 1686. <https://doi.org/10.21105/joss.01686>