Building the Research Innovation Workforce: Challenges and Recommendations from a Virtual Workshop to Advance the Research Computing Community*

THOMAS HACKER†, Purdue University, USA
PRESTON SMITH†, Purdue University, USA
DANA BRUNSON†, Internet2, USA
LISA ARAFUNE†, Forge Policy Solutions, USA
THOMAS CHEATHAM, III‡, University of Utah, USA
EWA DEELMAN‡, University of Southern California, USA

The workforce for research computing, cyberinfrastructure, and data analytics is a complex global ecosystem comprised of workers across academia, national laboratories, and industry. To explore the underlying factors that affect the growth and vitality of this workforce ecosystem, we conducted an NSF funded virtual workshop during the third quarter of 2020 attended by 100 participants. The workshop identified challenges affecting the workforce pipeline and ecosystem and generated recommendations to help address these challenges. This paper provides a summary of the workshop, challenges, and recommendations.

CCS Concepts: • **Social and professional topics**;

Additional Key Words and Phrases: workforce, research computing, cyberinfrastructure

ACM Reference Format:
Thomas Hacker, Preston Smith, Dana Brunson, Lisa Arafune, Thomas Cheatham, III, and Ewa Deelman. 2022. Building the Research Innovation Workforce: Challenges and Recommendations from a Virtual Workshop to Advance the Research Computing Community. In *Practice and Experience in Advanced Research Computing (PEARC '22), July 10–14, 2022, Boston, MA, USA.* ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3491418.3530288

1 INTRODUCTION

Modern high quality and innovative research relies on accurate and reproducible simulation, data analysis, and provenance preservation. Moreover, this research is frequently conducted by widely dispersed communities and virtual organizations that can cross institutional, industry, and national boundaries. Software must be written, adapted for use on ever-changing infrastructure, and used as an integral component of research projects. The level of training, skill, and necessary years of experience to develop high quality research computing and data infrastructure, services, and support, upon which the veracity of research results relies, requires a highly skilled and dedicated research innovation workforce (RIW) familiar with collaboration with domain science oriented researchers in developing, deploying, using, and supporting cyberinfrastructure.

*This paper is an extended and adapted version of the final workshop report [1], which is available at https://www.rcac.purdue.edu/ciworkforce2020/report/report.pdf.
†Workshop Co-Chair
‡Workshop Steering Committee Member

Manuscript submitted to ACM

The community of researchers, institutions, and supercomputer centers, as well as a recent (November 2019) U.S. government report from the National Science & Technology Council (NSTC) [8, p. 5-6] have reported the need for a trained cyberinfrastructure workforce and the availability of highly skilled and trained professionals who are necessary to create, deploy, and maintain the critical research computing and data infrastructure that is required for discovery and innovation. The NSTC report notes that the effort to put new computing technologies to work for applications that drive research, "…requires a skilled workforce, trained in the current state-of-the-art, that is able to anticipate and exploit future technologies and solutions. This involves a wide-range of training and skills, from cable laying to power and heat management to research and development to marketing and communication." [8, p. 5].

The ready availability of a skilled research innovation workforce is essential for the future of research. The vitality of the research and innovation enterprise rests on two pillars: advances in understanding driven by the quest for new knowledge; and putting this new knowledge to work through the creation of artifacts with a dual use of solving problems and fueling further advancement in new knowledge. Limitations in the size of the skilled workforce equipped with the knowledge and skills needed to transition knowledge to practice has the potential effect of significantly inhibiting the virtuous cycle of discovery and innovation that directly benefits humanity.

To explore the challenges and issues affecting the research innovation workforce, we held a virtual workshop across three sessions in August and September 2020. This paper is a summary of the workshop process, activities, findings, and recommendations identified by the workshop participants. This paper is an extended and adapted version of the final workshop report [1], which is available at https://www.rcac.purdue.edu/ciworkforce2020/report/report.pdf.

2 CYBERINFRASTRUCTURE, RESEARCH COMPUTING, AND DATA WORKFORCE

There has been a long history of computing within universities. A brief timeline of computing at the University of Michigan [11] provides some historic examples. A recent book by Peter Freeman, Et al. provides a chronological history of some of the National Science Foundation efforts in supporting education and research in computing [10]. Institutional organizations arose to provide support and training for the community of students, faculty, and staff who use computing to enhance or directly support teaching, research, and administration. Many of these university centered efforts led to the introduction of some of the first large-scale computer networks (e.g. the NSFnet - one of the predecessors of the Internet [9]) and software systems (such as Mosaic [19] and the World Wide Web at CERN [4]). Over time, as computers moved from specialized air conditioned rooms to desktops and laptops, many of the common commodity support functions were also transformed into what is now generally considered to be *Enterprise IT* [16].

Although many computing functions have been shifted to commodity, there remains a need for specialized support and training for the cadre of users for whom commodity enterprise IT resources are insufficient. Prof. Dan Atkins' report of the "National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure" [2] in 2003 - nearly 20 years ago - helped to refocus the research computing community on the need for a comprehensive approach to providing working infrastructure and support services to meet research user needs.

From the history of computing, it is abundantly clear that there is tremendous value and return on investment from academic and research computing efforts that spring from universities and research laboratories. These advances came from synergistic multidisciplinary efforts among research faculty, students, and highly skilled staff working at computer centers. It is critical to recognize that a successful and functional cyberinfrastructure must include *people* as an essential element that needs as much care and attention as the hardware and software. Moreover, it is clear from history that attention and investments focused on the people side of cyberinfrastructure - specifically the workforce who supports,

creates, and sustains the infrastructure - has significant potential impact on creating new technologies that can benefit humanity.

With this context in mind, there is a growing awareness that the growth in demand for skilled professionals has exceeded the available workforce, and that there is an immediate need to address factors that inhibit the growth and development of this workforce [3]. The NSF included the need for learning and workforce development in a recent (Dec 2020) report [14], and the National Science and Technology Council (November 2019) also recognized the need for a skilled and diverse workforce [8].

There are several existing efforts now underway that seek to address some of the challenges affecting the workforce. CaRCC (Campus Research Computing Consortium) [5] seeks to address issues affecting research computing and data (RCD) professionals in the workforce that include career development, building community, professionalization, and working to more clearly define roles within the workforce that are frequently present in the workforce today. Additionally, Neeman, Et. al. [13] describes an effort to operate a "Virtual Residency Program" to provide training to individuals involved in research computing. The paper by Neeman [13] also includes information that reflects the size of the need for skilled "CI Facilitators" [13] at research intensive (R1 and R2) institutions.

Two other efforts are focused on the role of *Research Software Engineer*. The US Research Software Engineer Association [20][6] aims to help develop a sense of community of research software engineering professionals and advocate for their involvement in research. In the United Kingdom, the Society of Research Software Engineering [15] seeks to promote the recognition and training for research software engineers who work with researchers and software experts, and seeks to promote the development of an academic career path for research software professionals.

3 WORKSHOP OVERVIEW

We conducted an NSF funded virtual workshop across three sessions in August and September 2020 to bring together the community with the overall aim of exploring issues and to devise potential solutions to the problems affecting the research innovation workforce that is essential to research computing and cyberinfrastructure. The workshop had two overarching goals. First, provide a set of recommendations to the National Science Foundation regarding actions to address the shortfall in availability of skilled research computing practitioners in the research innovation workforce. Second, provide a forum for community leaders to come together to discuss the issues and to document innovative strategies for consideration by the community to promote the professional and quantitative growth of the skilled research innovation workforce.

One item that the workshop co-chairs discussed at length was the question *what is the workforce?* We each had our individual backgrounds and history that informed our individual response to this question. It was clear from our discussions that the definitions of who makes up the workforce, where they work, and their roles is a complex question. The development of the workforce is often thought of as a *pipeline*, in which individuals over time progress through roles of increasing responsibility as they gain experience, education, and training. It became clear, however, that an *ecosystem* model provides a more accurate metaphor than a pipeline. The workforce evolves over time in response to changing technologies, demands, and needs of all members of the ecosystem. These members include: users, students, researchers, the public, funding agencies, industry and suppliers of hardware and software, teachers, and individuals in the workforce. Compared with the traditional career path for tenure track faculty (Assistant, Associate, to Full Professor), there are few well defined and stable career paths within the workforce ecosystem for which a new person embarking on a career can easily track and plan.

CaRCC developed a detailed collection of roles and role descriptions that reflect a concept they defined named *facings* [18]. A facing represents what a role focuses on. CaRCC defined[1] five facings [17][18]: researcher facing (directly supporting researchers); data facing (working with the data); software facing (focusing on software development and deployment); systems facing (focused on infrastructure); and strategy and policy facing (external and institutional relationship management). CaRCC defined a set of role definitions using these facings as a framework. Given this multifaceted nature, people in the workforce need training and development in skills and knowledge (to a varying degree) for all of these facings. Thus, a static pipeline metaphor fails to reflect the evolutionary and multifaceted nature of the roles and responsibilities of individuals in the workforce.

4 WORKSHOP PROCESS

The workshop brought together attendees who represented community *affinity groups*. Workshop attendees were identified and invited from affinity groups that included academics (from both computational science and other domain sciences), university and non-profit research computing leadership, governmental agencies, cyberinfrastructure practitioners and professionals, industry, as well as NSF observers. The attendee selection process was based on soliciting suggested attendees from the workshop co-chairs, the workshop steering committee, and from other invited attendees.

Our list of 223 invitations, based on our understanding of the attendee's affinity group, broke down as follows:

- 34 Practitioners and Skilled Professionals,
- 27 University and non-profit Research Computing Leadership,
- 27 from Industry,
- 14 from Academic Cyberinfrastructure (Computing focused),
- 34 from Academic Domain Sciences,
- 13 from Governmental Agencies,
- 16 Other, and
- 79 uncategorized.

A total of 120 responded to indicate they would attend. From this group:

- 13 self-identified as Cyberinfrastructure Practitioners and Professionals,
- 10 self-identified as Industry,
- 53 self-identified as University and non-profit Research Computing Leadership,
- 14 self-identified as Academic computer science/cyberinfrastructure research,
- 7 self-identified as Academic Domain sciences,
- 16 self-identified Government Agencies, and
- 7 self-identified as "Other"

Across all three virtual workshop sessions, there were 100 unique attendees. We did not expect the large number of respondents to self-identify as University and non-profit Research Computing Leadership. We examined the list, and determined that 4 could also be considered to be "Academic Domain Science", and we informed them via email that they were reclassified to that affinity group in order to balance between affinity groups based on our understanding of the dual roles of many of the attendees.

[1]CaRCC facings are described on the website https://carcc.org/rcd-professionalization/facings/

4.1 Conducting the virtual workshop sessions

All three virtual workshop sessions were conducted with the aid of KnowInnovation, an experienced workshop facilitator. The workshop facilitator and workshop co-chairs conducted the virtual workshop sessions using a multistage process. First, the workshop co-chairs and workshop facilitator collaborated to form a detailed agenda for each virtual workshop session. The first session on Wednesday August 19th kicked off the workshop with keynote presentations (summarized below), brainstorming challenges, and small breakout sessions to address and devise solutions to several thematic challenges. During Session 2 on Tuesday, August 25th, participants (again in breakout groups) synthesized ideas from Session 1 and created detailed recommendations. Participants then were given an opportunity in the virtual platform to tag and prioritize these recommendations. Following this session, the co-chairs and two steering committee members wrote up the recommendation summaries. In the final session on September 9th, the recommended actions and their votes were presented and reflected on by the participants. Then, in the spirit of the casual conversations following in person workshops, participants used the virtual platform to prepare to consider actions that could be taken in response to the recommendations together via self-selected topics and breakout sessions. The session wrapped up with brief summary reports from each of these groups and an opportunity for participants to respond with material objections to any of the recommendations and documented those in the virtual platform. An anonymous Qualtrics survey (reviewed and determined to be exempt by the Purdue University Human Research Protection Program Institutional Review Board (IRB-2020-814)) was conducted after the workshop and is described in the workshop report [1].

4.2 Summary of Keynote Presentations

The first virtual workshop session on August 19th began with keynote presentations from three speakers who discussed the current state of the problem space around the CI/RCD workforce.

The first speaker was Neil Chue Hong, Principal Investigator and founding Director of the Software Sustainability Institute at the University of Edinburgh. Hong spoke on "Sustaining a CI Workforce - Skills, Careers, and Diversity" [7]. He presented first on the Software Sustainability Institute, and its five goals around Software, Training, Policy, Community, and tied together by outreach, to enable engagement, delivery and uptake. Describing the history in the UK, from the 1970s, through the Parallel Applications Programme, UK eScience, the creation of Software Carpentry, and more recently the creation of the first Research Software Engineer (RSE) Fellowships in 2015.

The second keynote was from Sharon Broude Geva, Director of Advanced Research Computing at the University of Michigan, and chair of the Coalition for Academic Scientific Computation spoke on "Workforce Development : A Broad Perspective of (and on) Community Needs"[2]. Dr. Geva began by giving an overview of the CASC organization, then described what the cyberinfrastructure (CI) and advanced research computing workforce encompasses: research computing practitioners, center leadership, and thought leaders. For practitioners, several questions were posed around recruiting talent, providing career paths, and how leaders can describe the value of practitioners to enabling research.

Finally, Dana Brunson (a co-author of this paper and workshop co-chair), Internet2 Executive Director for Research Engagement and of CaRCC, Campus Champions and TrustedCI addressed the professionalization of the CI workforce with "Cyberinfrastructure Workforce: An Emerging Profession". Dr. Brunson presented an overview of existing communities, previous efforts, and ongoing work in the community. Dr. Brunson presented a summary of previous work on the professionalization of cyberinfrastructure work [3], describing previous workshops, the notions of "facings" towards researchers, software, strategy and policy, data, and systems that came out of those workshops, and the definition

[2]Sharon Broude Geva's presentation is available at https://www.rcac.purdue.edu/files/ciworkforce2020/geva.pdf

of "Facilitator" [12] - a researcher-facing role. Ongoing community-building activities such as Virtual Residency and CaRCC's people network were described. Finally, summaries of CaRCC-led work in professionalization and around HR considerations like job descriptions, job families, and career paths were presented to participants.

5 SUMMARY OF THEMATIC CHALLENGES IDENTIFIED DURING THE WORKSHOP

The keynote presentations were followed by the identification of challenges that were posted and voted on by participants using the KIstorm workshop tool that was managed by the workshop facilitator KnowInnovation. Small breakout groups then discussed the challenges and synthesized eleven critical "thematic challenges" and cross cutting issues that reflected common themes that emerged from the landscape of challenges identified by participants. For more details on these challenges, please refer to the full workshop report [1].

5.1 How do we recruit and sustain an inclusive and diverse workforce?

The diversity and inclusion of the CI/RCD workforce is negatively affected by several systemic factors. In terms of culture, there are rigid roles and expectations within the community that promotes the development of an 'in-group' that is difficult for new people to join. There is also a lack of role models within the community, and there is a computer science focused culture that is not open to perspectives from other fields. In terms of recruiting, there is a perceived focus on seeking new people from a limited number of majors and institutions. Finally, there is a lack of professional development and training in both technical topics as well and in areas related to diversity, inclusion, and working with people.

5.2 How do we create a workforce pipeline?

An element of this challenge is that using a metaphor of a *pipeline* fails to capture some of the critical essence of the realities of the workforce. Using an *ecosystem* metaphor more correctly represents an environment in which there are consumers and producers of knowledge and services, and that the workforce ecosystem provides an environment in which relationships are built and sustained, and transactions among participants occur with the goal of providing for the needs of all participants in the ecosystem. If we consider the workforce to be an ecosystem (rather than a pipeline), one challenge is that it is not clear where and how new people can enter into roles. Also, it is not clear what career paths are available or how a person can pursue advancement. Additionally, along with unclear roles and job descriptions, there is a lack of effective knowledge sharing and training opportunities to allow people to plan and pursue different roles within the workforce ecosystem.

5.3 How do we publicize, establish, and get recognition for CI careers and roles, both existing and as a viable career path?

The focus of this thematic challenge is on individuals who might be candidates to enter roles within the CI/RCD workforce. This challenge has several inherent elements. First, there is a lack of clear and well-defined job descriptions and roles within the CI/RCD workforce. There is a lack of role models and mentors within the community, and a poor understanding of reward structures that can serve to motivate individuals within the workforce. Moreover, there is an unmet need for common venues within which a community can gather to develop a group professional identity and to recruit talent from.

5.4 How do we communicate & sell value to various stakeholders?

This thematic challenge relates to stakeholders (such as institutional stakeholders) other than people who are in the CI/RCD workforce. In terms of identity and roles, there is a need for better understanding of who the stakeholders are, and the interests of the stakeholder groups. There is also the need for a clearer understanding of research computing and its role within the research enterprise. There is also a need for more representation of research computing professionals in advisory boards. In the area of communication, the inherent challenges include the need to clearly communicate the roles and value of CI/RCD staff, their technical capabilities, the inherent technical difficulty of the work, and successes achieved.

5.5 How to capture & communicate the state of the Ecosystem: successes, challenges, gaps?

This challenge relates to the workforce ecosystem and involves several areas: the need for clarity and a common understanding about CI/RCD; funding stability and priorities; and communication. For the first area, the challenge relates to the lack of a well communicated and common understanding of the definition and elements of the research computing and cyberinfrastructure workforce ecosystem, and the roles filled by cyberinfrastructure professionals within this ecosystem. Another area of challenge is the need for stable funding, and clear metrics that reflect what is important within the ecosystem. Finally, the communications about successes need a unified approach that can be addressed to the broader community and to the public.

5.6 How do we adequately compensate CI roles (both money and other compensation/experiences) to make them attractive?

This challenge relates to providing what is needed by the workforce in the near term (e.g. compensation and job stability/security), as well as to help individuals prepare for long term participation (e.g. reward mechanisms, career paths, and training) in the workforce.

Finally, in terms of skill development, there is a need for training and professional development, as well as more freedom to innovate and a model that would allow roles to evolve as the science changes.

5.7 How do we promote and ensure sustainability (funding / growth / retraining / size / non-volunteer) for the CI workforce?

The challenge of addressing sustainability is essential for the future health and growth of the CI/RCD workforce. This challenge relates to personnel issues and strategic planning. In terms of personnel, variability in funding for positions leads to attrition and turnover of skilled staff. Moreover, the paths for advancement and growth are limited. In the area of strategy, there has not been a parallel growth in funding for research infrastructure that has kept pace with increases in research funding. There also is a lack of focus on the role of research computing and cyberinfrastructure as a scientific instrument within the research enterprise.

5.8 Develop CI as a discipline that is independent of Computer Science.

A challenge that was discussed was whether the field of CI/RCD had reached the level of maturity at which it could be considered to be a discipline independent of computer science. It was not clear that this was major challenge. However, it is recognized that there is a need for parallel efforts to provide services to the research community as well as to advance the state of the art of the field of CI/RCD. The view of CI/RCD as a discipline depends on the context of

the discipline from which it viewed. From a domain science perspective, there is generally a lack of awareness and appreciation of software engineering and computing science fundamentals. CI/RCD is viewed as a support service. From a computer science perspective, many in computer science do not view CI/RCD as a well understood part of the discipline of computer science or as an academic discipline.

Another aspect of this challenge relates to individuals who rely on CI/RCD and those who provide services. There is a clear need for education and training programs to provide practical training for the graduate student research labor force, as well as more comprehensive education to produce a cohort of skilled research computing staff. The lack of clarity of roles and career paths for CI/RCD professionals is another issue.

5.9 Faculty, program managers at federal agencies, and CIOs need to be educated about the role of CI professionals.

There is a need for a greater awareness of the roles that faculty, federal agencies, and CIOs could play in aiding in the development and sustainment of the CI/RCD workforce.

5.10 Where the Research Computing function sits in the institution has a major effect in many of the issues listed as thematic challenges.

The organizational position of research computing can impact how research computing engages with the institutional research enterprise.

Placing research computing exclusively with the CIO or VPR office may not be ideal - there is no widely recognized "best practice" model. The unique history, role, and structure of individual institutions may require a unique organizational configuration to best serve the needs of the institution. The other facet is the role of researchers and faculty in research computing leadership and direction within an institution. Often, faculty and researchers have a clearer understanding of the need of research computing than institutional leadership.

5.11 How do we establish training & professional development for intermediate-level or staff transitioning careers? What is needed to support continuing development of people in CI careers?

Professional development and training is essential for the continuing education and development of the CI/RCD workforce. In terms of roles and training, there is no shared understanding across the community for the definitions of roles, career paths, and the competencies and skills needed for these roles. As the field evolves, there also needs to be recognition that the specifics of the roles will also change. There is also a lack of attention by leadership to the need to support (funding and staff) ongoing professional development and training, and that a cultural focus on training needs to be established within the community as a means to facilitating career paths and continued skills improvements of the CI/RCD workforce.

6 WORKSHOP RECOMMENDATIONS

Based on the discussion of the problems and thematic challenges, the workshop then focused on considering potential solutions. Twelve recommendations were generated during the workshop.

The detailed recommendations are described in the final workshop report [1]. The workshop co-chairs analyzed and reflected upon the recommendations, and we determined that the recommendations could be summarized by the communities or organizations to whom the recommendations seemed to be addressed. Details regarding the following material is also contained within the final workshop report [1].

6.1 Recommendations for the cyberinfrastructure and research computing and data (CI/RCD) community.

- Organize an "umbrella" professional organization (a "community of communities") that spans existing community groups to coordinate and agree on common standards, activities, and messaging to the public.
- Develop strategy and actions to communicate the impact of CI/RCD to institutional leadership, faculty, and students to raise the profile of CI/RCD and increase awareness of professional roles ("facings") and distinct career paths (that differ from enterprise IT) within the CI/RCD profession.
- Create a common set of job descriptions and career paths for CI/RCD.
- Develop a national census within the CI/RCD community to collect information on positions, pay grade, cost of living differences, and benefits to provide benchmarks for the profession, increase understanding of existing roles and compensation models for CI/RCD professionals, and improve retention and mobility options for the workforce.
- Develop a Diversity, Equity, and Inclusion action plan for the CI/RCD community. The research computing and data (RCD) community does not have the organized expertise within it that is needed to solve the problem, suggesting that efforts should involve outside experts (e.g. social psychologists and Diversity, Equity, and Inclusion professionals). The recommendation is to bring together the cyberinfrastructure (CI) community and experts in Diversity, Equity, and Inclusion (DEI) and social science to study the problem and develop a plan.

6.2 Recommendations for higher education institutions and other research organizations.

- Create or adopt common job descriptions and define a clear career path for research computing roles with compensation and promotion models that accommodates the dual research/service roles in the workforce ecosystem.
- Create multidisciplinary programs for CI/RCD that includes the involvement of institutional information and research organizations with a goal of developing a common core curriculum for CI/RCD.
- Improve communication about the value of CI/RCD to institutional stakeholders and leadership.
- Develop sustainable funding models for research cyberinfrastructure investments such as people, software, services, and resources on campus.
- Align research computing within the organization to ensure appropriate involvement and governance from the CIO, VP of Research, Deans, and other constituent groups.

6.3 Recommendations for the NSF.

- Advocate the inclusion of CI/RCD professionals in the proposal process in terms of budgeting, funding effort, recognition, and project leadership appropriate for their role in the effort.
- Encourage the establishment of institutional governance bodies for research computing, and encourage the establishment of cyberinfrastructure related activities at national annual meetings attended by leadership (e.g. EDUCAUSE, PEARC, and SC).
- Provide seed funding through unsolicited proposals, DCLs, and structured RFPs to incentivize community action to create working groups spanning institutions that could begin working on addressing the challenges identified in the workshop.

7 OBSERVATIONS AND LESSONS LEARNED FROM THE WORKSHOP

As described in the CI workshop report [1] and another paper by co-authors Cheatham and Brunson [18], this section summarizes a series of observations and lessons learned in the workshop. Please refer to the workshop report [1] for more information.

Overall, there is a clear need for a coherent, collective, and coordinated national strategy and action plan to address several factors that inhibit the expansion and sustainment of a healthy cyberinfrastructure and research computing workforce ecosystem. The specific factors that were often mentioned were:

- The roles and duties within the workforce are varied, and require skilled and trained personnel who are willing to invest years of their career in developing their expertise. These various roles and "facings" (a term originally developed at an NSF-funded workshop [3]) need to be clearly defined and recognized within the workforce;
- Given the commitment needed from individuals for these roles, the community needs opportunities for training and "upskilling" to build their skills and knowledge, and to allow individuals to build their career. There is also a need for stability in location and funding, as well as a career path to allow individuals to plan their training and aspirations over their career;
- A coordination group that would serve as an "umbrella" organization may be needed that could advocate for and help coordinate organizations (on an institutional through national level) and potentially encourage the sharing of activities and knowledge might be helpful;
- An organized and concerted effort is needed within the community to address diversity, inclusion, and equity issues affecting the workforce;
- There is a need to create an organized training and education program for the national workforce based on the definition of the material and skills needed by individuals; and
- Finally, a clear strategy is needed to communicate the impact of research computing and cyberinfrastructure and its role in research within institutions.

8 SUMMARY

Our reflection on the workshop has highlighted several aspects of the issues. It is clear that to attract individuals to the workforce, a coordinated effort is needed within the community to clarify the identity of the workforce, the relationship of the field of CI/RCD to computer science and domain sciences, and to articulate the value proposition of a career in the workforce. Compared with traditional academic roles, the entry points and paths within the workforce ecosystem are not clear.

For example, tenure track faculty positions have clear entry points with concomitant qualifications (Assistant Professor, Associate Professor, and Full Professor), a clear advancement *process* (promotion and tenure), and exit ramps. In contrast, for roles in the CI/RCD workforce, the entry points, promotion process, and advancement requirements are unclear and opaque. Further, incentive structures and performance expectations for CI/RCD professions are often not aligned with academic needs and service expectations as they are for faculty or research staff.

Individuals in the CI/RCD workforce have the potential to serve a significant role as a disciplinary bridge between computer science and the domain sciences that could greatly benefit both fields. As described in section two of this paper, the community has established a tremendous legacy of making significant contributions (e.g. the world-wide web) to society that goes well beyond traditional academic artifacts such as publications. Attention to the challenges

affecting the CI/RCD workforce is critical to address current needs, as well as to lay down a solid foundation to continue building the future workforce.

Being able to follow the recommendations made in the workshop will require a focused, sustained, and coordinated effort and engagement with the CI/RCD community, educational institutions, and funding bodies.

9 ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation under Grant No. 2036534. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. We would like to thank the workshop steering committee for their guidance and feedback during the workshop process. The members of the workshop committee were Linda Akli, Thomas Cheatham, Ewa Deelman, Dan Fay, Lauren Michael, Tsengdar Lee, and Frank Wurthwein. We would also like to thank the 100 workshop participants for their thoughtful comments and ideas shared during the workshop. The full list of workshop participants is listed in Appendix A of the final workshop report [1]. We would also like to thank KnowInnovation for their help in conducting the virtual workshop.

REFERENCES

[1] Lisa Arafune, Dana Brunson, Thomas Hacker, and Preston Smith. 2021. *Report of the Workshop Building the Research Innovation Workforce: A Workshop to Identify New Insights and Directions to Advance the Research Computing Community.* Technical Report. Retrieved March 2021 from https://www.rcac.purdue.edu/ciworkforce2020/report/report.pdf

[2] Daniel E Atkins, Kelvin K Droegemeier, Stuart I Feldman, Hector Garcia-Molina, Michael L Klein, David G Messerschmitt, Paul Messina, Jeremiah P Ostriker, and Margaret H Wright. 2003. Revolutionizing Science and Engineering Through Cyberinfrastructure. (2003). Retrieved February 2022 from https://nsf.gov/cise/sci/reports/atkins.pdf

[3] 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* (2017). https://doi.org/10.2139/ssrn.3138592

[4] Tim Berners-Lee, Robert Cailliau, Ari Luotonen, Henrik Frystyk Nielsen, and Arthur Secret. 1994. The world-wide web. *Commun. ACM* 37, 8 (1994), 76–82.

[5] CaRCC. 2022. *Campus Research Computing Consortium.* Retrieved May 4, 2022 from https://carcc.org/

[6] Jeff Carver, Ian Cosden, Julia Damerow, Charles Ferenbaugh, Sandra Gesing, Chris Hill, Daniel S Katz, Christina Maimone, and Lance Parsons. 2021. The US-RSE Association: Bringing Together Research Software Engineers. In *ACH 2021*. Association of Computers and the Humanities. https://doi.org/10.17613/b7gt-vv31 Poster presented at ACH 2021 virtual conference.

[7] Neil Chue Hong. 2020. Sustaining a CI Workforce: Skills, Careers, Diversity. https://doi.org/10.6084/m9.figshare.12827180.v1

[8] National Science & Technology Council. 2019. *National Strategic Computing Initiative Update: Pioneering the Future of Computing.* Technical Report. Retrieved January 2021 from https://www.nitrd.gov/pubs/National-Strategic-Computing-Initiative-Update-2019.pdf

[9] Karen D Frazer. 1996. *NSFNET: A partnership for high-speed networking: Final report, 1987-1995.* Merit Network. Retrieved February 2021 from https://www.merit.edu/wp-content/uploads/2019/06/NSFNET_final-1.pdf

[10] Peter A. Freeman, W. Richards Adrion, and William Aspray. 2019. *Computing and the National Science Foundation, 1950–2016: Building a Foundation for Modern Computing.* Association for Computing Machinery, New York, NY, USA.

[11] Bentley Historical Library. 2015. *A History of Computing at the University of Michigan.* Technical Report. The University of Michigan. Retrieved March 2021 from http://hdl.handle.net/2027.42/116206

[12] Lauren Michael and Bruce Maas. 2016. Research computing facilitators: The missing human link in needs-based research cyberinfrastructure. *Research bulletin. Louisville, CO: ECAR* (2016). https://library.educause.edu/resources/2016/5/research-computing-facilitators-the-missing-human-link-in-needs-based-research-cyberinfrastructure

[13] Henry Neeman, David Akin, Hussein Al-Azzawi, Kevin L Brandt, Jamene Brooks Kieffer, Dana Brunson, Dirk Colbry, Sandra Gesing, Anna Klimaszewski-Patterson, Claire Mizumoto, et al. 2020. Cyberinfrastructure Facilitation Skills Training via the Virtual Residency Program. In *Practice and Experience in Advanced Research Computing*. 421–428.

[14] Office of Advanced Cyberinfrastructure. 2020. *Transforming Science Through Cyberinfrastructure:National Data and Software Cyberinfrastructure.* Technical Report. https://www.nsf.gov/cise/oac/vision/blueprint-2019/Data-Software.pdf

[15] The Society of Research Software Engineering. 2022. *Society of Research Software Engineering.* Retrieved May 6, 2022 from https://society-rse.org

[16] Betsy Reinitz. 2016. What Is Enterprise IT? EDUCAUSE Blog. (22 April 2016). Retrieved May 4, 2022 from https://er.educause.edu/blogs/2016/4/what-is-enterprise-it

[17] Patrick Schmitz, Claire Mizumoto, John Hicks, Dana Brunson, Gail Krovitz, James Bottum, Joel Cutcher-Gershenfeld, Karen Wetzel, and Thomas Cheatham. 2020. *A Research Computing and Data Capabilities Model for Strategic Decision-Making*. Association for Computing Machinery, New York, NY, USA, 77–84. https://doi.org/10.1145/3311790.3396643

[18] 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 (2021), 19–27. https://doi.org/10.1109/MCSE.2021.3098421

[19] Charles Severance. 2013. Larry Smarr: Building Mosaic. *Computer* 46, 4 (2013), 6–8. https://doi.org/10.1109/MC.2013.140

[20] US-RSE. 2022. *United States Research Software Engineer Association*. Retrieved May 4, 2022 from https://us-rse.org/