Report of the Workshop

Building the Research Innovation Workforce:

A workshop to identify new insights and directions to advance the research computing community.

August - September 2020

Purdue University photo/Andrew Hancock

Workshop Co-Chairs
Lisa Arafune, Forge Policy Solutions
Dana Brunson, Internet2
Thomas Hacker, Purdue University
Preston Smith, Purdue University

NSF Award OAC-2036534

Facilitated by KNOWINNOVATION

FORGE POLICY SOLUTIONS

Building the research innovation workforce: a workshop to identify new insights and directions to advance the research computing community.

NSF Award OAC-2036534

Workshop Co-Chairs
Lisa Arafune, Forge Policy Solutions
Dana Brunson, Internet2
Thomas Hacker, Purdue University
Preston Smith, Purdue University

Executive Summary

A virtual workshop was conducted in August and September 2020 across three sessions with 100 unique attendees that focused on problems, challenges, and potential solutions to issues affecting the cyberinfrastructure/research computing and data (CI/RCD) workforce that develops software and operates and supports computing facilities for research. The workshop was conducted with the aid of KnowInnovation, an experienced workshop facilitator. This report provides a summary of the virtual workshop, and includes an overview of the workshop, a summary of the attendee selection and workshop processes, a summary of the keynote presentations, and an overview of the challenges and recommendations identified by workshop participants.

To summarize, workshop participants identified 12 thematic challenges that affect the research computing innovation workforce. These challenges included problem areas involving: diversity and inclusivity; fostering the development and support of the workforce ecosystem and talent pipeline; establishing viable career paths and normative role descriptions in the workforce; enhancing internal and external communications and education for stakeholders; compensation; workforce sustainability; the establishment of a identity of the field as a discipline; the position of research computing within institutional organizations; and the need for continuing training and education for professionals. The composition of the CI/RCD workforce is broad, and it is difficult to precisely define the boundaries of the workforce. In the broadest sense, the CI/RCD workforce includes highly skilled and experienced individuals who create, support, and facilitate access to digital information, computing infrastructure, software, and communities of practice that function in service to the broad research communities who rely on advanced research computing to advance discovery in their respective fields.

To address these challenges, workshop participants generated 12 recommendations. Upon analysis and reflection on these recommendations, the workshop co-chairs found that the recommendations could be summarized and grouped by the communities or organizations to whom these recommendations could be addressed. These communities are:

1. The cyberinfrastructure and research computing and data (CI/RCD) community itself;
2. Higher education institutions and other research organizations; and
3. The National Science Foundation and other government science agencies.

Notable recommendations that potentially could be easily implemented in a short timeframe include: developing an "umbrella" professional organization to span across existing community groups; creating a common set of job descriptions and career paths for the cyberinfrastructure/research computing workforce across institutions; conducting a national census within the research computing community to illuminate the current state of the national workforce; and encouraging the development of institutional governance bodies for research computing. The workshop discussions did not identify specific institutions (beyond the NSF) to whom these recommendations would be addressed. It is assumed that organizations or groups within those organizations (such as Research Computing or the Human Resources group) will consider the recommendations within this report, and adopt and adapt recommendations to address problems that groups recognize within their own institution.

Workshop Report

Introduction

Modern 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, Industrial, and national boundaries. Software must be written, adapted for use, or used as an integral component of research projects. Large, complex computing systems must be designed, built, and operated; all while effectively communicating the value of these investments to stakeholders. Developing high quality software, building and operating cyberinfrastructure systems, and aligning with multiple stakeholders, all of which the veracity of research results relies upon, requires a highly skilled and dedicated **workforce in Cyberinfrastructure and Research Computing and Data (CI/RCD)**.

A report from the NSF-sponsored "Professionalization in Cyberinfrastructure" workshop in 2017 (Berente, 2017) describes examples of these workforce roles as "*...'cyberinfrastructure engineer', 'research software engineer', 'research technology facilitator', 'research applications support', and 'data application specialist'.* "[1]

The Campus Research Computing Consortium (CaRCC) provides a definition of the area of Research Computing and Data (CaRCC, 2020), within which the workforce labors (the following quote is from the CaRCC webpage at http://carcc.org/about) :

> "**Research computing and data**" involves people, scholarship, and resources supporting the needs of researchers and research leveraging compute, data, networking, and software, broadly defined, including the professionals who execute and support these efforts. Whereas entities supporting research computing and data historically emerged from operating and supporting high performance computing, the needs, capabilities and technologies have sufficiently broadened the scope of research information technology to include virtualization, support for the cloud, containers, middleware, workflows, data management, data movement, compliance and security, user training, support of instruction using advanced research

[1] http://managingcenters.net/wp-content/uploads/2018/05/Berente-et-al-2017-professionalization.pdf

computing and data, on-boarding into new technologies, and deep engagement ("facilitation") to help guide researchers."

Overview of the Workshop

The workshop sought to bring together the community, composed of affinity groups, that included research computing leaders, practitioners, industry, academic domain scientists, academic computing scientists, and attendees from federal agencies. The overall aim of the workshop was to explore the issues and to devise potential solutions to the problems affecting the research computing innovation workforce that is essential to research computing and cyberinfrastructure. The goals of the workshop were to: 1) provide a set of recommendations to the National Science Foundation regarding actions to address the shortfall in the availability of skilled research computing practitioners in the research innovation workforce; and 2) 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.

Attendee Selection Process

The workshop attendee selection process was built on identifying and inviting attendees who represented *affinity groups* that included academic (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.

Workshop co-chairs and steering committee members compiled a list of 223 potential attendees from the international community to invite. E-mailed invitations included a link to a Google form that requested self-identification of the individual's affinity group. The following set of affinity groups were provided on the form and attendees were asked to select one: Industry, Government Agency, Cyberinfrastructure Practitioners and Professionals, University and non-profit Research Computing Leadership, Academic computer science/cyberinfrastructure research, Academic Domain science (e.g. physical sciences, life sciences, social sciences, engineering, etc.), and Other. This 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 141 invitees completed the form and responded to the question "Will you attend the workshop?" with 120 "Yes, I'll be there", and 21 responded with "Sorry, can't make it". Of the 120 who indicated they would attend:

- 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 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.

Workshop Process

All three virtual workshop sessions were conducted with the aid of KnowInnovation, an experienced workshop facilitator, on (all times in eastern time) Wednesday, August 19, 2020 (from 10am to 6pm) with 96 attendees; Tuesday, August 25, 2020 (from noon to 4pm) with 77 attendees; and Wednesday, September 9, 2020 (from noon to 2pm) with 77 attendees. [2]

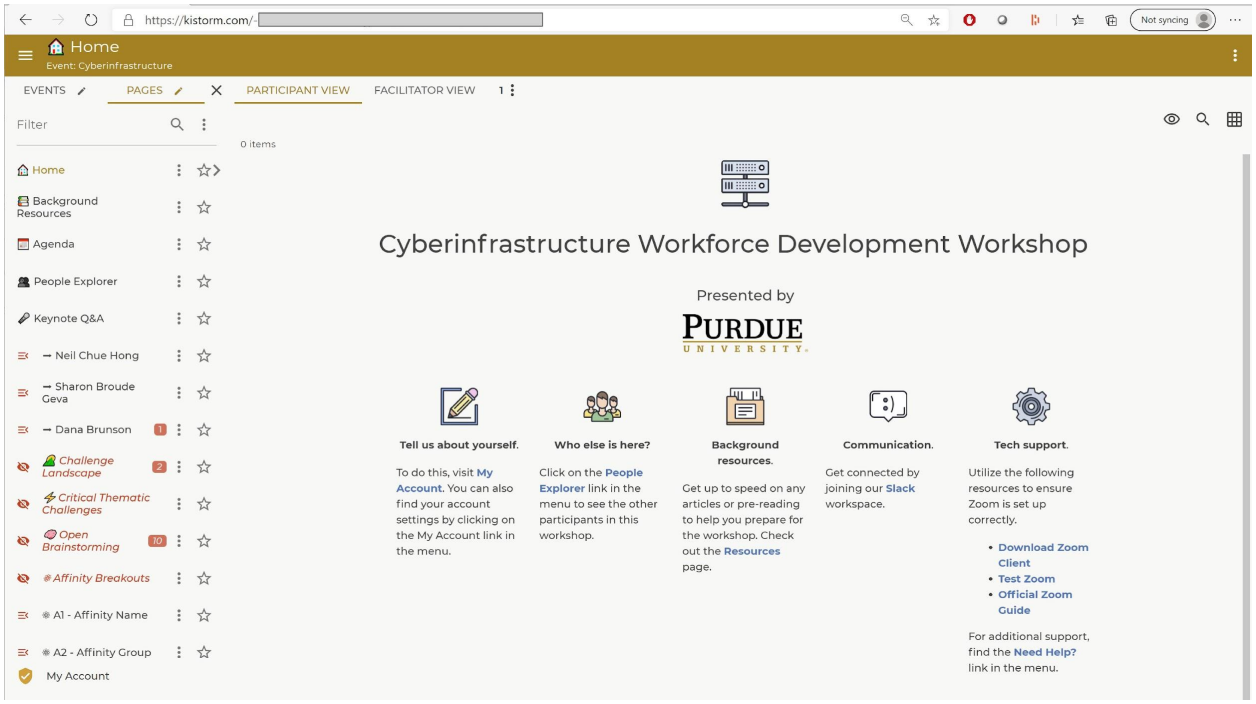

Figure 1. Virtual Workshop webpage within KIstorm platform.

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

[2] The number of attendees was determined using Zoom logs to count the number of people who joined the Zoom session at least once for the workshop (we did not take into account the length of time the person was in the Zoom session), and included NSF observers but not KnowInnovation workshop facilitators.

agenda for each virtual workshop session[3]. The sessions are described in detail in later sections of this report. 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.

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, founding Director and Principal Investigator of the Software Sustainability Institute and the University of Edinburgh. Hong spoke on "Sustaining a CI Workforce - Skills, Careers, and Diversity". 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.

Hong reported that the RSE movement was growing quickly - from just a workshop in 2012 to a full professional society in 2020. This professionalization has seen marked growth in RSE job postings, steady increases in workshops teaching the skills necessary to be an RSE, and even citing the recognition in public policy of the need for skilled software and data workforce.

The RSE movement still has room to improve, however - RSEs require career paths, and moving from industry to research is still difficult. A pipeline for junior RSEs is still required, and funding to support those RSEs remains a challenge. In the UK, diversity is a challenge - UK RSEs are largely from physics and computer science, and only 14% female and 5% non-white.

The presentation concluded with Hong's description of the CI workforce as an ecosystem, where the various roles and facings overlap. This notion of the "workforce as an ecosystem" (shown in Figure 2 - figure reused with permission from OECD from a report from a project workshop unrelated to our workshop (OECD, 2020)) continued to re-emerge as a theme throughout the workshop[4].

[3] The agenda for the workshop sessions are included as an appendix at the end of this report.
[4] Neil Chue Hong's presentation is available at https://doi.org/10.6084/m9.figshare.12827180

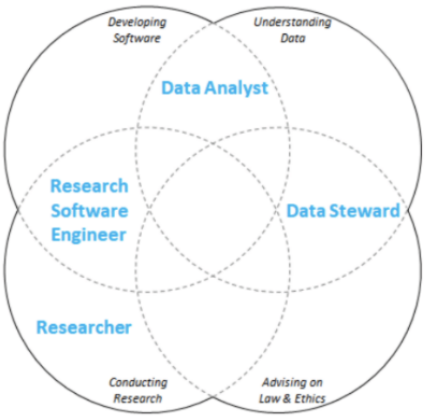

Figure 2. Venn diagram of roles and responsibilities. *Source: adapted by authors from an original diagram by Simon Hettrick (SSI) that was developed at the project workshop* (OECD, 2020). Used with permission from OECD.

The second keynote was from Sharon Broude Geva, Director of Advanced Research Computing (ARC) in the Office of Research (UMOR) at the University of Michigan and chair of the Coalition for Academic Scientific Computation (CASC) spoke on "Workforce Development : A Broad Perspective of (and on) Community Needs"[5]. Dr. Geva began by giving an overview of the CASC organization, then described what the 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.

On center leadership, Dr. Geva described several paths to become a center leader - from an RCD practitioner, as an IT manager changing streams, or from a faculty member. Dr. Geva made the case that center leadership should have a graduate degree, potentially a PhD - a statement which spurred lively discussion throughout the workshop.

Finally, the category of thought leaders was laid out - describing the function (often performed by center leadership) where academic pursuits around cyberinfrastructure are conducted. Challenges here are convincing universities that this activity is important and academic, and the need to treat CI thought leaders as scholars in their own right.

Finally, Dana Brunson, 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"[6]. Dr. Brunson presented an overview of existing communities, previous efforts, and ongoing work in the community. The presentation first discussed the Campus

[5] Sharon Broude Geva's presentation is available at https://www.rcac.purdue.edu/ciworkforce2020/schedule/geva.pdf
[6] Dana Brunson's presentation is available at https://www.rcac.purdue.edu/ciworkforce2020/schedule/brunson.pdf

Champions program, a community of practice sponsored by the XSEDE project. There are over 700 Campus Champions from over 300 institutions.

Following this, Dr. Brunson presented a summary of previous work on the professionalization of cyberinfrastructure work, describing previous workshops, the notions of "facings" (researcher-facing, systems-facing, data-facing, software-facing, and strategy/policy-facing) that came out of those workshops, and the definition of "Facilitator" - 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.

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. 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 [7].

The next section provides a summary of these thematic challenges.

Summary of Thematic Challenges Identified

Thematic Challenge 1. How do we recruit and sustain a diverse and inclusive workforce?

Today, the CI workforce has several characteristics that inhibit the development of a diverse and inclusive workforce. These characteristics include: a homogeneous workforce that is not representative of the overall society or academic population; a rigid clique-like culture with predefined expectations that is difficult to break into; drawing recruits from a narrow set of majors and schools; a lack of openness to perspective from outside computer science and closely-adjacent fields; gaps in training for professional development; and a lack of role models. The impact of these characteristics on the diversity of the workforce is increased attrition and loss of knowledge, decreased innovation, fostering insular and rigid thinking that limits the ability to adapt to changing circumstances, increased difficulty in hiring, and loss of new talent and new perspectives. The root causes of this challenge include: lack of acknowledgement of the importance of diversity; candidate searches limited in scope and locale to where existing group members are already present; lack of diversity in leadership; lack of funding and training for disseminating and promoting the adoption of inclusive practices; a small hiring pool; a lack of partnerships with minority serving schools; and a male oriented culture in Computer Science and Software Engineering. Successful progress on this problem would be reflected in workforce diversity that is similar to the composition of the research community; increased retention; and increased awareness of diversity and inclusion in hiring practices.

[7] We note that no breakout discussion groups formed about one of the thematic challenges (number 12 below), which we include for completeness. This challenge is closely related to thematic challenges 10 ("Where the RC [Research Computing] function sits in the institution has a major effect in many of the issues listed as thematic challenges."), and 4 ("How do we communicate & sell value to various stakeholders? (stakeholder other than people occupying the cyberinfrastructure workforce - INSTITUTIONAL stakeholders)").

Thematic Challenge 2. How do we create a workforce pipeline?

The discussion around this thematic challenge recast the workforce problem as a “workforce ecosystem” in place of a “pipeline” to reflect the reality that people enter and move throughout the profession at different points, and that an “ecosystem” represents an environment that sustains growth and nourishment of the workforce. The workforce ecosystem currently lacks training, effective knowledge sharing among institutions and projects, clarity about the roles available within the workforce, clearly defined entry points for new people, and has ill-defined opportunities for advancement. Overall, what is needed is well defined entry points, exit points, and required qualifications and expectations for different roles within the ecosystem. The impact of this challenge is several-fold: confusion about roles within CI organizations; poor job mobility between institutions; confusing or missing compensation structures; poor availability of highly qualified candidates for new roles. It also results in a less creative workforce that does not draw talent from greater diversity; an inadequate number of qualified people available for roles; poor visibility of the profession to leadership and the public; difficulties hiring and retaining skilled workers; and limited professional recognition. The root causes discussed included: need for clearly defined career paths; unclear value proposition for career paths; the relative newness of the field; a poor awareness of the critical need for skilled workers; and the lack of clear communication about the field to the general public. Successful progress in this challenge area would be indicated by a clearly defined profession that was widely understood with a vibrant and diverse workforce that would be supported by institutional leadership and the general public.

Thematic Challenge 3. How do we publicize, establish, and get recognition for CI careers and roles, both existing and as a viable career path? (focusing on people who might take on careers in cyberinfrastructure)

The recognition of CI careers and roles is inhibited by the lack of: clearly defined CI professional roles and career paths; understanding of motivational reward structures; role models/mentors; and venues in which to develop a group professional identity and from which to recruit talent. The impact of this is poor rate of entry into the profession, loss of people to industry, poor diversity, invisible career paths that impedes recruitment; lack of investments by institutions in training for long-term careers; non-competitive compensation compared with existing IT institutional roles; and a general lack of awareness of the field. The root causes of these problems include: roles of professionals change over time; lack of shared understanding across institutions on the identity of job roles; lack of co-authorship on papers; lack of sustained core funding for the positions, which results in a reliance on unstable soft funds for the positions; lack of understanding of the role of research computing in academic research as a team effort; and lack of university leadership awareness of the role and importance of research computing and CI professionals. Successful progress on this challenge would be reflected in increased recruiting and retention, clear pathways for career development, existence of a professional society recognized by ACM and/or IEEE, and the recognition of research computing and the profession by institutional leadership.

Thematic Challenge 4. How do we communicate & sell value to various stakeholders? (stakeholders other than people occupying the cyberinfrastructure workforce - INSTITUTIONAL stakeholders)

The challenge of communicating the value of research computing to stakeholders outside of the research computing workforce has several inherent characteristics. The lack of clear and well defined

understanding of who are the stakeholders, need for a clear career path for CI/RCD professionals difficulties in communicating the role and value of staff and their technical capabilities, lack of understanding about the role of research computing, difficulties communicating success and the inherent difficulty of the work, absence of research computing professionals from advisory boards, and an overall lack of clear definitions and understanding of the entire field of research computing and CI contribute to several negative impacts. These impacts are: CI is not viewed as essential to science; high turnover of CI professionals; being viewed as a cost center instead of a critical strategic investment with sustainable funding; and a lack of a coordinated institutional approach to research computing. The root causes for this include: the organizational position of research computing under the CIO/IT organization, which is viewed as less research oriented; no well defined common methodology for measuring the impact of effective CI in the research process; poor support from institutional leadership; and the lack of recognition of the roles and accomplishments of individuals in the workshop and of research computing generally within the institution. Successful progress on this challenge would be reflected by the inclusion of cyberinfrastructure in institutional planning and funding, recognition of cyberinfrastructure by institutional leadership as an essential element similar to teaching and general IT, and the availability of common reported metrics reflecting research computing efforts and impacts.

Thematic Challenge 5. How to capture & communicate the state of the Ecosystem: successes, challenges, gaps?

The challenge of capturing and communicating the state of the research computing and cyberinfrastructure ecosystem in the context of successes, challenges, and gaps has several salient characteristics. The overall challenge of communicating a clear message is influenced by: the lack of a comprehensive definition of a research computing and cyberinfrastructure ecosystem and where the roles of cyberinfrastructure professionals fit in this ecosystem; poor understanding of the elements of the ecosystem among stakeholders and how these elements align and work together; instability in funding for these elements over time; the lack of a “unified vehicle” or single outlet for shaping and disseminating successes to the broader community and to the public; and a lack of clarity and fragmentation about what is important: domain science successes, technology advancements, or big new machines - clear metrics of success are missing. The impacts of this challenge are: disparities in capabilities available for the research enterprise when comparing institutions that integrate research computing into their strategy and planning versus institutions that lack an integrative view; difficulties in communicating the value and impact of research computing to stakeholders, which creates a barrier to entry for new people from outside the community; and an overall lack of cohesive direction. The root causes of this include several elements. The very broad set of skills present in the research computing ecosystem that is one of its strengths is also a weakness, in that it impairs the development of clear descriptions of professional roles within the ecosystem. Other root causes include: poor communication overall among researchers and CI professionals, and with groups outside the institution; lack of institutional funding for this communication; the high rate of change of technologies and user communities; political factors (such as competition) that inhibits the development of common definitions and principles; lack of national leadership toward a clear definition and strategic plan; and cyclical funding availability fostering “churn” within the ecosystem. Successful progress on this challenge would be indicated by the development of a clear map of institutions and stakeholders that reflects the current ecosystem state and future plans; a clear set of deliberate strategies within the overall ecosystem; more clear and prominent communications; and a coherent singular research computing and cyberinfrastructure roadmap for the ecosystem.

Thematic Challenge 6. How do we adequately compensate CI roles (both money and other compensation/experiences) to make them attractive?

The challenge of adequately compensating roles within the research computing and cyberinfrastructure ecosystem, in terms of money and other types of compensation and experiences, is a complex challenge. The characteristics of this challenge include: lack of information about what institutions are paying for similar roles; the lack of clear understanding of attractive reward mechanisms for individuals at different career stages within the workforce; lack of clear career paths with well defined stages and compensation models for each stage; reasonable job security; the need for training and professional development; the lack of freedom to innovate and to evolve roles as the science changes; compensation packages that are not competitive compared with industry. All of these lead to impacts that include: poor retention and recruitment with job openings that are difficult to fill; loss of talent to industry or to peer competitor institutions; organizational identity struggles that seek to juggle mission and securing funding sources; and poor levels of recognition for workforce professionals. The root causes of this challenge include: lack of flexible working arrangements; poor role definitions and irregular and unclear job titles; the organizational location of the research computing unit when it falls under the CIO with no reporting to the VP of Research office; lack of consistent recurring hard funding; poor understanding of the role of research computing within the institution, which can lead to difficulties in the HR classification of roles; and the perception of research computing and cyberinfrastructure being a cost center instead of a strategic institutional investment. Successful progress on this challenge would be indicated by increased opportunities for professionals to move laterally across or up to a higher level between institutions; better compensation; and the ability for professionals to move between industry and academic careers.

Thematic Challenge 7. How do we promote and ensure sustainability (funding / growth / retraining / size / non-volunteer) for the CI workforce?

The challenge of promoting and ensuring sustainability in terms of funding, growth, retraining, and size of the research computing and cyberinfrastructure workforce has several characteristics. These characteristics include: turnover and attrition of skilled staff based on the changing availability of funding; limited career advancement in terms of leadership and projects; lack of investment in research infrastructure to keep up with growth in funding activity for research projects; and the lack of attention to the development of cyberinfrastructure and research computing as a scientific instrument. The impacts of this challenge include: reliance on the need to continually search for and secure grant funding sources for funding positions; division of institutions into well funded vs. poorly funded when institutions lack hard funds to provide a stable base; lack of continuity of projects (stale/abandoned); limits to progress in research due to limitations in research computing staff; difficulty in retention of highly skilled staff; inefficiencies in the workforce ecosystem; and poor strategic planning between research groups and research computing staff. The root causes of this challenge include: poorly defined career paths; attrition due to lack of diversity in soft funding sources for staff; misalignment of the research computing organization within the institution; lack of strategic planning and commitment; pace of change within institutions is too slow; and lack of understanding of stakeholders of the role of research computing. Successful progress on this challenge would be indicated by: increased retention of research computing professionals; inclusion of research computing in institutional policy and planning; clear institutional mechanisms to include research computing professionals in research grants; better alignment of the research computing organization within the institution; and acknowledgement of cyberinfrastructure as a scientific instrument.

Thematic Challenge 8. Develop CI as a discipline that is independent of Computer Science

The challenge of understanding and differentiating cyberinfrastructure as a discipline independent of computer science has several inherent characteristics. It is not clear that this is a major challenge, as many in the field of computer science do not view cyberinfrastructure as a part of the discipline of computer science. Characteristics of the challenge include: poor coding expertise in the graduate student research labor force; low level of awareness of computing science fundamentals in the domain science community; lack of understanding or appreciation within the science community for software engineering; a disconnect between the needs for educational programs to produce skilled research computing staff and computer science programs that can provide the necessary skills; lack of acknowledgement of cyberinfrastructure as an academic discipline; the view of cyberinfrastructure as a support service; difficulty in defining the roles and identity of cyberinfrastructure professionals; and the duality of efforts in research computing in both providing services (e.g. computing infrastructure) and the need to advance the state of the art in the field as an academic discipline. The impacts of this challenge include: lack of awareness of cyberinfrastructure roles from people outside of computer science, which impacts the size of the research computing workforce; difficult recruiting people from computer science into research computing; limited technical depth of students entering the research computing field; lack of cross-training in research methods for application to discovery in cyberinfrastructure; and the lack of a well established career pipeline for cyberinfrastructure. The root causes of this include: poor understanding of cyberinfrastructure outside the community; lack of requirements for programming in domain science courses; fracturing and siloing of expertise across different science domains; different scholarship expectations for computer science vs. cyberinfrastructure; lack of an underlying theoretical framework for cyberinfrastructure to provide a foundation for a new discipline; lack of a curriculum for cyberinfrastructure; lack of professional certifications for roles in cyberinfrastructure; schism between faculty dominated computer science organizations and staff dominated research computing groups. Successful progress on this challenge would be indicated by: funded internships in research computing centers; a pipeline of positions for new graduates with a cyberinfrastructure oriented degree; a more clearly understood role for cyberinfrastructure as a field of science; and recognition of cyberinfrastructure by computer science departments.

Thematic Challenge 9. Faculty, program managers at federal agencies, and CIOs need to be educated about the role of CI professionals.

The challenge of educating external stakeholders (such as law makers and federal agencies) about the role of cyberinfrastructure professionals has several characteristics. These characteristics include: lack of a “home” with ownership in a federal agency (such as NSF) for cyberinfrastructure professional roles which often apply to research across many disciplines; lack of a common understanding of roles within cyberinfrastructure either within cyberinfrastructure or within academic disciplines; difficulties in securing funding that is specific for supporting cyberinfrastructure related positions and support; lack of recognition of the expertise required for cyberinfrastructure professionals; lack of career paths; emphasis on funding faster equipment instead of cyberinfrastructure professionals; and poor retention and career progression due to uncertainties in the availability of research funds. The impacts of this challenge include: misidentification of research computing professionals as “IT people” by institutional computing leadership; early-career struggles for cyberinfrastructure professionals due to uncertainties in the availability of research funding and

institutional commitment to funding cyberinfrastructure; retention of skilled professionals; underutilized cyberinfrastructure due to a lack of skilled professionals; and growing software “debt” - which refers to the need to expend effort to document and harden software for long term viability - over time as projects suffer from the lack of skilled software professionals. The root causes of this challenge include several factors: the lack of a clear value proposition for research computing targeted to the different stakeholders (leadership, faculty, etc.); lack of recognition for research computing staff; bias to funding hardware instead of people; the traditional assignment of graduate students to the professional roles in research computing; the need for faculty, CIOs, and program managers to be educated on how research computing is utilized and valued at institutions; and poor communications about research computing to outside communities. Successful progress on addressing this challenge would be indicated by: the presence of consistent job and role descriptions for research computing across institutions; stable institutional funding for research computing; effective communications; and recognition for research computing as an essential element of the institutional research enterprise.

Thematic Challenge 10. Where the Research Computing function sits in the institution has a major effect in many of the issues listed as thematic challenges.

A challenge facing research computing organizations within institutions relates to the position of the group within the overall institutional organization. The characteristics of this challenge include: placing the organization exclusively within the VP of Research or the CIO office is often not ideal; research computing often is placed in the office of the CIO, VP of Research, an academic unit, or a distinct center; leadership by tenured faculty often lacks administrative standing within the institution; research leadership may better understand the need for research computing; and there is no single “best” model - the unique history and structure of each institution and the people involved may lead to different organizational configurations across institutions. The root causes of this challenge include: unclear expectations about how roles fit within the institution;poor flexibility and ability to adapt to change in academic leadership structures; the context of central vs. distributed institutional funding sources; disagreement among institutional leadership about which group should “own” research computing; and differences in tasks and needs addressed by institutional IT vs. research computing. Successful progress on this challenge would be indicated by: the presence of research computing leadership as a part of institutional planning; recognition of cyberinfrastructure and research computing as a “core” research infrastructure element; and successful collaboration between the CIO and VP Research offices in describing and supporting the role of research computing within the institution.

Thematic Challenge 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?

The challenge of establishing training and development for research computing staff is a critical issue that affects the continuing development of the workforce. There are several characteristics of this challenge.

Although there are many CI professionals working within academia and national laboratories, there is no clear definition of career paths, necessary competencies and skill, and the job descriptions often do not reflect the changes that the CI field is undergoing over time. There are efforts, such as those

championed by campus research computing professionals, that aim to define the various roles that CI professionals play within the research ecosystem, focusing on the definitions of “researcher-facing, systems-facing, stakeholder-facing, and software/data-facing” professionals. However, there is no clear understanding or agreement across the broader CI community about the definition of these roles. Additionally, there is a lack of culture and mechanisms to train CI staff (both in technical and management areas) as they progress through their careers. In many cases, the careers are plateauing. Some leave their current positions to pursue more clearly defined careers.

Within the current environment there is also a lack of funding and of staff who can focus on providing training and experience for teaching training. Because of the lack of such understanding, there are no well-defined job titles, no progressive career paths that can be followed. As a result, it is difficult for CI professionals to change jobs within and across institutions and no structure to support the recognition of accomplishments and appropriate compensation. The overall impact is that opportunities for CI professionals to mature and to find multifaceted job satisfaction within and across organizations is limited. Moreover, there is a lack of connections between academic and cyberinfrastructure organizations. The root causes for this challenge include overworked staff, a tradition of groups working autonomously within a university rather than cooperating across the institution, and a lack of recognition of the role of the research computing professional as a partner and collaborator in a research project.

Thematic Challenge 12. How do we get resources and credibility for CI organizations and roles from funding agencies and university leadership?

The challenge was initially identified as a thematic challenge during the first workshop. However, no breakout groups formed about this thematic challenge. This challenge is closely related to thematic challenges 10 (“Where the RC [Research Computing] function sits in the institution has a major effect in many of the issues listed as thematic challenges.”), and 4 (“How do we communicate & sell value to various stakeholders? (stakeholder other than people occupying the cyberinfrastructure workforce - INSTITUTIONAL stakeholders)”).

Recommendations

The second virtual workshop session turned from problems to considering potential solutions to the problems and challenges identified during the first workshop session. Launching from the thematic challenges, the workshop broke into small affinity groups for a ‘deep-dive’ into the thematic challenges to describe the characteristics and impacts of these challenges, and also to seek definitions for what successful solutions to the challenge may look like.

Following a lunch break, an open brainstorming session gave an opportunity for individual reflection and the proposing of ideas. The workshop then transitioned into generating ideas for solutions to the thematic challenges using two charrette breakout small groups sessions: the first one for affinity groups; and the successive charrette consisting of mixed groups that included a range of affinities. This second virtual workshop session formed new mixed breakout groups that focused on synthesizing solutions (conveyed as recommendations) from the thematic challenges and ideas identified in the first session.

Upon analysis and reflection on these recommendations, the workshop co-chairs found that the recommendations could be summarized and grouped by the communities or organizations to whom these recommendations could be addressed. These communities are:

1. The cyberinfrastructure and research computing and data community (CI/RCD);
2. Higher education institutions and other research organizations; and
3. The National Science Foundation and other government agencies.

We prepared a brief presentation that summarizes the recommendations and grouped them by the communities or organizations to which we feel that the recommendations seemed to be addressed.

The material within this section was summarized in slide form in a presentation to the NSF CI Strategy Committee on October 20, 2020.

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

a. 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.
b. 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.
c. Create a common set of job descriptions and career paths for CI/RCD.
d. 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 existing roles and compensation models for CI/RCD professionals, and improve retention and mobility options for the workforce.
e. Develop a Diversity, Equity, and Inclusion action plan for the CI/RCD community.

Recommendations for higher education institutions and other research organizations.

a. 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.
b. 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.
c. Improve communication about the value of CI/RCD to institutional stakeholders and leadership.
d. Develop sustainable funding models for research cyberinfrastructure investments such as people, software, services, and resources on campus.
e. Align research computing within the organization to ensure appropriate involvement and governance from the CIO, VP of Research, Deans, and other constituent groups.

Recommendations for the NSF.

a. 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.
b. 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).
c. 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.

Detailed List of Recommendations

The next section of this report provides a detailed list and summary of the recommendations identified during the workshop. The recommendations are ordered by the number of votes from the highest to lowest number of votes. Each recommendation below is presented in the following format:

- **Thematic challenge area identified by attendees**
- *Recommendation summary*

The material from this section was also provided to the NSF as part of an interim workshop report in September, 2020.

Recommendation I.

Thematic challenge: How do we publicize, establish, and get recognition for CI careers and roles, both existing and as a viable career path? (focusing on people who might take on careers in cyberinfrastructure)

Recommendation Summary:
For cyberinfrastructure (CI) / research computing and data (RCD) to be a viable career choice, workshop participants recommend increasing the community awareness of CI/RCD work as a unique career path with an identity distinct from enterprise IT. A community organization should be funded to communicate and support existing work in HR job families for CI/RCD, and to take the next steps in developing materials such as defining career arcs, identifying good examples of internship and training programs, and disseminating these to the community. Participants recommend providing funds (perhaps from NSF) to institutions to bootstrap programs that take advantage of these community resources, with a commitment to contribute experience gained back to the community.

Recommendation II.

Thematic challenge: Faculty, program managers at Federal agencies, and CIOs/Institutional leadership need to be educated about the role of CI professionals. (Note there are three sub-themes - Federal Agencies, CIOs/VPRs/Center Leader, and Faculty)

Recommendation Summary:
The workshop identified a need to educate management about the role of HPC/CI/RCD professionals and to communicate the impactful value added by this group of people to research and researchers. Advocacy should include success stories, facts/figures, and demonstrated impact of HPC/CI/RCD. Also, a great desire was expressed by this community of professionals to increase opportunities to contribute to/publish research. Managers and science domain researchers tend to continue to view the research computing aspects as part of commodity IT or as a service rather than a partner. Ideally, managers should see CI staff as an investment in research to help increase research funding, and not as a cost center.

Ideas for Specific Solutions:

- The NSF should encourage grant applications to include RCD professionals in the proposal process, and a budget appropriate for the level of RCD professional participation in the grant. The NSF should ask for annual reports and strategic plans to include a section on RCD participation. The NSF should review the CaRCC HR standards document and consider its application to NSF grants. The NSF might also consider the development of a policy change

white paper.

- The academic community and teaching faculty should make a proactive effort to advocate for the important role RCD professionals can play in publishing, training and supporting graduate students, and in developing fundable proposals. Academic departments should include RCD professionals as a part of the faculty recruitment (e.g. in faculty campus visits) and retention plans, and be involved in the faculty onboarding process.
- Centers should include an effective communications strategy for faculty, administration, and funding agencies, and publicize tools, leading/best practices, and results. The strategy should include "Leading/Best Practice" methods and models for how research computing centers can be structured into sustainable central parts of institutions and create example funding models and shared use models that allow supporting long term RCD career paths within those organizations. Centers should be encouraged to have broad conversations within university administration (e.g., between VPR and CIO) to communicate how centers support research, by showing value of the investment to the university's output. Ultimately, the strategy should communicate that Centers are investments instead of cost-centers and are clearly tied to research results.
- Centers should engage with HR professionals to institute consistent cross-institutional job descriptions for RCD and should review the CaRCC HR Standards document.
- Centers should create sustainable funding models for investments (people, software, services (such as training and workforce development), and resources) on campus. Solid funding for RCD roles should be a part of standard and sustainable budgets.
- The RCD community should organize itself to enable more effective and consistent communication to promote the community. The community should also commit to funding community positions that can help realize these activities instead of relying on volunteers.

Recommendation III.

Thematic challenge: How to capture & communicate the state of the Ecosystem: successes, challenges, and gaps?

Recommendation Summary:

The workshop attendees identified the need to capture and communicate information related to the cyberinfrastructure (CI) "ecosystem" and including the CI workforce ecosystem (a term more accurate than 'pipeline') that reflects an up-to-date coherent state of the research computing field, opportunities for entering and progressing through roles and careers, and a common definition of roles and responsibilities within the ecosystem. To address this in the near term, a comprehensive effort is needed to: contact the community to determine what information is needed, and to collect and report these data; create a central information resource to disseminate ecosystem information; create a visionary "umbrella" organization to serve as a "community of communities" to coordinate initiatives; develop a shared understanding within the community of the roles within the ecosystem; and creating a cross-institutional advisory group. The goals of these efforts are to: generate actionable information; create a clear map of stakeholders describing the roles of organizations; clearer strategies and roadmaps across the ecosystem that can adapt with changing technology; and a coordinated communication strategy that produces publications with highlights as well as actionable quantitative metrics.

Recommendation IV.

Thematic challenge: How do we adequately compensate CI roles (both financially and through other compensation/experiences) to make them attractive?

Recommendation Summary:

The recruitment and retention of a diverse and skilled cyberinfrastructure (CI) workforce relies on

incentives and adequate competitive compensation. To address this, organizations need to create a compensation, career path, and promotion model that accommodates the dual research/service roles present in the workforce ecosystem. To kick off this effort, a national census of positions, pay grade, cost of living differences, and benefits should be taken of the research computing and cyberinfrastructure workforce, with comparative information included between roles in industry and non-profit organizations. NSF could provide seed funding to incentivize the community to undertake this census, and also consider including requirements in solicitations that would encourage proposals to include direct or indirect budgetary items for cyberinfrastructure professionals. The goals of this effort are to: encourage the adoption of the practice of including support for CI professions as direct or indirect costs in funded proposals; encourage the production of an annual national cyberinfrastructure workforce survey (for the national industrial and non-profit workforce) that can be used to provide benchmarks for the profession; and help to increase retention and mobility options for the workforce.

Recommendation V.

Thematic Challenge:How do we recruit and sustain a diverse and inclusive workforce? How do we create a workforce pipeline?

Recommendation Summary:
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). However, framing this as a systems problem and identifying the key forces within that system will allow us to reason more effectively about it than our current model which focuses on a loose collection of single efforts and metrics. 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. The plan should be shared for community input through extensive community engagement to get buy-in and further develop incentives and metrics. A DEI plan and metrics would be included as a requirement similar to a data management plan by agency CI funding programs.

Recommendation VI.

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

Recommendation Summary:
Foster and fund a visionary umbrella organization to serve as a “Community of Communities” that nurtures, facilitates, and helps coordinate self-organized initiatives through their respective lifecycles. This requires decadal-scale planning/resources and national commitment to create and support such a core enabling organization. The effort can also benefit from drawing upon the experiences and models developed by the UK Society for Research Software Engineers. The organization should understand the community of community needs, advocate for its community members and develop programs to grow and enhance the CI workforce.

Recommendation VII.

Thematic challenge: How do we create a workforce pipeline?

Recommendation Summary:
Establish a formalized professional organization (e.g. Campus Research Computing Consortium (CaRCC) or within EDUCAUSE) that encompasses formalizing the profession, training, etc. to improve the workforce ecosystem. The workshop identified a need for a national organization that represents and advocates for RCD professionals that are distributed across campuses and research units. This

organization could take on issues that are vital to the RCD professionals' community, advocate for them to the funding agencies, campus administration, and other bodies. This organization could also provide direct benefits to their members including training, fellowships, internships, job description standardization, job advertising board, and many others. This organization can create a curriculum that can be adopted across institutions, hold annual meetings, awards, and produce publications. The language should also reflect the numerous pathways to entry, by changing 'pipeline' to 'workforce ecosystem.'

Recommendation VIII.

Thematic challenge: How do we communicate and sell value to various institutional stakeholders and where the RC [research computing] function sits in the institution has a major effect in many of the issues listed as thematic challenges.

Recommendation Summary:

Higher education institutions need to recognize and prioritize their research computing and data (RCD) capabilities, and place them organizationally somewhere appropriate for such a core set of capabilities for the institution, with funding at a level to support the institution's goals. Appropriate governance for RCD should be in place, connecting (for example) the CIO, VP for Research, Deans, libraries, Provost, CFO, or research centers with faculty stakeholders, and aligning the RCD function with institutional priorities. Where this function resides will vary between institutions, but it must exist, and needs to align at the strategic level. Successful funding models range from institutionally funded to researcher supported to anything in between. Institutionally funded or subsidized models tend to provide stability especially needed when starting an CI/RCD initiative at a campus, while faculty funded models tend to scale out more easily once critical mass is reached.

As with the CC* program's requirement for campus CI plans, the NSF could encourage the establishment of institutional governance bodies for networking, computing, and data in funding solicitations. NSF may consider encouraging the development of a national annual meeting (similar to EDUCAUSE) focused on research computing and cyberinfrastructure.

Recommendation IX.

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

Recommendation Summary:

This challenge focuses on the need for clearly defined career paths for experts and academic professionals that includes institutional mobility (addressing staffing scarcities) and funding stability for cyberinfrastructure positions. To address this challenge, national scale funding (e.g. NSF) similar to federal initiatives such as the BRAIN initiative are needed to establish seed funding to jumpstart a community wide effort. As a component of this initiative, higher education institutions should be encouraged to create multidisciplinary academic departments that include the involvement of institutional information and research organizations (e.g. CIO, Library, VPR). Stable institutional career paths (providing paths to tenure similar to faculty or librarians) are needed. An effort to identify the root causes of this challenge to investigate policy changes may be beneficial. The goals of this recommended effort are to: increase in the number of skilled personnel; encourage excellence in the scholarship and practice of cyberinfrastructure; and foster the creation of a common organizational structure and job descriptions across universities and research laboratories.

Recommendation X.

Thematic challenge: Develop CI as a discipline separate from Computer Science

Recommendation Summary:

The workshop identified the need to respond to the arrival of Cyberinfrastructure (CI) as a recognized field by developing CI as a discipline that is independent of Computer Science. To address this: multi-disciplinary (e.,g. domain sciences, social sciences, engineering, and library/information science) groups within educational institutions, in partnership with industry and professional societies, should begin an iterative process to develop a common core curriculum for Cyberinfrastructure that could be adopted for developing degree, certificate, badging, and specialization programs; implement badging and certification to current training programs; and expand virtual residencies across institutional locations across multiple roles. The goal of this effort is to establish a sustainable academic program and research ecosystem to advance the field, provide sufficient education/training opportunities sufficient to meet demand; and develop the workforce needed for current and future needs to aid discovery across all disciplines that rely on research computing and information sciences. There was also a need identified for informal education programs focused on certifications for CI professionals.

Recommendation XI

Thematic challenge: Cross cutting issues

Recommendation Summary:
The workshop identified several cross-cutting challenges, which included: diversity, equity, and inclusion; need for establishment of professional networks; smaller institutions lacking resources necessary for participation; and the need for a central resource for disseminating information about the research cyberinfrastructure ecosystem. To address these challenges, existing regional networks (e.g. Rocky Mountain Advanced Computing Consortium, Big10 Alliance, Great Plains Network, etc.), consortia (e.g. Internet2, CASC, CaRCC, Campus Champions, etc.), institutions (including R1/R2, HSIs, MSIs, HBCUs, and TCUs), and industry along with federal and state government agencies should establish a coordination program. The combined effort should seek to establish exchanges of people, information, expertise, and best practices based on needs identified by the community. The goals of this effort are to: establish career paths in research computing spanning regional institutions; foster the creation of new knowledge, increase the number of CI professionals, increase engagement with non-traditional users of cyberinfrastructure, and encourage the development of educational tracks aligned with activities related to cyberinfrastructure.

Recommendation XII.

Thematic challenge: How do we establish training & professional development for intermediate-level or staff transitioning careers?

Recommendation Summary:

The workshop considered the problem determining what is needed to support the continuing development of people in CI careers. The approaches recommended to address this are to: include ongoing skill development as a key job activity in job descriptions; ensure that NSF or NIH solicitations include requirements for plans for continued staff skill development; encourage the creation of a central information source to disseminate training and development opportunities specifically for cyberinfrastructure personnel; support the creation of intermediate training and workforce development events at conferences (such as SC) with high attendance from the CI community; and explore the development of apprenticeship or rotation programs to foster knowledge and skill sharing as well as to develop cohort communities across institutions. The goals of this effort are to: increase the number of trained people; build a deep pool of skilled candidates for mid- and upper-level CI positions; and create a culture of continual improvement and training across the CI workforce.

Final Workshop Session - Action Groups

The final workshop session turned to discussion of what actions could be taken in response to the challenges and recommendations identified in the prior two virtual workshop sessions. Seven action groups self-organized in response to recommendations:

- Community of Communities;
- Core Curriculum Development;
- Career Arcs;
- Professionalization Article;
- CI Job Description and Census
- Diversity, Equity, and Inclusion; and
- Regional Cross Cutting Groups.

The Community of Communities action group identified that the purpose of their group was to address the needs for sustainable facilitation of the professional community, education and training, and community and advocacy. Their stated objective is to create a blueprint for establishing a “community of communities”. The first action group discussion was attended by approximately 10 workshop attendees.

The second action group, **Core Curriculum Development**, sought to address the need for formal and information education for professionals across the many roles and “facings” within the cyberinfrastructure research computing community. The discussion for this group was attended by approximately 13 attendees. The needed activities identified by the action group were: defining a framework for curriculum; exploring the use of “badges” to recognize skills and training; and to define core competencies necessary for a successful graduate.

The third action group, **Career Arcs**, attended by approximately five participants, sought to understand the career paths taken by successful cyberinfrastructure professionals as part of the process of defining and communicating to potential new professionals and institutions the potential career paths for research computing professionals.

The fourth action group, **Professionalization Article**, with approximately 13 participants, discussed the need to define the profession of cyberinfrastructure and research computing as a 21st century entity that recognizes the existence of a workforce “ecosystem” instead of a “pipeline”.

The fifth action group, **CI Job Description and Census**, with approximately four participants, explored the steps and some of the challenges in establishing an annual Cyberinfrastructure Census.

The sixth action group, **Diversity, Equity, and Inclusion** (with approximately six participants), discussed the need to create an effort with experts in DEI to generate a report with recommendations centered on DEI to help improve recruitment and retention for a diverse workforce.

The final (seventh) action group, **Regional Cross Cutting Groups**, discussed the need to create regional coordination groups. There were approximately three participants in this group.

Observations and Lessons Learned from the Workshop

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 need for recognition and a clear definition of the different roles (or “facings”) reflecting duties in the workforce;
- The need for 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 creation of an “umbrella” organization spanning existing institutional, regional, and national organizations to advocate and coordinate actions and knowledge across the entire spectrum of roles, communities of practice, and institutions;
- The need for a concerted effort to address diversity, equity, and inclusion as a systems problem;
- The need to formalize education and training to create a coherent body of transferable knowledge for the national community; and
- The need for an effective communication strategy to increase understanding, awareness, and acknowledgement of the essential role of cyberinfrastructure and research computing as a core element within the institutional research enterprise.

REFERENCES

OECD (2020), "Building digital workforce capacity and skills for data-intensive science", *OECD Science, Technology and Industry Policy Papers*, No. 90, OECD Publishing, Paris, https://doi.org/10.1787/e08aa3bb-en.

Campus Research Computing Consortium (CaRCC) (2020). What is CaRCC? http://carcc.org/about. Accessed December 2020.

Berente, N., Howison, J., Cutcher-Gershenfeld, J., King, J.L., Barley, S.R. and Towns, J., “Professionalization in Cyberinfrastructure” (February 15, 2017). Available at SSRN: https://ssrn.com/abstract=3138592

ACKNOWLEDGEMENT

This workshop was supported by the National Science Foundation Award OAC-2036534. Any opinions, recommendations, findings, or conclusions expressed are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Appendix A: List of Participants

Workshop Co-Chairs

Lisa Arafune, Forge Policy Solutions
Dana Brunson, Internet2
Thomas Hacker, Purdue University
Preston Smith, Purdue University

Steering Committee

Linda Akli, Southeastern Universities Research Association (SURA)
Thomas Cheatham, University of Utah
Ewa Deelman, University of Southern California, Information Sciences Institute (ISI)
Dan Fay, Microsoft Research
Tsengdar Lee, NASA
Lauren Michael, University of Wisconsin - Madison
Frank Wurthwein, University of California, San Diego

All Participants

Name	Institution/Organization
Adrian Feiguin	Northeastern University
AJ Lauer	University Corporation for Atmospheric Research (UCAR)
Alan Blatecky	RTI International
Alan Sussman	NSF
Alex Younts	Purdue University
Amy Neeser	UC Berkeley
Anita Nikolich	UIUC
Ann Kovalchick	University of California Merced
Anna Klimaszewski-Patterson	California State University, Sacramento
Barr von Oehsen	Rutgers
Blake Joyce	University of Arizona
Bob Freeman	Harvard Business School
Bradley Tipp	Microsoft Corp
Brian Haymore	University of Utah

Bruce Lincoln	Silicon Harlem
Christina Koch	UW Madison
Christina Maimone	Northwestern University
Christine Harvey	MITRE
Claire Mizumoto	UC San Diego
Claudiu Farcas	Amazon Web Services (AWS)
Dan Fay	Microsoft Research
Dan Stanzione	TACC/UT-Austin
Dana Brunson	Internet2
Daniel S. Katz	UIUC
Danny Tang	University of Iowa
Deborah Dent	Jackson State University
Dhruva Chakravorty	Texas A&M University
Diana Dugas	New Mexico State University
Dingwen Tao	Washington State University
Dirk Colbry	Michigan State University
Douglas Jennewein	Arizona State University
Eli Dart	ESnet / LBNL
Ewa Deelman	University of Southern California, ISI
Fernanda Foertter	BioTeam
Frank Wuerthwein	UC San Diego
Gwen Jacobs	University of Hawaii
Hakizumwami Birali Runesha	University of Chicago
Henry Neeman	University of Oklahoma
Ilkay Altintas De Callafon	San Diego Supercomputer Center, UC San Diego
Jarek Nabrzyski	University of Notre Dame
Jay Alameda	NCSA
Jianjun Xu	Amazon Web Services
Jim Basney	NCSA

Joe Whitmeyer	NSF
Joel Gershenfeld	Brandeis University
John Blaas	University of Colorado Boulder
John King	University of Michigan
John McGee	University of North Carolina at Chapel Hill
John Towns	NCSA/UIUC
Julie Ma	MGHPCC
Karen Tomko	Ohio Supercomputing Center
Kelly Gaither	TACC/UT-Austin
Kevin Thompson	NSF
Laura Carriere	NASA Center for Climate Simulation
Laura Christopherson	RENCI, UNC
Lauren Michael	UW Madison
Lauren Rotman	ESnet
Linda Akli	SURA
Lisa Arafune	Forge Policy Solutions
Lorna Rivera	Georgia Institute of Technology
Marisa Brazil	Arizona State University
Marla Meehl	University Corporation for Atmospheric Research (UCAR)
Martin Halbert	NSF
Mats Rynge	University of Southern California
Maytal Dahan	Texas Advanced Computing Center
Neil Chue Hong	Software Sustainability Institute
Nicholas Berente	University of Notre Dame
Nicole Weibel	Purdue University
Nina McCurdy	NASA/ASRC Federal
Patrick Schmitz	Semper Cogito Consulting
Paul Morin	Polar Geospatial Center
Piyush Mehrotra	NASA Ames Research Center

Preston Smith	Purdue University
Ray Leto	TotalSim US
Rebecca Hartman-Baker	NERSC / Lawrence Berkeley Lab
Richard Knepper	Cornell University
Rick Friedman	Microsoft
Rose Robinson	Center for Minorities and People with Disabilities in IT (CMD-IT)
Ruth Marinshaw	Stanford University
Sandra Gesing	University of Notre Dame
Sarvani Chadalapaka	University of California, Merced
Scott Michael	Indiana University
Scott Yockel	Harvard
Scotty Strachan	University of Nevada, Reno
Shafaq Chaudhry	University of Central Florida
Sharon Broude Geva	University of Michigan
Shelley Knuth	University of Colorado Boulder
Stefan Robila	NSF
Stephen Harrell	Texas Advanced Computing Center
Terry Hogan	National Center for Women & Information Technology (NCWIT)
Thomas Cheatham	University of Utah
Thomas Hacker	Purdue University
Timothy Middelkoop	Internet2
Tracy Woods	Microsoft
Tsengdar Lee	NASA
Verónica G. Melesse Vergara	Oak Ridge National Laboratory
Von Welch	CACR, Indiana University
Wendy Whitcup	University of Southern California
William Thigpen	NASA
Wu He	NSF

Appendix B: Results of Post-Workshop Assessment

After the final workshop session, an invitation to an anonymous Qualtrics survey was sent to participants and conducted from November 12, 2020 to December 4, 2020. A total of 39 responses to the survey were recorded.

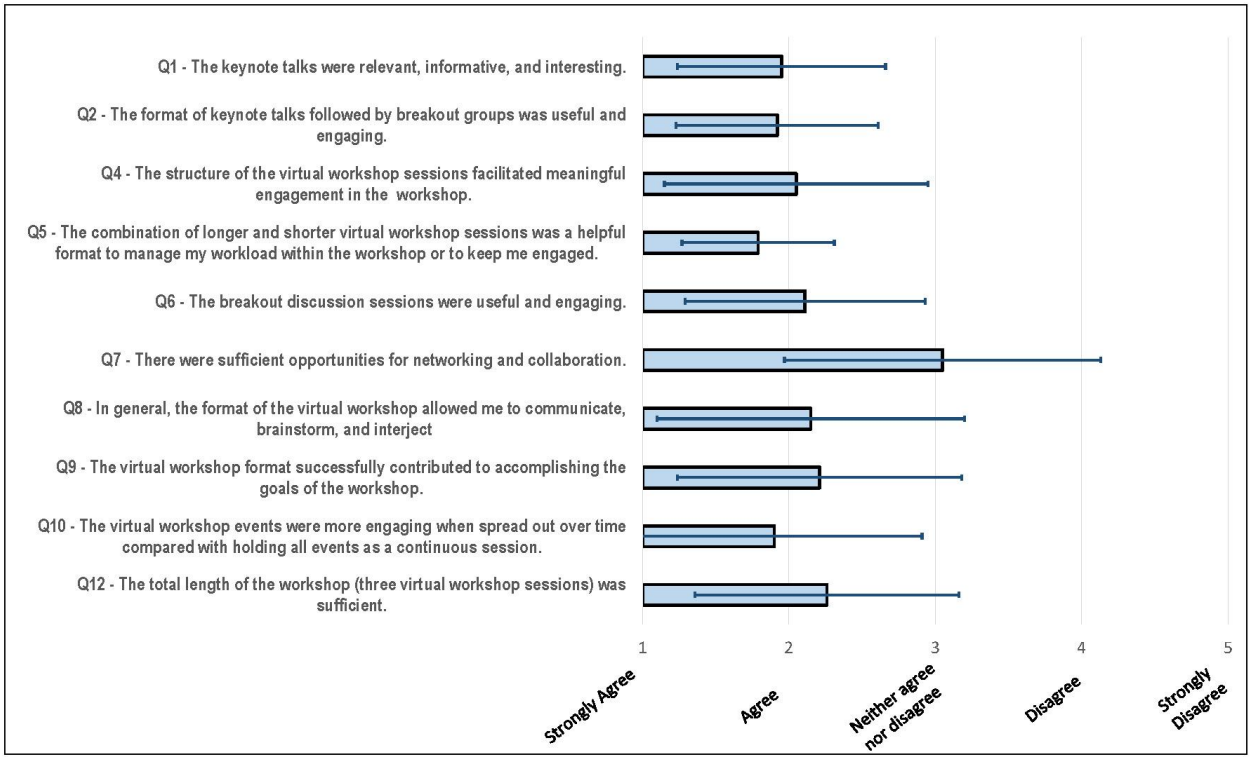

Figure B1. Summary of Workshop Survey Questions with Likert scale response.

Several survey questions requested a response on a Likert scale ranging from 1 (Strongly Agree) through 3 (Neither agree nor disagree) to 5 (Strongly Disagree). Figure B1 shows the mean and standard deviation of the responses to these questions. Attendees were most positive in Q5 about the helpful format of longer and shorter workshop sessions (the workshop session length decreased from the first day to the last day) to help them to manage their workload or to stay engaged, and in Q10 attendees indicated that the sessions were more engaging when spread out over time. Attendees were least positive (neutral) in Q7 to the question of whether there were sufficient opportunities for networking and collaboration. To improve the virtual workshop experience, it may help to include more free form unstructured "social" time to allow attendees to participate in pop-up discussions. For the remainder of the questions, attendees were generally positive (hovering around "Agree") in their responses.

In response to the question Q3: “The number of attendees was appropriate”, responses across five categories were “Too small” (0%), “Almost too small” (2.6%), “Appropriate” (69.2%), “Almost too large” (23.1%) and “Too large” (5.1%). The majority of attendees indicated that the workshop size was appropriate. However, almost ¼ of the attendees indicated that the workshop was too large. Future virtual workshops may consider either limiting the overall total number of attendees or structuring the workshop to be composed of smaller groups.

In response to question Q11 “Which workshop sessions did you attend?” Of the 39 survey respondents, 38 selected “Wednesday, August 19, 2020”, 38 selected “Tuesday, August 25, 2020”, and 36 selected “Wednesday, September 9, 2020”. The survey was not structured to identify individual participants who attended for multiple days.

In response to question Q13, “In my opinion, the successful virtual workshop sessions had the following characteristics: (Mark as many as may apply)”, of the 39 survey respondents, 18 indicated “Shorter sessions”, 7 indicated “Longer sessions”, 21 responded to “Previously had met other attendees in person”, 16 indicated “Time available at the beginning of the session to chat with other attendees”, 19 responded to “Time available between sessions to chat with other attendees”, and finally 9 indicated other characteristics not mentioned. Combining these characteristics based on responses, attendees indicated that successful virtual workshop sessions: 1) were shorter; 2) included individuals attendees had met in person before the workshop; 3) provided some unprogrammed time between sessions to allow attendees the opportunity to chat. Future virtual workshops may consider adding some structured time between structured sessions (using an agenda) using the ability to allow attendees to create and move between virtual breakout rooms to provide an opportunity for pop-up and free-form conversations.

There were many responses to the request in the survey for question Q15 seeking suggestions to increase the free flow of communications. The responses primarily focused on the virtual workshop process, with suggestions for improvement including: clarifying the role of the breakout group moderator, more structure within breakout groups, and adding more time within the agenda for free form discussions.

In response to question Q16 seeking comments on the virtual workshop experience, some responses indicated a positive experience, and there were also comments on improving the virtual workshop process.

Question Q17 sought comments on broad new areas or topics that attendees would like to see covered in a follow on workshop. Responses included: collaborations between agencies and universities, attracting high school students to the field, and diversity.

Responses to Question Q18 seeking suggestions for community building activities outside of the workshop included birds-of-a-feather at conferences, maintaining a Slack channel, and facilitating continued discussions and meetings on the topics discussed at the workshop.

Appendix C: Minority report

As a component of the workshop process, we also requested input from workshop attendees on any material objections that would represent a “minority report”.

One important point that was raised focused on the need to avoid adding more requirements (with associated administrative effort) to the request for proposal (RFP) frameworks often used by NSF. A workshop attendee summarized the brief discussion of this point. Essentially, there is opposition to any new required activities or sections in pre- or post-award for PIs or institutions related to addressing workforce development. The rationale for this request is that currently a significant amount of effort is already required from PI(s) and institutions to write high quality proposals and to administer existing awards.

For the specific recommendation in question, i.e. instituting a general workforce development or DEI required section analogous to Data Management plans, there are likely better alternatives to achieve the goal of a thriving and diverse workforce that would be more substantive than introducing more administrative work.

1. First, solicitations that are specifically focused on workforce development can, and should, have solicitation specific requirements that address DEI and workforce development.
2. Second, the NSF could provide clearer guidelines about specific intended outcomes and goals with the existing “Broader Impacts” requirement. There is a sense that there is no common understanding among PIs on what constitutes a “Broader Impact”. As a result, too many NSF proposals have less stellar Broader Impact sections than they could have. While the NSF statement [2] is broad and descriptive, it is much less clear and actionable advice than, e.g. the one Penn State provides online [1]. The goal of a diverse workforce is better served by PIs and institutions spending effort on their broader impact programs in their awards than on more paperwork.

[1] https://broaderimpacts.psu.edu/nsf-guidelines-and-trends/
[2] https://www.nsf.gov/pubs/2007/nsf07046/nsf07046.jsp

Appendix D. Workshop Agenda

Session 1 - Wednesday, August 19, 2020

10:00 Welcome, Introductions, Orientation
Thomas Hacker and Costa Michailidis

10:15 Keynote Presentations
Neil Chue Hong, Sharon Broude Geva, Dana Brunson

11:30 Asynchronous Q&A
All participants asked questions of speakers, speakers answering.

11:45 Voting On Critical Questions
Participants voted on challenge questions provided by the group.

12:00 Mixed Groups - Capturing Thematic Challenges in Breakout Groups
Participants discussed keynotes and challenge questions, and sought to write down a thematic challenge based on the keynotes and challenge questions.

12:20 Break

13:00 Affinity Groups - Defining Challenges, Defining Success
Participants discussed the thematic challenges in affinity group breakouts, and for each determined by their group to be significant, described the characteristics, impacts, and underlying root causes of each from the perspective of their affinity group. Then, for each thematic challenge they determined to be significant, their task was to discuss and write a statement that answers the question: "What does success look like?" to help articulate criteria for brainstorming potential solutions to these thematic challenges.

14:15 Lunch

15:15 Open Brainstorm
Participants were asked to individually review the thematic challenges, and to add their ideas in the workshop webpage for the thematic challenges.

15:35 Idea Generation Charrette Instructions

15:40 Idea Generation Charrette A
Participants broke out again into affinity groups to generate ideas for solutions for thematic challenges.

16:40 Break

17:00 Idea Generation Charrette B
Participants broke out into assigned mixed breakout groups to generate ideas for solutions for thematic challenges.

17:45 Closing Remarks

Session 2 - Tuesday, August 25, 2020

12:00 Welcome and Orientation
Thomas Hacker and Annemarie Boss

12:10 Browsing Ideas
Participants were asked to individually read through ideas generated during the first workshop session.

12:30 Synthesizing Recommendation in Groups
Participants were asked to break out into new mixed groups, and to work as a group to synthesize multiple ideas into more concrete and thorough recommendations.

13:35 Break

14:00 Report Outs from the Breakouts
Each breakout group briefly presented their recommendations and sought feedback.

15:00 Tagging Recommendations
Participants were asked to review the recommendations within the KIStorm platform, and to assign one or more "tags" to recommendations if they felt that the recommendation was highly feasible, had the potential for high impact, needed to grow/might be great, or represented an urgent matter.

15:15 Prioritizing Recommendations
Participants were asked to individually prioritize up to five recommendations.

15:30 Closing Remarks

Session 3 - Wednesday, September 9, 2020

12:00 Welcome and Orientation
Thomas Hacker and Costa Michalidis

12:05 Recommendations Summary
Thomas Hacker presented a summary of the recommendations developed during previous workshop sessions.

12:20 Individual Reflection
Participants were asked to individually think about actions they might take individually.

12:30 Teaming Up (Slack Session & Breakouts)
Participants were asked to identify an action for one of the recommendations that they would be willing to take a lead on if they had some help. Participants posted a note on the Actions page in the KIStorm platform for a potential action, and Zoom Breakout rooms were set up for each posted action note for participants to join if they were interested in discussing the potential action as a group.

13:15 Quick Report Out
Each breakout group gave a brief presentation on the possible actions that the group considered, and how participants could follow up.

13:45 Material Objections
Participants were asked to bring up for discussion any potential material objections to the problems, challenges, or recommendations discussed at the workshop.

13:55 Closing Remarks