Understanding Factors that Influence Research Computing and Data Careers

Shafaq Chaudhry
University of Central Florida
New York, USA
Shafaq.Chaudhry@ucf.edu

Arman Pazouki
Northwestern University
Evanston, USA
pazouki@northwestern.edu

Patrick Schmitz
Semper Cogito
Berkeley, USA
patrick@sempercogito.com

Elizabett Hillery
Purdue University
New York, USA
eahillery@purdue.edu

Kerk Kee
Texas Tech University
New York, USA
Kerk.Kee@ttu.edu

ABSTRACT

Research Computing and Data (RCD) professionals play a crucial role in supporting and advancing research that involve data and/or computing, however, there is a critical shortage of RCD workforce, and organizations face challenges in recruiting and retaining RCD professional staff. It is not obvious to people outside of RCD how their skills and experience map to the RCD profession, and staff currently in RCD roles lack resources to create a professional development plan. To address these gaps, the CaRCC RCD Career Arcs working group has embarked upon an effort to gain a deeper understanding of the paths that RCD professionals follow across their careers. An important step in that effort is a recent survey the working group conducted of RCD professionals on key factors that influence decisions in the course of their careers. This survey gathered responses from over 200 respondents at institutions across the United States. This paper presents our initial findings and analyses of the data gathered. We describe how various genders, career stages, and types of RCD roles impact the ranking of these factors, and note that while there are differences across these groups, respondents were broadly consistent in their assessment of the importance of these factors. In some cases, the responses clearly distinguish RCD professionals from the broader workforce, and even other Information Technology professionals.

CCS CONCEPTS

• Social and professional topics → Computing occupations.

KEYWORDS

Research Computing and Data, RCD, Cyberinfrastructure, CI, Professionalization, Career Arcs

1 INTRODUCTION

A 2018 survey of nearly 12,000 scholars from various scientific disciplines across more than 60 countries indicated that 67% of scientific production results in new data or code (34% data only, 9% code only, and 24% both) [7]. Research Computing and Data[1] (RCD) Professionals play a crucial role in supporting and advancing research that involves data and/or computing, however, there is a critical shortage of RCD workforce, and organizations face challenges in recruiting and retaining RCD Professional staff [3]. Data indicates that institutions have been more successful in providing computational resources than providing necessary catalysts for effective use of those resources [5]. A Department of Energy Technical Meeting reported that workforce development is a major concern in High Performance Computing (HPC) and a priority for supporting National Strategic Computing Initiative (NSCI) objectives [2]. Similarly, workforce management is one of the key competencies identified in the NSF Research Infrastructure Guide [20] for the success of any scientific facility that runs and manages research cyberinfrastructure (CI). This includes programs to recruit and retain staff through continued professional development. In a recent Coalition for Academic Scientific Computation (CASC) survey, issues related to RCD workforce including staff skill sets, diversity, retention, recruiting, and workforce development were identified as one of the top four priorities facing the RCD profession[2].

Distinct RCD roles are relatively new and it is not obvious to people outside of RCD how their skills and experience map to the RCD profession [25]. This is a serious barrier to recruitment and to building a pipeline into these roles. In addition, current RCD staff lack resources to help them grow and create a professional

---

[1]"Research Computing and Data" (abbreviated as RCD) includes technology, services, and people supporting the needs of researchers and research, and is intended as a broad, inclusive term covering computing, data, networking, and software. The National Science Foundation uses the term "cyberinfrastructure," and others use "Research IT."
[2]CASC (https://casc.org/) Fall 2021 Membership Meeting, member discussion, October 19-21 2021.

development plan. This undermines retention of these skilled professionals, which is an issue that is indicated in several studies: e.g., LinkedIn reports that employees consider having learning and development opportunities as the top indicator of an exceptional workplace, which helps in retention as well as attracting and recruiting [17]; an EDUCAUSE report indicated that professional development contributes to job satisfaction [10]; a 2016 Gallup report revealed that 59% of millennials say opportunities to learn and grow are extremely important to them when applying for a job [1].

Studies pertaining to the RCD workforce have similarly indicated the significance of workforce development. An NSF-funded workshop in 2018 identified professional development as the top priority [9], and a more recent NSF-sponsored workshop in 2021 identified the need for recognition and a clear definition of the different roles in the workforce as well as the need for viable, reasonably funded career paths with location stability, and the availability of training, advancement, and upskilling [3]. Neeman et al. reported on the gap in national formal education curriculum to cultivate the needed workforce [21]. Goodhue et al. identified the lack of access to cyberinfrastructure workforce at small and mid-size institutions and the importance of a trained workforce that can bridge those gaps [13].

All of this points to the importance of researching the challenges and opportunities of the RCD workforce and has been the subject of several Campus Research Computing Consortium (CaRCC)[3] working groups. The RCD Professionalization working group within CaRCC recently conducted a survey to understand the demographic characteristics of the RCD workforce, and in 2021, the CaRCC RCD Career Arcs working group embarked upon an effort to gain a deeper understanding of the paths that RCD professionals follow across their careers. These efforts will enable recruiters to market these positions better, improve targeting of recruitment activities and show the various possibilities for the potential future workforce. Studying these patterns will also help existing RCD workforce gain a better understanding of their future paths in the field and help them identify areas of growth, professional development, and opportunities, thereby improving workforce retention.

In the first phase of the Career Arcs work, we organized a birds-of-a-feather session at the Supercomputing (SC) 2021 conference where we invited a panel of experts from various RCD roles to share their journeys. This was followed by facilitated conversations with the session participants around issues pertaining to the RCD career arcs. Common themes revolved around the importance of good mentors and networking, the importance (and challenges) of taking risks, and the need for skills in self-promotion and storytelling as one ascends the ladder. Collaboration and teamwork, communication and technical skills, and willingness to listen and learn were identified as skills that contribute to getting promoted. Feedback from this session provided valuable input to our future effort of one-on-one interviews with RCD Professionals.

In the next phase, we designed a survey for RCD professionals to gather data on the key factors that influence decisions in the course of their careers. This survey was opened from November 16, 2021 through February 1, 2022 and gathered responses from more than 200 respondents at institutions across the United States (US).

---

[3]https://carcc.org/

This paper presents our initial findings and analyses of the data gathered. The rest of the paper is organized as follows: Section 2 presents related work; Section 3 describes the research and analysis methodology; results and analysis are presented in Section 4; and the last section presents conclusions and plans for future work.

2 RELATED WORK

A report from the NSF-sponsored *Professionalization in Cyberinfrastructure* workshop in 2017 [8] outlined two main issues concerning RCD personnel: RCD personnel are scarce, and cyberinfrastructure is different from traditional Information Technology (IT). The workshop highlighted the critical role that RCD professionals play in supporting next-generation scientific research. With that knowledge, a second NSF-sponsored workshop, *Building the Research Innovation Workforce*, was conducted in Fall of 2020 [3]. Lessons learned outlined the clear need for a coherent, collective, and coordinated national strategy and action plans to address several factors that inhibit the expansion and sustainment of a healthy cyberinfrastructure and research computing workforce ecosystem. These included:

- The need for recognition and a clear definition of the different roles (or "facings") reflecting duties in the workforce;
- The need for a viable career path with a reasonable level of funding and location stability, and the availability of training and education necessary for advancement and upskilling;
- The need for a concerted effort to address diversity, equity, and inclusion;
- The need to formalize education and training to create a coherent body of transferable knowledge for the national community; and
- The need to increase understanding, awareness, and acknowledgment of the essential role of cyberinfrastructure and research computing as a core element of the research enterprise.

A third effort, a three-month study done in June of 2021 by the CaRCC RCD Professionalization working group was also influential in our work [18]. This survey inquired about staffing across the RCD facings, associated pay ranges, institutional demographic, job turnover, and position types (e.g., permanent vs. term-based). We relied upon this work to align the demographic data requested and obtained by our survey, as the focus of their survey was to provide a high-level description of who is in the field of RCD in the US today. Most Career Arcs survey respondents had worked in academia (94%); a slight majority (54%) had also worked in a corporate environment; and significant numbers had worked elsewhere: 22% for a Non-academic Non-profit, 20% in Government, 10% in Federal Labs, and 17% had been self-employed. The Workforce survey only asked for the current employer, but was also heavily weighted toward Academia (87%) with only 1% each from Corporate and Government, and 10% Other. The Career Arcs survey had a 58%/35%/4% male/female/other splits (3% did not state), while the Workforce study was more heavily weighted toward men with a 70%/24%/2% distribution (4% did not state). Additionally, the Career Arcs survey had 71% heterosexual responses, 17% other, and 12% choosing not to say, while the Workforce Study had an 88%/4%/8% distribution. The two studies had very similar representation in Race/Ethnicity,

Citizenship, and Disability categories, indicating that our data is broadly representative of the RCD community in the United States. Those numbers are as follows: Race/Ethnicity: White (83%), Asian (8%), Pacific (1%), Black (1%), Native (1%), Other (1%), Did not say (6%).

Our paper focuses on what attracts people to RCD roles and influences the decisions they make over the course of their career. This topic was partly inspired by the research literature on "organizational socialization", which, according to Saks and Ashforth [24] received much attention in the 1990s. In [16], Kramer divides organizational socialization into three stages: a) pre-arrival (anticipatory socialization), b) entry (initial participation or encounter), and c) metamorphosis (becoming/being an active and established full member). The third stage of metamorphosis is also when organizational members learn to manage their roles and/or change and acquire new roles. Spagnoli explains that organizational socialization is key to career growth and work outcomes [26]. Furthermore, Nifadkar talks about how newcomers could be "blank slates" [22]. This is where Fang et al.'s argument [11] of organizational insiders is key. These insiders play an important role in newcomers' experiences, especially in effective mobilization and access to social capital and social networks in their organizations. While the first three stages make up the bulk of members' experiences with a work organization, Kramer also talks about a fourth stage of exit (or disengagement) - when a member leaves an organization. The research literature on organizational socialization is important to this paper because our overall goal in this study is to understand what attracts individuals to the RCD field throughout the different stages of their careers.

Given the literature referenced above, our particular study examines the key concepts of organizational socialization in the RCD context. More specifically, we included questions related to pre-arrival (and the possible introduction to RCD by existing RCD insiders/contacts), metamorphosis (including satisfaction, career growth, role expansion, and upward mobilization), and exit (leaving the RCD field). Specific questions that get at these concepts can be found in Section 3 of this paper. One of our goals is to build on previous work done in the field, including describing what facilitates entry into, movements within, and potential disengagement from RCD.

3 METHODOLOGY

An IRB approval was obtained to run the RCD Career Arcs Survey in November 2021. Given the specialized nature of the target population, we employed purposive sampling techniques [6] in data collection. More specifically, invitations to complete the survey were distributed through a variety of distribution lists and RCD communities between November 2021 and late January 2022. The target population was working professionals in RCD, whether in a research university, research lab, self-employed, or working in industry. Specifically, we targeted CaRCC groups, various EDUCAUSE distribution lists, XSEDE Campus Champions, SIGHPC, Women in HPC, Data Science lists, Regulated Research Community, and SURA among others. For these target lists, we worked with list administrators to ensure that the invitation was relevant to the group.

Our outreach plan included posting to diverse communities that represent RCD professionals in various RCD roles using the CaRCC *facings* model[4]. These include: *Researcher Facing* Roles (research computing and data staffing, outreach, and advanced support, etc.), *Data Facing* Roles (data creation; data discovery and collection; data analysis and visualization; research data curation, etc.), *Software Facing* Roles (software package management, research software development, optimization, and troubleshooting, workflow engineering, etc.), *System Facing* Roles (infrastructure systems, systems operations, and systems security and compliance), and *Strategy and Policy Facing* Roles (RCD program leadership and management).

The RCD Career Arcs survey includes data from 225 respondents residing in the US, collected in an anonymous manner - respondents were not asked any personal questions or institutional affiliations, and no compensation was offered for survey participation. Of the 225 respondents, 159 answered all the questions, and 66 provided partial data with some missing responses. The survey consisted of 31 questions with a median completion time of 12.06 minutes.

Fig. 1 shows the distribution of respondents based on the RCD roles they have had throughout their career, using the CaRCC *facings* model (note that respondents could select more than one facing). For a number of questions, we aggregated all the results to rank the factors, and then considered the responses through the lens of these facings. In many cases, we saw relatively little variation across the facings, which may in part be due to the fact that respondents could indicate more than one facing as a part of their job. Indeed, only 16.3% of respondents indicated a single facing, with *Software Facing* the least likely to be distinct. All five facings were marked by 18.3% of respondents and after this, the most common patterns were *Research Facing + Strategy and Policy Facing* (7.7%) and all except *Strategy and Policy Facing* (6.7%). For
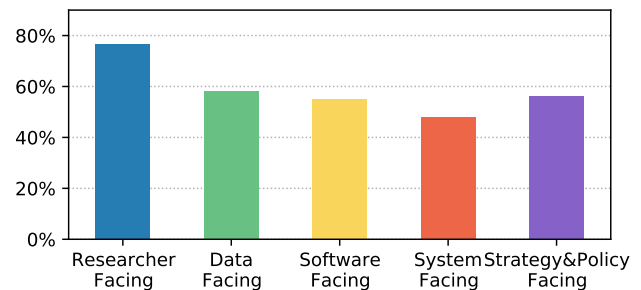

Figure 1: Proportion of respondents by facing.

any given facing, the most likely second facing was *Researcher Facing*, which is not surprising given that in general, RCD roles are distinguished (e.g., from traditional IT roles) by the close engagement with researchers. *System Facing* was least likely to be paired with another facing, possibly indicating that *System Facing* roles are mostly likely to specialize. The exception was *Software Facing*, for which the *Strategy and Policy* was least likely to also be part of their role.

Disciplinary backgrounds range from arts and humanities (14%), engineering, computer and information sciences (57%), life and

---

[4]https://carcc.org/facings

health sciences (13%), mathematical and physical sciences (32%), social, economic and behavioral sciences (8%) and other (4%) (respondents could select more than one, so percentages sum to more than 100%). The gender distribution of the survey respondents was 34.6% *female*; 58.5% *male*; 3.8% *gender non-conforming/genderqueer, non-binary/third gender* or *other*; and 3.1% preferred not to state. Most respondents were 35 to 54 years old (57%), with the rest skewing slightly older. Although we set a minimum age of 20 to participate in the survey, no one under 25 responded (or they preferred not to state their age). We set the minimum age as well as US residency to simplify data collection by excluding special handling of data that is required of data related to minors and for example, data subject to General Data Protection Regulation (GDPR). A majority (93%) of our respondents have worked in academia, over half have worked in the private sector, and quite a few have worked at federal labs (10%), for the government (20%), or for non-profits (22%). Nearly 17% have been self-employed at some point during their career. Respondents have had an average of 2.8 RCD positions, and moved often: 43% stayed less than 3 years in positions, and 75% less than 6 years.

3.1 Analysis Methodology

The primary objective of the RCD Career Arcs survey questions was to identify factors that attract professionals into the field of research computing and data, factors that influence their career advancements or changes within the RCD field, and factors that cause them to leave and pursue other fields. Questions of the CaRCC Career Arcs survey[5] fall under the following broad areas of research questions we are exploring:

(1) **General characteristics of RCD professionals and their RCD career**
This set of questions explores the familiarity of survey respondents with the RCD field at the time they got hired and at present. We ask how many years they have been in the field and how much of their RCD career was at an academic institution. We also ask what roles they have taken on over the course of their RCD careers.

(2) **Factors for RCD career entry, advancement, and satisfaction**
These questions ask respondents to rank the importance of factors that lead to being hired in RCD roles, both from the point of view of the applicant and the hiring manager. We also investigate how RCD professionals view career advancement and satisfaction.

(3) **Factors for changing jobs, and/or leaving RCD for other domains**
These questions ask respondents to consider factors that may cause one to consider a change of jobs, such as opportunities to move into management or to have more influence; higher salary or better benefits; professional development; and other factors. For each factor, the respondents rated the importance of each factor on a five-point Likert scale.

We asked respondents how many years of total experience they had, as well as how many years of experience they had in RCD roles. We noted that the overall experience was quite a bit higher than the RCD experience, and where the RCD experience was mostly less than 10 years, overall experience was distributed across a much wider range. We resolved to focus on RCD experience and created three bins to define *career stages* that we could use in slicing the data for analysis. We defined the cuts such that we got three nearly equally-sized groups among the respondents; this resulted in: "Early Career" (up to 7 years), "Mid-Career" (7 to 14 years), and "Advanced Career" (over 14 years).

Several of the questions asked respondents to rank factors from a list in order of importance; they were required to rank at least 5 factors, but could rank more (up to as many as 17 factors). We combined the answers by assigning a weight to each rank that was the inverse of the position (i.e., *number of factors in the list − respondent's rank*) and then summing the rank-weights for all respondents, for each factor (unranked factors were assigned a rank-weight of 0). We sorted the factors in decreasing weight order to get an overall ranking of the factors for a given question, and then sliced the data along various dimensions such as gender, RCD facing, and RCD career stage to compare how different sub-groups ranked the factors. Most of this analysis methodology was implemented as a Jupyter notebook[6] using the pandas data analysis library[7].

In the next section, we highlight some initial findings from the survey data, show the correlation of the various factors above and how demographics play a role in these factors. Additional data and visualizations are available in [23].

4 RESEARCH FINDINGS

4.1 What *Advancement* means across Career Stages

We posed the question "What does advancement in your current RCD role mean to you?" and asked respondents to rank at least 5 of 12 factors. The rankings are presented in Fig. 2, sliced by RCD Career Stage. Overall, the data show how important recognition is to RCD Professionals (this factor was ranked well above all others, and influence was also in the top three. Second highest was salary and benefits (although this was the top factor for early career professionals). Somewhat lower in the rankings, there is a cluster of professional development factors (and early career respondents value these more than those with more experience).

Of note is that "Progressing up a series of titles" and "Rising to management" both rank quite low overall, although "Rising to management" is comparatively more important to early career respondents, and much less important for advanced career respondents. Also worth noting is that although respondents ranked salary quite high in their definition of *Advancement*, they ranked it much lower in importance as a factor motivating them to switch jobs (see next section). It seems that people recognize that salary is often a marker of advancement, but it is not as important to RCD Professionals in making decisions about their career.
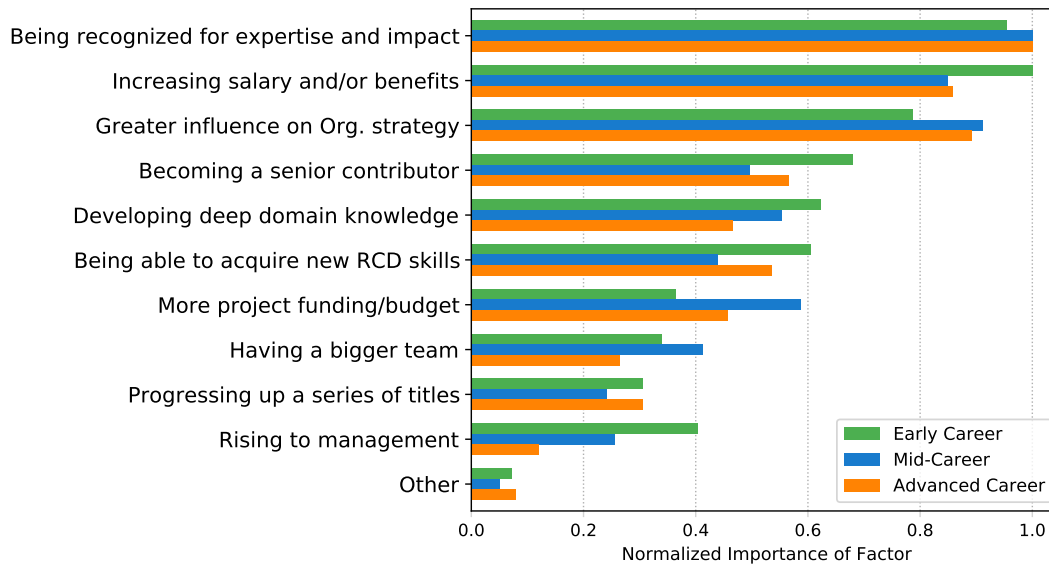

Figure 2: Ranking of factors defining "Advancement", sliced by RCD Career Stage.

4.2 The Importance of Factors motivating job changes for different RCD Facings and Career Stages

One of our goals was to understand why people change jobs over the course of their career. We asked respondents: "How important were the following factors in motivating you to make a previous job transition, or that would motivate you to consider a future job transition, to or within the RCD field (i.e., to a new RCD position or role)" using a 5-point Likert scale (see Fig. 3). The top three factors are associated with impact and personal growth, i.e., opportunities for: a more meaningful contribution (1), professional development (2), and joining a more innovative organization using cutting-edge tools (3). The next two factors in order were more practical: more flexible hours and/or a better work-life balance (4), and a higher salary (5). It is worth noting that ranking salary 5th, the ability to work remotely 10th, and better benefits 12th stands in contrast to the current discussion of how many workers in general are leaving jobs in search of higher pay, better benefits, and remote work opportunities [19]. Relocation is also cited as a major reason people are changing jobs these days, but it was the lowest ranked factor in this survey.

We saw a broad agreement across the facings, although *Strategy and Policy Facing* roles put much more weight on the opportunity to have more influence, and somewhat more on opportunities for promotion or advancement, for greater community engagement, and to relocate. Similarly, we saw relatively little divergence by Career Stage, with a few exceptions. The ranking of *Opportunity to have more influence* increases with career experience while the interest in working remotely decreases with career experience. Mid-career professionals are more likely to move for a better cultural fit, or due to a loss of funding for their position, but they are much less interested in gaining experience in other domains. Early career

professionals are the most interested in relocation, although it is still ranked low for them.

4.3 The Role of Gender in the RCD Profession

In this section, we explore what factors lead women into the RCD profession as compared to men and non-binary people. We also look at how they define advancement and career satisfaction and how it is different. Thirdly, we want to look at what factors motivate women to leave this field.

We had 34.6% women respondents in our survey with the majority pursuing researcher-facing, data-facing and strategy & policy-facing roles. Women represented only 26% and 22% respectively in the software- and systems- facing roles, and are markedly under-represented in these roles as compared to men who dominate (∼70% each) in these two facings. This finding is in alignment with studies that show that men get to have innovative assignments (HPC, software engineering, systems administration) while women are pushed into project management, business analysis, quality assurance, and technical recruiter type roles [12, 14, 15].

According to our survey, the most important factor that women consider for pursuing a career in RCD is interpersonal skills while the top factor for men is technical skills – even though the overall order of most important ranked factors was the same: interpersonal, communication and related skills; experience in academic research projects and technical skills. Men were much more likely to believe that technical skills, projects they had worked on, and their years of experience were important in getting hired into their RCD roles, while women were much more likely to believe that interpersonal and communication skills, leadership skills, and a referral from someone were important. Women were also somewhat more likely to believe that their experience and understanding of academic research projects, and their degree, were important.
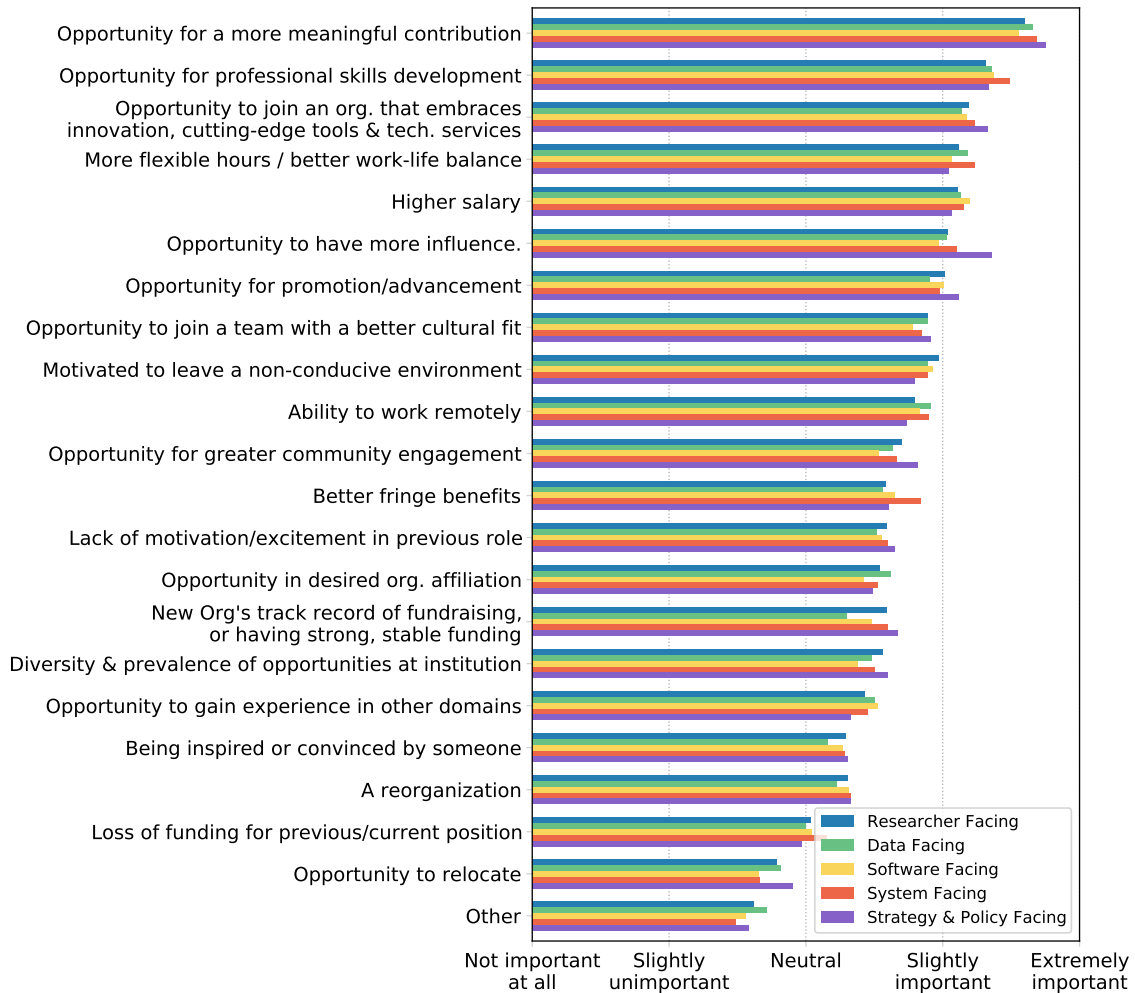
Figure 3: Ranking of factors in making a job move, sliced by RCD Facing.

However, when observed from the perspective of a hiring manager, both men and women consider technical skills to be the most important factor for hiring people into an RCD role (see Fig. 4). Compared to men, female hiring managers give a higher importance to interpersonal and communication skills; experience in understanding of research projects; previous projects that candidates had worked on; degree and domain of degree; years of overall experience; places worked in the past; and technical certifications. On the other hand, men rely more heavily on referrals from someone when hiring as well as experience working with a group as a student.

The top five factors that people consider for switching jobs were opportunity for a more meaningful contribution, opportunity for professional skills development, opportunity to join an innovative organization, better work-life balance, and higher compensation as seen in Fig. 6. An interesting observation here was that men rated better work-life balance slightly higher than women while women rated better compensation slightly higher than men as a factor for making a switch. Women were also more likely to consider

switching jobs to join a team with a better culture fit, as well as seeking to move by being inspired or convinced by someone.

The top three important defining factors for career advancement for all people were recognition for expertise and impact, increasing salary and/or benefits, and greater influence on organizational strategy. However, it is to be noted that women consider recognition and impact to be a far more important factor than compensation. Men rated both better compensation and recognition/impact as equally important. Men were also more likely to rate becoming a senior contributor, developing deep domain knowledge, and being able to acquire new RCD skills as factors of career advancement, as compared to women.

Furthermore, we observed that of our respondents, there are relatively fewer women than men who have advanced experience in the RCD field as compared to early and mid-stage RCD experience as shown in Fig. 5. Here, we defined early career stage to be < 10 years, mid-career as having 10 − 20 years of experience in the RCD field and > 20 years as advanced experience. This is a different split than other career stage charts included in this paper to be more
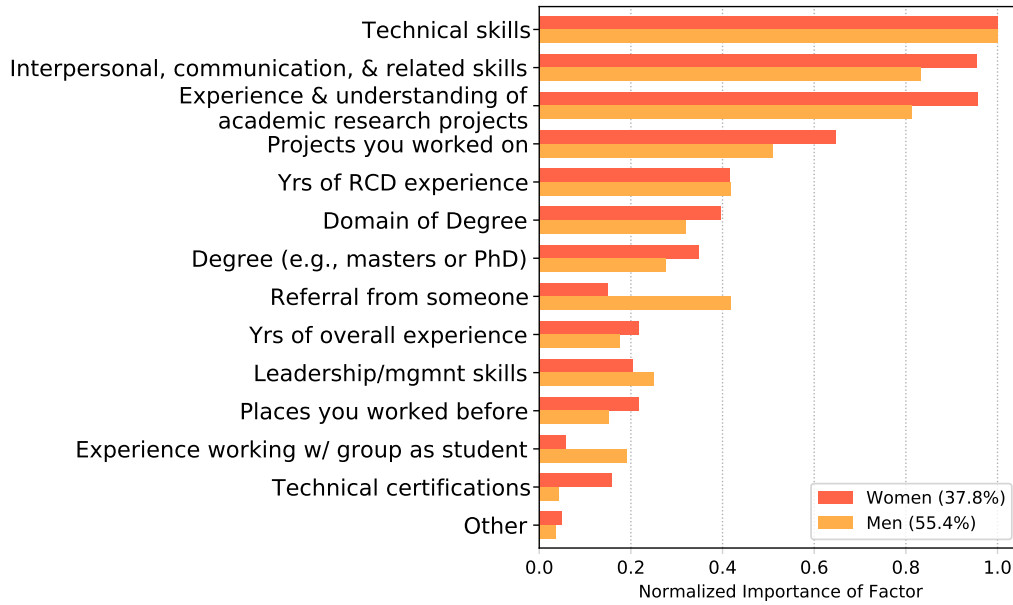
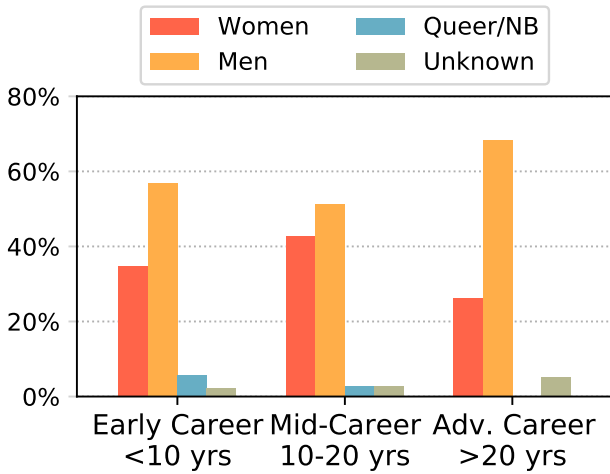**Figure 4: Importance of factors when hiring candidates into RCD roles - by hiring manager gender.**

Figure 5: Genders by RCD career stage.

aligned with the technology career stage definition used in [4, 14]. This female-to-male imbalance observed in Fig. 5 may be due to female attrition and women not working in the RCD long enough to reach the advanced experience stage. This needs to be explored further.

Thus, we observe two main future efforts needed here: there is a need to attract more women into the RCD field, especially in systems and software-facing roles. And secondly, there is possibly a need to propel more women from mid-career into advanced-career levels.

5 CONCLUSION AND FUTURE WORK

We presented an analysis of factors that can influence the career arcs of RCD Professionals, based upon data gathered in a recent survey. We described how gender, career stage, and different RCD roles impact the ranking of these factors, and note that while there are differences across these groups, respondents were broadly consistent in their assessment of the importance of these factors. In some cases, the responses clearly distinguish RCD Professionals from the broader workforce, and even other IT professionals.

This work is part of a larger project to describe different possible paths for RCD roles, and to help hiring managers recruit and retain people in these roles. In the next phase of our work, we are looking to gather deeper individual narratives through a series of interviews of RCD professionals in various phases of their RCD career journeys. Our hope is that these narratives will illuminate possibilities for individuals considering or already in RCD careers, as well as helping hiring managers understand fruitful domains and populations for recruitment. The survey questions and response data presented here will be used to refine our interview questions.

There is more data and additional analyses that we plan to conduct and share in future papers. In particular, we have questions about factors that influence someone to leave an RCD role for a different kind of job, but we are also interested in more finely segmenting the data to see if other interesting patterns emerge (we provide our current data and more visualizations as auxiliary data in [23]). We believe that there is value in repeating this survey in future years to understand how these factors change over time, and as the RCD profession evolves.

We note a couple of limitations of this study. First, given that participation is voluntary, the sample represents RCD professionals who were willing and available to fill out our survey. Thus, the sample may not fully represent the entire community in some ways.
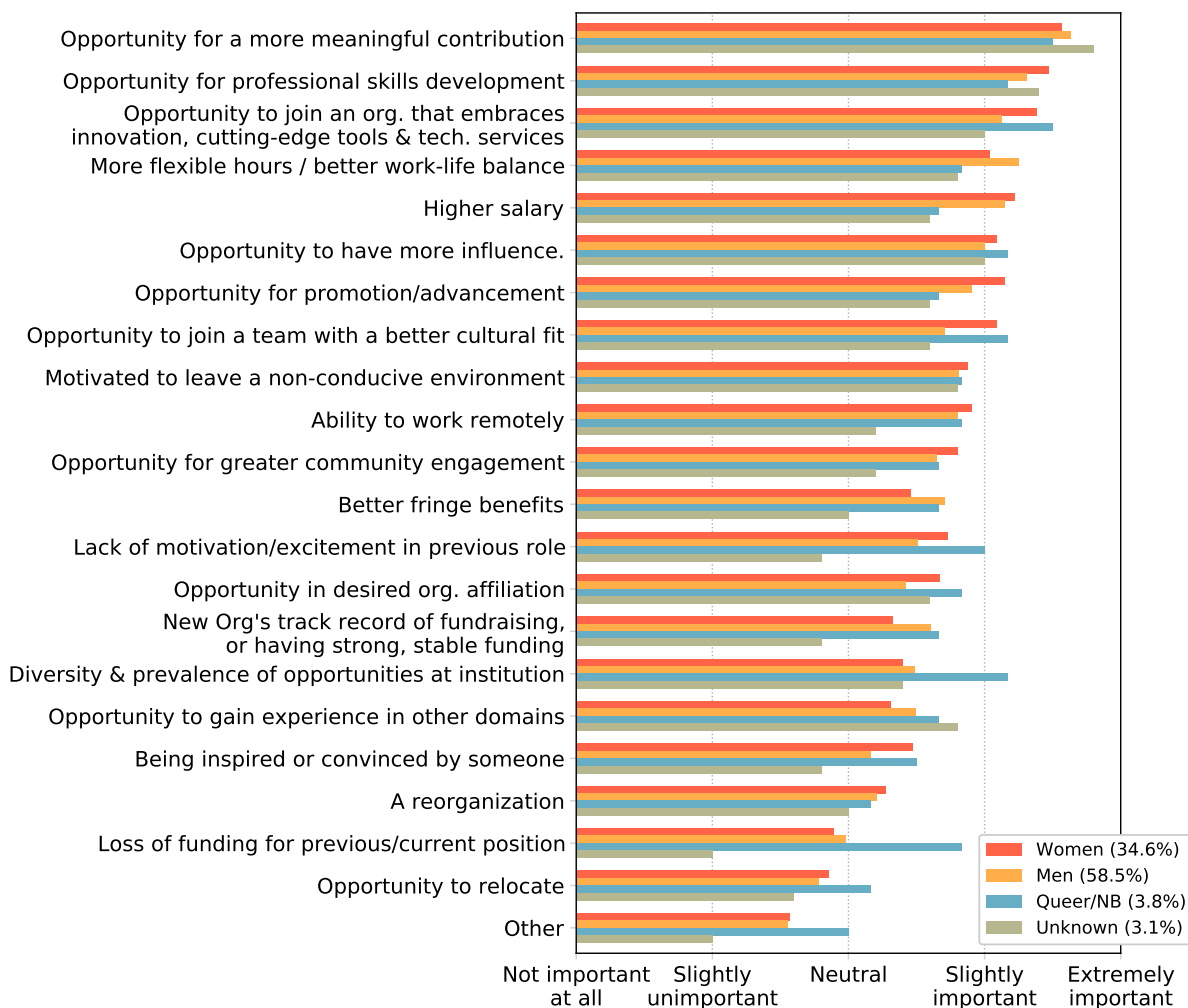
Figure 6: Importance of factors for considering a job switch from the perspective of various genders.

Second, our analysis is based on the sample without calculating the statistical significance of comparisons and correlations. Therefore, we are cautious about the generalizability of our findings. We will work to increase the sample size in future surveys, perhaps by combining our questions with other broad surveys of the RCD professional community. Third, since the study was completed during the COVID-19 pandemic, the findings may not be a representation of the feelings the survey participants would have had during non-pandemic period. Nonetheless, we believe the findings reported in this paper represent one of the first and most comprehensive attempts to date to describe the complex RCD career arcs.

REFERENCES

[1] Amy Adkins and Brandon Rigoni. 2016. *Millennials want jobs to be development opportunities*. Gallup, Inc. Retrieved May 26, 2022 from http://news.gallup.com/businessjournal/193274/millennials-jobs-development-opportunities.aspx

[2] National Security Agency and the Department of Energy. 2016. *U.S. Leadership in High Performance Computing (HPC) A Report from the NSA-DOE Technical Meeting on High Performance Computing*. Technical Report. NITRD. https://www.nitrd.gov/nitrdgroups/images/b/b4/NSA_DOE_HPC_TechMeetingReport.pdf

[3] Lisa Arafune, Dana Brunson, Thomas Hacker, and Preston Smith. 2021. *Building the research innovation workforce: a workshop to identify new insights and directions to advance the research computing community*. Technical Report. https://www.rcac.purdue.edu/files/ciworkforce2020/report.pdf

[4] Catherine Ashcraft, Brad McLain, and Elizabeth Eger. 2016. *WOMEN IN TECH: THE FACTS 2016 Update.* Technical Report. NCWIT Research. https://wpassets.ncwit.org/wp-content/uploads/2021/05/13193304/ncwit_women-in-it_2016-full-report_final-web06012016.pdf

[5] Lindsay Barone, Jason Williams, and David Micklos. 2017. Unmet needs for analyzing biological big data: A survey of 704 NSF principal investigators. *PLOS Computational Biology* 13, 10 (10 2017), 1–8. https://doi.org/10.1371/journal.pcbi.1005755

[6] L.A. Baxter and R. Babbie. 2003. *The Basics of Communication Research.* Cengage Learning. https://books.google.com/books?id=PTE9AAAAQBAJ

[7] Michela Bello and Fernando Galindo-Rueda. 2020. Charting the digital transformation of science: Findings from the 2018 OECD International Survey of Scientific Authors (ISSA2). *OECD Science, Technology and Industry Working Papers* No. 2020/03 (2020). https://doi.org/10.1787/1b06c47c-en

[8] Nicholas Berente, James Howison, Joel Cutcher-Gershenfeld, John L. King, Stephen R. Barley, and John Towns. 2017. *Professionalization in Cyberinfrastructure.* SSRN Scholarly Paper ID 3138592. Social Science Research Network, Rochester, NY. https://doi.org/10.2139/ssrn.3138592

[9] Nicholas Berente, James Howison, John L. King, Stanley Ahalt, and Susan Winter. 2018. *Organizing and the Cyberinfrastructure Workforce.* SSRN Scholarly Paper ID 3260715. Social Science Research Network, Rochester, NY. https://papers.ssrn.com/abstract=3260715

[10] Christopher Brooks, Joseph Galanek, Dana Gierdowski, and Mark McCormack. 2019. *The IT Workforce in Higher Education.* Technical Report. EDUCAUSE Research. https://library.educause.edu/resources/2019/2/the-it-workforce-in-higher-education-2019

[11] Ruolian Fang, Michelle K. Duffy, and Jason D. Shaw. 2011. The Organizational Socialization Process: Review and Development of a Social Capital Model. *Journal of Management* 37, 1 (Jan. 2011), 127–152. https://doi.org/10.1177/0149206310384630 Publisher: SAGE Publications Inc.

[12] Athina Frantzana. 2019. *Women's Representation and Experiences in the High Performance Computing Community.* Ph.D. Dissertation. The University of Edinburgh, Edinburgh, Scotland. https://era.ed.ac.uk/bitstream/handle/1842/36127/Frantzana2019.pdf

[13] John Goodhue, Julie Ma, Adrian Del Maestro, Sia Najafi, Bruce Segee, Scott Valcourt, and Ralph Zottola. 2020. Northeast Cyberteam Program–A Workforce Development Strategy for Research Computing. *The Journal of Computational Science Education* 11, 1 (Jan. 2020), 8–11.

[14] Sylvia Hewlett, Carolyn Luce, Lisa Servon, Laura Sherbin, Peggy Shiller, Eytan Sosnovich, and Karen Sumberg. 2008. The Athena Factor: Reversing the Brain Drain in Science, Engineering, and Technology. *Harvard Business Review Research Report* (May 2008), 108 pages.

[15] Dawn Kawamoto. 2016. Parity for Some, But Tech Women's Pay Lags Overall. https://insights.dice.com/2013/03/22/it-salaries-for-women/

[16] M. Kramer and M.W. Kramer. 2010. *Organizational Socialization: Joining and Leaving Organizations.* Wiley.

[17] LinkedIn Learning. 2022. *2022 Workplace Learning Report.* Technical Report. LinkedIn. https://learning.linkedin.com/resources/workplace-learning-report

[18] 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).* ACM, Boston, MA, USA. https://doi.org/10.1145/3491418.3530289

[19] Lucas Mearian. 2022. *No end in sight for the Great Resignation; workers keep quitting for better pay, benefits.* Retrieved May 26, 2022 from https://www.computerworld.com/article/3646390/no-end-in-sight-for-the-great-resignation-workers-keep-quitting-for-better-pay-benefits.html

[20] National Science Foundation. 2021. Research Infrastructure Guide. https://www.nsf.gov/pubs/2021/nsf21107/nsf21107.pdf

[21] Henry Neeman, Hussein M. Al-Azzawi, Aaron Bergstrom, Zoe K. Braiterman, Dana Brunson, Dirk Colbry, Eduardo Colmenares, Akilah N. Fuller, Sandra Gesing, Maria Kalyvaki, Claire Mizumoto, Jeho Park, Anita Z. Schwartz, Jason L. Simms, and Rustomji Vania. 2018. Progress Update on the Development and Implementation of the Advanced Cyberinfrastructure Research & Education Facilitators Virtual Residency Program. In *Proceedings of the Practice and Experience on Advanced Research Computing* (Pittsburgh, PA, USA) *(PEARC '18).* Association for Computing Machinery, New York, NY, USA, Article 71, 7 pages. https://doi.org/10.1145/3219104.3219117

[22] Sushil S. Nifadkar. 2020. Filling in the "Blank Slate": Examining Newcomers' Schemas of Supervisors During Organizational Socialization. *Journal of Management* 46, 5 (May 2020), 666–693. https://doi.org/10.1177/0149206318807288 Publisher: SAGE Publications Inc.

[23] Patrick Schmitz, Shafaq Chaudhry, and Arman Pazouki. 2022. *Extended data and graphs for the 2022 Career Arcs survey.* Technical Report. CaRCC. https://doi.org/10.5281/zenodo.6503005

[24] Alan M. Saks and Blake E. Ashforth. 1997. Organizational Socialization: Making Sense of the Past and Present as a Prologue for the Future. *Journal of Vocational Behavior* 51, 2 (Oct. 1997), 234–279. https://doi.org/10.1006/jvbe.1997.1614

[25] 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 Computing in Science & Engineering.

[26] Paola Spagnoli. 2020. Organizational Socialization Learning, Organizational Career Growth, and Work Outcomes: A Moderated Mediation Model. *Journal of Career Development* 47, 3 (June 2020), 249–265. https://doi.org/10.1177/0894845317700728 Publisher: SAGE Publications Inc.