Cyberinfrastructure Facilitation Skills Training via the Virtual Residency Program

Henry Neeman
hneeman@ou.edu
University of Oklahoma
Information Technology
Norman, Oklahoma, USA

David Akin
david.akin@ou.edu
University of Oklahoma
Information Technology
Norman, Oklahoma, USA

Hussein Al-Azzawi
azzawi@unm.edu
University of New Mexico
Ctr for Adv Research Computing
Albuquerque, New Mexico, USA

Kevin L. Brandt
kevin.brandt@sdstate.edu
South Dakota State University
Technology & Security
Brookings, South Dakota, USA

Jamene Brooks-Kieffer
jamenebk@ku.edu
University of Kansas
KU Libraries
Lawrence, Kansas, USA

Dana Brunson
dbrunson@internet2.edu
Internet2
Stillwater, Oklahoma, USA

Dirk Colbry
colbrydi@msu.edu
Michigan State University
Computational Math, Sci & Engr
East Lansing, Michigan, USA

Sandra Gesing
sandra.gesing@nd.edu
University of Notre Dame
Center for Research Computing
Notre Dame, Indiana, USA

Anna Klimaszewski-Patterson
anna.kp@csus.edu
California State Univ, Sacramento
Department of Geography
Sacramento, California, USA

Claire Mizumoto
claire@ucsd.edu
University of California, San Diego
Research IT Services
La Jolla, California, USA

Joy A. Pine-Thomas
japineth@uncg.edu
Univ of North Carolina Greensboro
School of Nursing
Greensboro, North Carolina, USA

Anita Z. Schwartz
anita@udel.edu
University of Delaware
IT - Research CyberInfrastructure
Newark, Delaware, USA

Horst Severini
severini@ou.edu
University of Oklahoma
Information Technology
Norman, Oklahoma, USA

Mohammed Tanash
tanash@ksu.edu
Kansas State University
Department Of Computer Science
Manhattan, Kansas, USA

Daniel Voss
dan.voss@beaumont.org
Beaumont Health Department
Beaumont Research Institute
Royal Oak, Michigan, USA

ABSTRACT

Cyberinfrastructure (CI) Facilitation is the process of helping researchers to use research computing systems and services to advance their computing/data-intensive research goals. The growing need for CI Facilitation isn't being met by traditional academic degree/certificate programs, so informal education is required. Since 2015, the Virtual Residency Program (VRP) has been teaching key CI Facilitation skills to pre-service and in-service CI Facilitators, at introductory, intermediate and advanced levels, via a combination of (a) workshops, (b) conference calls, and (c) apprenticeships. The demand for CI Facilitation has expanded exponentially, with the known CI Facilitator population growing from dozens in the late 2000s to over a thousand presently, so the need for this kind of training has become increasingly acute. The 2019 VRP workshop was presented at a combined introductory/intermediate level, with content suitable for both new and experienced CI Facilitators, to attract both of these populations, in many cases presenting the same topic first at the introductory level and then intermediate, to maximize impact. Since 2015, the VRP has served 924 participants from 370 institutions in every US state and 3 US territories plus 11 other countries.

PEARC '20, July 26–30, 2020, Portland, OR, USA

ACM ISBN 978-1-4503-6689-2/20/07…$15.00
https://doi.org/10.1145/3311790.3396629

CCS CONCEPTS

• **Social and professional topics** → **Computing occupations**; **Adult education**.

KEYWORDS

Cyberinfrastructure workforce development, research computing facilitation, Cyberinfrastructure facilitation

ACM Reference Format:
Henry Neeman, David Akin, Hussein Al-Azzawi, Kevin L. Brandt, Jamene Brooks-Kieffer, Dana Brunson, Dirk Colbry, Sandra Gesing, Anna Klimaszewski-Patterson, Claire Mizumoto, Joy A. Pine-Thomas, Anita Z. Schwartz, Horst Severini, Mohammed Tanash, and Daniel Voss. 2020. Cyberinfrastructure Facilitation Skills Training via the Virtual Residency Program. In *Practice and Experience in Advanced Research Computing (PEARC '20), July 26–30, 2020, Portland, OR, USA.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3311790.3396629

1 INTRODUCTION

Cyberinfrastructure Facilitation means working directly with researchers to provide Cyberinfrastructure (CI) expertise and mentoring, in order to advance the computing-intensive and/or data-intensive aspects of researchers' investigations. CI Facilitators are individuals with a complex mix of skills: not only do they need to be technically proficient in the use of advanced digital systems and services, they are also expected to be familiar with the full research lifecycle. In addition, they need excellent professional/interpersonal (soft) skills, because they work with a broad variety of researchers across the full spectrum of research disciplines. The national community of CI Facilitators has been growing rapidly, in terms of both new CI Facilitators and new institutions (see Section 2, below). This has led to a large number of inexperienced CI Facilitators, so there is a pressing need for CI Facilitation skills training.

The CI Facilitators Virtual Residency Program (VRP) [15–17, 19], founded in 2015, teaches CI Facilitation skills to prospective and current CI Facilitators, focusing primarily on professional/interpersonal (soft) skills development, because many of the relevant technical skills, for example parallel programming, already have well established training programs [13, 14, 23]. The VRP consists of four main components: (1) an annual workshop on CI Facilitation and CI leadership skills; (2) regularly scheduled conference calls on planning and development of the upcoming VRP workshop; (3) a multi-month virtual Grant Proposal Writing Apprenticeship; (4) a multi-month virtual Paper Writing Apprenticeship. In addition, the University of California system ran mini-workshops in 2017 and 2018 that were based on the VRP model (which are included in the overall VRP participant and institution counts, above and below).

To date, the VRP has had 924 participants from 370 institutions in every US state and 3 US territories plus 11 other countries, including:

- 56 Minority Serving Institutions (15% of VRP institutions);
- 94 non-PhD-granting institutions (25%), among them 3 high schools and 5 community colleges;
- 101 institutions (27%) in 27 of the 28 Established Program to Stimulate Competitive Research (EPSCoR) jurisdictions;
- 121 (93%) of the 130 Carnegie Classification [18] Very High Research Activity (R1) institutions that have central on-premise CI resources (out of 131 R1 institutions total);
- 54 of 82 (66%) High Research Activity (R2) institutions that have central on-premise CI resources and 21 of 53 (40%) R2s that don't have central on-premise CI resources (out of 135 R2 institutions total);
- on the US News rankings "2020 Best National Universities" [20], all of the Top 10 (100%), 23 of the Top 25 (88%), 45 of the Top 50 (85%), 83 of the Top 100 (81%), 116 of the Top 150 (76%), and 137 of the Top 200 (68%) (some percentages are due to ties at the last position in the group);
- 243 of the 370 VRP institutions are XSEDE Campus Champion (CC) [2, 5] institutions (75% of CC institutions, 66% of VPR institutions);
- 119 of 141 Campus Research Computing Consortium (CaRCC) [6] Researcher-Facing Track institutions (84% of CaRCC Researcher-Facing Track institutions, 32% of VRP institutions).

The VRP has been highly successful, as shown by its repeat rate: of the 330 institutions that started participating in the VRP before the 2020 workshop (and therefore could have participated in multiple activities), 257 of those institutions (78%) have participated in multiple VRP activities, and 237 VRP institutions (72%) have participated in multiple types of VRP activities (workshops, VRP workshop planning calls, grant proposal writing apprenticeship, paper writing apprenticeship). In total (including institutions that were new as of the 2020 workshop), 348 of the 370 VRP institutions (94%) have participated in the VRP workshops, 239 VRP institutions (65%) have participated in the VRP planning calls, 138 VRP institutions (37%) have participated in the Grant Proposal Writing Apprenticeship, and 111 VRP institutions (30%) have participated in the Paper Writing Apprenticeship.

The VRP annual summer workshops have been presented at multiple levels, varying year by year in response to input from Virtual Residents as well as to other relevant factors. These workshop levels are: (a) introductory, designed around pre-service and new CI Facilitators; (b) intermediate, designed around experienced CI Facilitators; (c) advanced, designed around extant and emerging institutional CI leaders. Specifically, the 2015-17 workshops were introductory (but included some advanced topics, such as grant proposal writing), the 2018 workshop was at a combined intermediate/advanced level, the 2019 workshop was at a combined introductory/intermediate level, and the 2020 workshop was at a combined intermediate/advanced level.

The motivations for providing these multiple levels of workshops are focused around several goals. For introductory and intermediate level content, the motivations are focused around developing the national CI Facilitator workforce, including both (a) expanding the population of CI Facilitators and (b) training them to be more effective. By contrast, the motivation for providing advanced content is to help experienced CI Facilitators to begin the transition to institutional CI leadership, and ultimately to national CI leadership in some cases. That is, the pool of potential CI leaders is being expanded via CI professionals who have extensive experience working directly with researchers to advance their computing/data-intensive research goals, and who therefore have a deep understanding of these researchers' needs, across the full spectrum of disciplines. This demand for institutional CI leaders is being exacerbated by (a) the growth in the number of institutions that have researchers engaged in computing/data-intensive research, as evidenced by the growth in the number of institutions that have CI Facilitators (see Fig. 2), and (b) the retirement rate of the Baby Boom generation (roughly 5% of Baby Boomers per year [8]), which risks creating a CI leadership gap across the US.

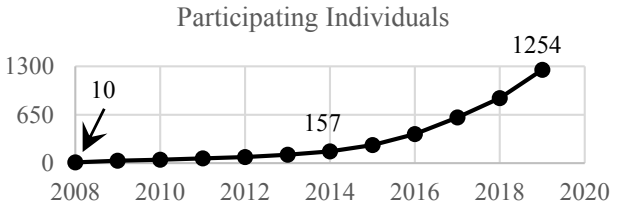

Figure 1: The number of distinct individuals participating in the Campus Champions and/or Virtual Residency and/or CaRCC Researcher-Facing Track, 2008-19.

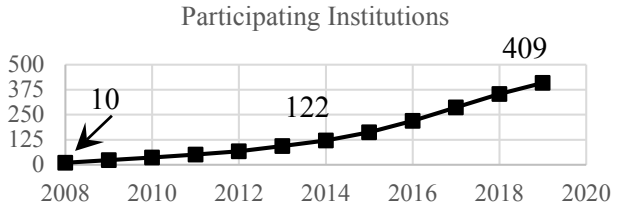

Figure 2: The number of distinct institutions participating in the Campus Champions and/or Virtual Residency and/or CaRCC Researcher-Facing Track, 2008-19.

2 DEMAND FOR CI FACILITATORS

CI Facilitation has become a key to success in computing/data-intensive research, because of (a) the increasing complexity of CI systems and software, (b) the increased ubiquity and applicability of CI to new disciplines and new categories of problems within disciplines, and (c) the growing mindset distance between consumer computing (e.g., handhelds) versus advanced CI resources (command line, Linux, remote, shared, batch).

The National Cyberinfrastructure Coordination Service Conference report [1] recommends the following: "Incentivize the development of new/ongoing efforts that bring together CI professionals to learn from one another and generate community efforts to identify and improve leading practices."

Currently, most high-ranked academic institutions have centralized on-premise CI capability: 130 of 131 R1s (99%) and 82 of 135 R2s (61%), and likewise all of the Top 25 institutions on the US News ranking of national universities, 49 of the Top 50, 95 of the Top 100, 132 of the Top 150 (88%), and 159 of the top 200 (80%). Many non-PhD-granting institutions are also pursuing these capabilities (for example, 25% of VRP institutions and 28% of CC institutions are non-PhD-granting).

CI organization leaders — for example, High Performance Computing (HPC) center directors — recognize the need for CI facilitation. In our recent survey of CI organization leaders at R1s and R2s that have on-premise CI (see section 3, below), 87% of responding CI organization leaders said that CI Facilitators are very important to their computing/data-intensive researchers. The subset of these CI organization leaders who are at R1s reported a mean of 7.3 in-service CI Facilitators, and those at R2s reported a mean of 2.7 in-service CI Facilitators. (These are distinct individuals currently employed as CI Facilitators, not Full Time Equivalents — many CI professionals do CI Facilitation as only a portion of their responsibilities.) In the coming five years, the R1 subset of these CI organization leaders expect to need a mean of 5.5 more CI Facilitators, and at R2s a mean of 2.2 more CI Facilitators. Thus, five years from now, the total number of distinct individuals employed as full- or part-time CI Facilitators at R1s and R2s that have on-premise CI is expected to be at least 2065 (plus a more modest number at other institution types and at R2s that add on-premise CI resources in the meantime). We use CI organization leaders as a proxy of researcher need, because, despite their chronically limited budgets for advanced computing infrastructure (hardware, software, systems-facing personnel, etc.), CI organization leaders at R1s and R2s choose to invest, in many cases, six figures per year for salary and fringe of CI Facilitators, recognizing the value that CI Facilitators provide to computing/data-intensive researchers.

The need for CI Facilitation is also seen in our recent survey of CI Facilitators (see section 3, below), who estimated a mean of 1802 researchers served per CI Facilitator over their entire CI Facilitation career (mean of 16.4 years). To the nearest order of magnitude (because each researcher may work with multiple CI Facilitators), CI Facilitators as a group are expected to serve on the order of a million researchers over their CI Facilitation careers.

The growth of both the number of CI Facilitators and the number of institutions that they serve has been exponential: from 2008 to the end of 2019, considering, as a proxy for the number of CI Facilitators, participants in one or more of (a) the VRP, and/or (b) the XSEDE Campus Champions, and/or (c) the CaRCC Researcher-Facing Track, distinct individuals grew from 10 to 1254 (see Fig. 1), a Compound Annual Growth Rate (CAGR) of 41% (doubling every 1.7 years), and distinct institutions grew from 10 to 409 (see Fig. 2), a CAGR of 32% (doubling every 2.2 years), because many institutions have multiple CI Facilitators.

3 SURVEY DATA COLLECTION

The VRP team conducted two anonymous surveys during the 2019-20 Grant Proposal Writing Apprenticeship (see section 1, above): the first of CI Facilitators about their CI Facilitation experience and longevity and the second of CI organization leaders such as HPC center directors.

3.1 CI Facilitator Questions and Responses

(University of Oklahoma Institutional Review Board exemption # 11320, 10/28/2019)

(1) In a typical year, what is your BEST ESTIMATE of the number of DISTINCT INDIVIDUAL RESEARCHERS (NOT research teams) for whom you provide Cyberinfrastructure (CI) Facilitation?
Include undergraduates, graduate students, postdocs, faculty, staff, non-academic professionals and any other relevant categories of researchers.
Include those who were/are/will be at your institution, as well as any others even if at other institutions.
Count each individual researcher ONLY ONCE, even if you provide CI Facilitation to that same person multiple times in

that typical year.
Mean: 199; n: 137 (of 899 contacted, 15% response rate)

(2) How many years have you ALREADY been a CI Facilitator?
Mean: 7.1; n: 136

(3) What is your BEST ESTIMATE of the number of ADDITIONAL years that you expect to continue to be a CI Facilitator?
EXCLUDE the number of years you've already been a CI Facilitator, which was covered in question Q2, just above.
Mean: 9.3; n: 136

(4) What is your BEST ESTIMATE of the TOTAL NUMBER of distinct individual researchers to whom you have provided, and will provide, CI Facilitation, during your ENTIRE CAREER as a CI Facilitator (past and future)?
Consider your answers to ALL of the questions above (Q1, Q2, Q3), and bear in mind that you will provide CI Facilitation to some individual researchers repeatedly during multiple years (so it's not as simple as multiplying total career years by researchers per year).
As before, include undergraduates, graduate students, postdocs, faculty, staff, non-academic professional and any other relevant categories of researchers, both at your institution and at other institutions.
Mean: 1802; n: 136

3.2 CI Organization Leader Questions and Responses

(University of Oklahoma Institutional Review Board exemption # 11498, 12/2/2019)

(1) Which of the following is the type of institution you're at?
If you aren't sure, please look up your institution at:
http://carnegieclassifications.iu.edu/lookup/lookup.php
1. Academic: Doctoral Very High Research Activity (R1)
2. Academic: Doctoral High Research Activity (R2)
...
R1 73%, R2 27%; n: 47 R1, 15 R2 (of 181 contacted, 34% response rate)

(2) Thinking about CI Facilitators employed by your institution: How many distinct individuals are CURRENTLY employed as CI Facilitators at your institution (whether full time or part time)?
R1 Mean: 7.3, R2 Mean: 2.7; n: 47 R1, 15 R2

(3) Thinking about CI Facilitators employed by your institution: How many ADDITIONAL distinct individuals do you EXPECT to need as CI Facilitators at your institution within the COMING 5 YEARS (whether full time or part time)?
R1 Mean: 5.5, R2 Mean: 2.2; n: 47 R1, 15 R2

(4) How important to your computing-intensive/data-intensive researchers (e.g., supercomputer users) are the services that your CI Facilitators provide? Here, researchers includes all of faculty, staff, postdocs, graduate students, undergraduates, non-academic professionals, etc.
Very: 87%, Moderately: 11%, Slightly: 2%; n: 62

(5) How many individual computing-intensive/data-intensive researchers (e.g., supercomputer users) does your institution have? Please give your BEST ESTIMATE.
Here, researchers includes all of faculty, staff, postdocs, graduate students, undergraduates, non-academic professionals, etc.
R1 Mean: 1332, R2 Mean: 1024; n: 47 R1, 15 R2

(6) How valuable would it be, to those researchers, for your CI Facilitators to have training in key CI Facilitation skills (including both technical/hard and especially interpersonal/soft skills)?
Very: 74%, Moderately: 21%, Slightly: 5%; n: 62

(7) If there were a high quality CI Facilitation skills certification program available, would you ENCOURAGE CI Facilitators ALREADY IN SERVICE AT YOUR INSTITUTION to integrate that certification into their professional development trajectory?
Yes: 90%, No: 10%; n: 62

4 VIRTUAL RESIDENCY WORKSHOPS 2015-2018

Since 2015, the VRP has held annual summer workshops, funded primarily by National Science Foundation (NSF) grants (see Acknowledgements, below). The 2015-17 workshops [15, 17] were all at the introductory level, designed for pre-service and in-service CI Facilitators, including CI professionals for whom CI Facilitation was only part of their duties (e.g., CI system administrators who help researchers to use local CI resources).

The 2018 VRP workshop was listed as intermediate, but had a substantial fraction of CI Leadership content (roughly 2/3 of total content), so it was actually a combination of intermediate and advanced. Workshop attendance 2015-18 was as follows:

- 2015: 50 participants (28 onsite, 22 remote) from 39 institutions in 26 US states and 1 US territory;
- 2016: 100 participants (43 onsite, 57 remote) from 60 institutions in 33 US states and 3 other countries;
- 2017: 196 participants (51 onsite, 145 remote) from 134 institutions in 47 US states and 2 US territories plus 3 other countries;
- 2018: 216 participants (38 onsite, 178 remote) from 147 institutions in 43 US states and 1 US territory plus 2 other countries.

Topics for the 2015-18 workshops are found in [17], [15] and [16].

5 VIRTUAL RESIDENCY WORKSHOP 2019

The 2019 VRP workshop had a record-breaking 254 attendees (37 onsite, 217 remote) from 162 institutions in 45 US states and 3 US territories plus 5 other countries, including:

- 27 Minority Serving Institutions (17% of participating institutions);
- 44 non-PhD-granting institutions (27%);
- 52 institutions (32%) in 24 of 28 (86%) EPSCoR jurisdictions;
- 130 Campus Champion institutions (80%).

The 2019 VRP workshop was a mix of introductory and intermediate material, because of limited intermediate content in past workshops. This workshop used a two-stage approach: for a specific topic, an introductory session (typically a talk) was held first, and

then an intermediate session (typically a panel) on the same or similar topic was held later in the week. While most of the topics had been presented at previous VRP workshops, most of the sessions contained substantial new content, because of a mix of (a) new topics, (b) extant topics at a new level, and (c) new contributors with fresh perspectives. Of the 2019 VPR workshop's 43 presenters, moderators and panelists, 24 (56%) hadn't presented at previous (2015-18) VRP workshops, so were able to provide novel points of view.

5.1 2019 Workshop Content

5.1.1 CI User Support. This session focused on facilitating the use of CI resources by computing/data-intensive researchers. It introduced activities that resulted in a dramatic increase in CI users and research productivity.

CI users come from a broad variety of disciplines, both Science, Technology, Engineering and Mathematics (STEM) and non-STEM, that rely on CI resources for simulation, visualization, experiments, and processing of tremendous volumes of data. Consequently, researchers can face technical challenges when working on their own, which can be a high barrier to research success. This session introduced best practices for resolving users' issues, outreach, and how to educate CI users about mission critical skills for computational investigations. The session also introduced successful policies for maintaining CI resources in a manner that best serves the needs of CI users, without excessively burdening systems-facing professionals.

5.1.2 The Structure of a Cluster Supercomputer. This session focused on introducing new CI Facilitators to the main components of HPC clusters, to help them build better understanding of how HPC systems work. The session started with defining what an HPC cluster is, followed by a few national examples. The need for HPC systems was discussed and explained with simplified mathematical calculations showing the impact of simulations and parallelism, justifying substantial investments. That was followed by academic HPC challenges (staffing, funding, etc). Then, the primary HPC components were discussed, including: compute, network, storage, resource manager/scheduler, and software, as well as benchmarks, backups, data center resources, and security and monitoring tools. Real life technical diagrams of HPC systems showed all the HPC components connected and working together in a production environment. Next, concepts of cloud, on-premise, and hybrid HPC models were discussed, including advantages and disadvantages. Finally, the session closed with a conversation about future HPC technologies such as quantum computing.

5.1.3 Introduction to the CyberAmbassadors Program. In this session, participants were introduced to the CyberAmbassadors program, an NSF-funded project to develop modular, open source, professional skills training [3]. The curriculum helps CI Professionals strengthen communication, teamwork and leadership skills, in order to more effectively contribute to computing/data-intensive, interdisciplinary STEM projects [11]. This included parts of the module "First Contact: Communicating with Purpose." Emphasis was on the importance of making a positive and productive first impression, with practice via scenarios covering networking, forming teams, and one-on-one consulting.

5.1.4 Setting Up an XSEDE Account and Submitting a Batch Job. This session focused on XSEDE user onboarding, with step-by-step instructions for creating an XSEDE user account [22]. After creating the new account, attendees were given information about Multi-Factor Authentication (MFA) and how to use this feature in the XSEDE Single Sign-On (SSO) Hub. In the context of MFA and the Hub, the goal is to streamline user access to computing and storage resources, safeguarding against unauthorized system access [10]. After successfully logging into an XSEDE system via the SSO Hub, attendees were shown how to create Slurm batch scripts for Message Passing Interface (MPI) jobs, containing the commands to be executed to run a parallel compute job within an XSEDE cluster compute service [12]. After job submission, users were shown how to monitor a job in the cluster queue manager and how to view program output after the compute job completed [21].

5.1.5 Effective Communication. This intermediate panel focused on the following questions:

- What are some "tricks of the trade" that you use to communicate effectively with researchers?
- What kinds of ideas, concepts, technologies, techniques etc do you find are the most challenging to communicate or explain? What do you do about that?
- How do you follow up after the conversation and how can that followup act as a verification of understanding by all parties?
- Who are the speakers or writers who you find the most effective at communicating? What approaches of theirs do you like to use?
- You will enter conversations where there are preconceived notions. What techniques do you use to move beyond these preconceptions?
- How do you deal with situations where there's asymmetric knowledge between the researcher and the facilitator (and unfair expectations a researcher may have of a facilitator)?
- Suppose you need to persuade a researcher to do something that they otherwise wouldn't choose to do. How do you approach that? What has been the most effective?

5.1.6 Working Effectively with Systems-Facing Professionals. This session began by defining the position and mindset of a systems-facing professional, exploring questions such as:

- What roles does a systems-facing professional play in CI?
- What does it mean to be in the role of a technical expert in CI?

The goal was to help the attendees understand the perspective of systems-facing professionals.

Then, the discussion was steered to role-playing the interaction between researchers and systems-facing professionals. We used the opportunity to expand beyond the norms of how each role typically interacts with the other. Specific focus was given to shared values (the actual research) and challenges (both technological and interpersonal). The goal was to stimulate conversations that could continue offline after the workshop.

5.1.7 CyberAmbassadors: Communicating a Problem. The "Let's Talk: Communicating about Problems" module of the CyberAmbassador curriculum focuses on building participants' capacity to engage in meaningful, one-on-one conversations about challenging topics [7]. The interactive session includes practice for communicating about problems with subordinates, peers and superiors, and outlines communication "algorithms" that can be used to solve a problem while also maintaining relationships. The exercises in this module give participants tools they can apply to difficult conversations in their professional and personal lives.

5.1.8 Experiences Working with Faculty. This panel included highly experienced members of the CI Facilitator community sharing their experiences, as a springboard to conversation throughout the workshop, providing strategies to expand workshop attendees' toolkits and to stretch their thinking to include next-level research support. Questions included:

- Talk through the evolution/development of a CI Facilitator's relationship with a PI at your institution.
- Share strategies to deal with scope creep: How do you clearly communicate what your scope of work will include? Are there certain scenarios where scope creep is necessary to accomplish the CI Facilitator mission?
- How much influence do you have in addressing gaps in researcher support, championing the researchers effectively, and promoting change to the institution's administration?

5.1.9 CyberAmbassadors: Problem Solving and Decision Making in a Group. This module of the CyberAmbassadors curriculum focuses on helping CI Facilitators become effective at problem solving and decision-making activities. Specific skills for catalyzing brainstorming sessions are discussed, along with practical tools for helping groups distill the results of brainstorming into actionable solutions and activities, to help spark creativity. This session allowed participants to practice skills they can use to lead teams and to help make effective decisions.

5.1.10 The CI Milieu: Organizations. This session was comprised of CI leaders who gave a high-level overview of the organizations they're involved with, sharing their unique knowledge and backgrounds. The organizations that were presented included: Campus Champions, CaRCC, Women in HPC, The Carpentries, Science Gateways Community Institute, Working Towards Sustainable Software for Science: Practice and Experience, US Research Software Sustainability Institute, UK Research Software Engineer Association, US Research Software Engineer Association, Coalition for Academic Scientific Computation, and the Linux Clusters Institute (LCI).

The panel also addressed the following questions:

- How do I join and what are the requirements to be part of the organization?
- How many members are there, and what types of CI roles and responsibilities do the members have?
- What is the time and effort commitment?
- How often does the group meet? Virtual or in person?
- Is there a cost to join?

5.1.11 Facilitation from Other Perspectives. This session brought perspectives from facilitators in scientific visualization, data science/machine learning, statistics, and research data management. It focused on questions such as:

- How do you do an "intake interview" with a new researcher/team? What are the techniques you use to learn what a researcher/team is trying to accomplish and how best to accomplish their goals?
- How do you learn the lingo and needs of a new research discipline that you haven't worked with before?
- How do you balance what you should teach the researcher/team to do for themselves, versus what you should do for them?
- How did you learn to do your job? What advice would you have for someone getting started in the same line of work today?
- What do you find is the balance between what the researcher/team believes is the right way to do things, versus ideas/concepts/techniques that you're aware of but that they aren't aware of?

5.1.12 How to Help Researchers Identify the Right Resources for Their Research. This panel consisted of CI Facilitators discussing their perspectives on helping researchers identify the best resources for their research. This included approaches for researcher engagement, institutional support, education and training, relationship building, and communication. Questions included:

- Where do you start (one-on-one versus group strategies, e.g., surveys, group meetings, benchmarking)?
- What are strategies for supporting research in a field you (or your team) do not understand?
- How do you support technical capability limitations, especially homegrown code versus community codes?
- Where do you point users for additional information (e.g., private/public clouds, XSEDE, regional resources, local resources, IT resources like VM farms, etc.)?

5.1.13 Research Data Management. This panel focused on presenting enough information about Research Data Management (RDM) [4, 9] concepts, techniques, tools, and resources so that CI Facilitators could begin navigating the RDM landscape. The panelists originated from research-intensive universities, a liberal arts college, a federal institute, and a not-for-profit RDM software provider. The session was an opportunity for these panelists to discuss RDM experiences in institution-specific and broader contexts.

First, each panelist was asked to talk about one or two RDM-related services or programs that they have facilitated in their local context. The goal was to demonstrate to attendees the variety of shapes that RDM can take, depending on the institution, the specific unit(s), and the people involved. A question to stimulate discussion was, "Where have you already encountered the need for RDM in the work that you do?"

Next, panelists were asked to discuss RDM beyond the bounds of their individual institutions:

(1) cultivating communication across differing priorities of researchers, RDM specialists, and computing specialists;
(2) training citizen scientists in RDM;

(3) automating workflows and RDM at scale;
(4) promoting open science, data sharing where possible, and reproducible research;
(5) facilitating data as a research object meriting publication;
(6) creating and branding tools specific to RDM.

The goals for this section were to show applications of RDM tools and techniques, and to demonstrate the array of expertise available. Questions prepared for this session included:

- How will you translate RDM needs and requirements between researchers and HPC specialists?
- What is your plan for identifying RDM specialists at your organization and communicating with them?
- What have you been surprised to discover that your researchers don't know about RDM?
- What are the key issues that you've seen work well to get your researchers to pursue/collaborate with you on RDM?

At the end of the panel, two readings were suggested, for understanding basic RDM practices and for communicating with each other, with RDM specialists, and with researchers.

5.1.14 Community Engagement. This session was developed during the workshop, based on attendees voting to suggest topics for a Friday timeslot. Two example outreach presentations were discussed, to illustrate why certain topics are covered in outreach talks, depending on the audience and goals. Following that, the presenters facilitated discussions on knowing who to engage (researchers, funding agencies, administrators, students, etc.), how to credibly engage stakeholders, and the challenges and successes they've had engaging their communities.

5.1.15 Sessions Described Elsewhere. The following sessions were held at the 2019 workshop and are described in [17], [15] and/or [16].

- "Virtual Residency Workshop Overview"
- "Effective Communication: How to Talk to Researchers about Their Research"
- ''Faculty: Tenure, Promotion, and Reward System"
- "Research Networking Overview"
- "How to Do an Intake Interview"
- "Speed Dating - Practicing the Intake Interview"
- "The CI Milieu: Systems, Service Providers, Technologies"
- "Working with Researchers from Non-traditional Disciplines"
- "Stories from the Trenches"

6 FUTURE WORK

The 2020 VRP workshop was held June 1-5 2020 at a combined intermediate/advanced level. The workshop had 430 attendees from 225 institutions in 49 US states and 3 US territories plus 7 other countries, including 31 Minority Serving Institutions (14% of this workshop's institutions), 44 non-PhD-granting institutions (20%), and 65 institutions in 26 of 28 EPSCoR jurisdictions (29%).

Registration opened in January 2020, with 50 registrations in just over a day, 100 in under three days, 150 in two weeks, 200 in under a month, 300 in under 3 months and 450 in under 4 months, even though the agenda was still in active development during much of the preregistration period and therefore hadn't yet been released during most of that period. This shows the CI Facilitator community's high faith in the VRP.

Looking deeper into the future, the Grant Proposal Writing Apprenticeship has resubmitted an NSF CyberTraining proposal to create a professional development certification program for CI Facilitators. Under this program, content will be split into modules covering not only technical topics but especially professional/interpersonal (soft) skills. Each module will consist of a badge, including an examination instrument and a scoring rubric for each badge. Certification will come from specific combinations of badges that collectively demonstrate mastery of key skills.

ACKNOWLEDGMENTS

This work was partially supported by National Science Foundation grant nos. 1440783 ("CC*IIE Engineer: A Model for Advanced Cyberinfrastructure Research and Education Facilitators"), 1548562 ("XSEDE 2.0: Integrating, Enabling and Enhancing National Cyberinfrastructure with Expanding Community Involvement"), 1620695 ("RCN: Advancing Research and Education Through a National Network of Campus Research Computing Infrastructures – The CaRC Consortium"), 1649475 ("Cyberinfrastructure Leadership Academy"), and 1730137 ("CyberTraining: CIP - Professional Skills for CyberAmbassadors"). This paper was developed collaboratively by a total of 81 participants in the 2019-20 Paper Writing Apprenticeship (see section 1, above). Portions of this paper were originally developed for various grant proposals and grant reports. Several workshop sessions were presented by guest lecturers, moderators and panelists. The authors are grateful to all of these collaborators.

REFERENCES

[1] Alan Blatecky, Dana Brunson, Thom Cheatham, Joel Gershenfeld, Angel Hedberg, Jim Bottom, and Dan Reed. 2019. *National Cyberinfrastructure Coordination Service Conference: Rethinking NSF's computational ecosystem for 21st century science and engineering.* https://www.rti.org/publication/national-cyberinfrastructure-coordination-service-conference.
[2] Marisa Brazil, Dana Brunson, Aaron Culich, Lizanne DeStefano, Douglas Jennewein, Tiffany Jolley, Timothy Middelkoop, Henry Neeman, Lorna Rivera, Jack Smith, and Julie Wernert. 2019. Campus Champions: Building and Sustaining a Thriving Community of Practice Around Research Computing and Data. In *Proceedings of the Practice and Experience in Advanced Research Computing on Rise of the Machines (learning).* ACM. https://doi.org/10.1145/3332186.3332200
[3] Astri Briliyanti, Julie Rojewski, T. J. Van Nguyen, Katy Luchini-Colbry, and Dirk Colbry. 2019. The CyberAmbassador Training Program. In *Proceedings of the Practice and Experience in Advanced Research Computing on Rise of the Machines (learning).* ACM. https://doi.org/10.1145/3332186.3332218
[4] Kristin Briney. 2015. *Data Management for Researchers: Organize, maintain, and share your data for research success.* Exeter Pelagic Publishing.
[5] Campus Champions. [n.d.]. *Campus Champions webpage.* https://www.xsede.org/community-engagement/campus-champions.
[6] Campus Research Computing Consortium. [n.d.]. *Campus Research Computing Consortium website.* https://carcc.org/.
[7] Katy Colbry and Dirk Colbry. 2019. *Mastering Complex Communication.* https://www.inscits.org/training-and-development.
[8] Richard Fry. 2019. *Baby Boomers are staying in the labor force at rates not seen in generations for people their age.* https://www.pewresearch.org/fact-tank/2019/07/24/baby-boomers-us-labor-force/.
[9] Celeste Headlee. 2017. *We Need to Talk: How to have conversations that matter.* Harper Wave.
[10] Xinyi Huang, Yang Xiang, Elisa Bertino, Jianying Zhou, and Li Xu. 2014. Robust multi-factor authentication for fragile communications. *IEEE Transactions on Dependable and Secure Computing* 11, 6 (2014), 568–581. https://doi.org/10.1109/tdsc.2013.2297110
[11] Katy Luchini-Colbry, Dirk Joel-Luchini Colbry, Julie Rojewski, and Astri Briliyanti. 2019. Partners in Professional Development: Initial Results from a Collaboration Between Universities, Training Programs, and Professional

Societies. In *Proceedings of the 2019 ASEE Annual Conference & Exposition*. https://peer.asee.org/33159

[12] MPI Forum. [n.d.]. *MPI Forum website*. https://www.mpi-forum.org/.

[13] National Computational Science Institute. [n.d.]. *National Computational Science Institute website*. http://computationalscience.org/.

[14] Henry Neeman. [n.d.]. *Supercomputing in Plain English webpage*. http://www.oscer.ou.edu/education/.

[15] Henry Neeman, Hussein M. Al-Azzawi, Aaron Bergstrom, Zoe K. Braiterman, Dana Brunson, Dirk Colbry, Eduardo Colmenares, Akilah N. Fuller, Sandra Gesing, Maria Kalyvaki, Claire Mizumoto, Jeho Park, Anita Z. Schwartz, Jason L. Simms, and Rustomji Vania. 2018. Progress Update on the Development and Implementation of the Advanced Cyberinfrastructure Research & Education Facilitators Virtual Residency Program. In *Proceedings of the Practice and Experience on Advanced Research Computing*. ACM. https://doi.org/10.1145/3219104.3219117

[16] Henry Neeman, Hussein M. Al-Azzawi, Dana Brunson, William Burke, Dirk Colbry, Jeff T. Falgout, James W. Ferguson, Sandra Gesing, Joshua Gyllinsky, Christopher S. Simmons, Jason L. Simms, Mohammed Tanash, Daniel Voss, Jason Wells, and Scott Yockel. 2019. Cultivating the Cyberinfrastructure Workforce via an Intermediate/Advanced Virtual Residency Workshop. In *Proceedings of the Practice and Experience in Advanced Research Computing on Rise of the Machines (learning)*. ACM. https://doi.org/10.1145/3332186.3332204

[17] Henry Neeman, Aaron Bergstrom, Dana Brunson, Carrie Ganote, Zane Gray, Brian Guilfoos, Robert Kalescky, Evan Lemley, Brian G. Moore, Sai K. Ramadugu, Alana Romanella, Johnathan Rush, Andrew H. Sherman, Brian Stengel, and Daniel Voss. 2016. The Advanced Cyberinfrastructure Research and Education Facilitators Virtual Residency: Toward a National Cyberinfrastructure Workforce. In *Proceedings of the XSEDE16 Conference on Diversity, Big Data, and Science at Scale - XSEDE16*. ACM Press. https://doi.org/10.1145/2949550.2949584

[18] The Carnegie Classification of Institutions of Higher Education. [n.d.]. *The Carnegie Classification of Institutions of Higher Education website*. https://carnegieclassifications.iu.edu/.

[19] Virtual Residency Program. [n.d.]. *Virtual Residency Program Workshops webpage*. http://www.oscer.ou.edu/virtualresidency/.

[20] US News & World Report. [n.d.]. *US News 2020 Best National University Rankings webpage*. https://www.usnews.com/best-colleges/rankings/national-universities.

[21] SchedMD. [n.d.]. *Slurm Workload Manager website*. https://slurm.schedmd.com/.

[22] XSEDE. [n.d.]. *Extreme Science and Engineering Discovery Environment (XSEDE) website*. http://www.xsede.org/.

[23] XSEDE. [n.d.]. *XSEDE Training webpage*. https://www.xsede.org/for-users/training.