The Advanced Cyberinfrastructure Research and Education Facilitators Virtual Residency: Toward a National Cyberinfrastructure Workforce

Henry Neeman
University of Oklahoma
301 David L. Boren Blvd.
Norman Oklahoma 73019
405-325-5386
hneeman@ou.edu

Aaron Bergstrom
University of North Dakota
NDUS IT Building, James Ray Drive
Grand Forks North Dakota 58202
701-777-2075
aaron.bergstrom@und.edu

Dana Brunson
Oklahoma State University
MS102
Stillwater Oklahoma 74078
405-744-4455
dana.brunson@okstate.edu

Carrie Ganote
Indiana University
2709 E. 10th St..
Bloomington IN 47408
812-855-9255
cganote@iu.edu

Zane Gray
University of Oklahoma
905 Asp Avenue
Norman Oklahoma 73019
405-325-7564
zgray@ou.edu

Brian Guilfoos
Ohio Supercomputer Center
1224 Kinnear Road
Columbus Ohio 43212
614-292-2846
guilfoos@osc.edu

Robert Kalescky
Southern Methodist University
3225 Daniel Avenue
Dallas Texas 75275
214-768-3070
rkalescky@smu.edu

Evan Lemley
University of Central Oklahoma
100 N. University Dr.
Edmond Oklahoma 73034
405-974-5473
elemley@uco.edu

Brian G. Moore
South Dakota State University
UNSS-Box 2201
Brookings South Dakota 57007
605-688-4679
brian.moore@sdstate.edu

Sai Kumar Ramadugu
The University of Iowa
15 Lindquist Center
Iowa City Iowa 52242
319-335-5330
sramadug@iowa.uiowa.edu

Alana Romanella
Virginia Tech
3050 Torgerson Hall
Blacksburg, VA 24061
540-231-0284
aromanel@vt.edu

Johnathan Rush
University of Illinois at Urbana-Champaign
1205 W. Clark St.,
Urbana Illinois 61801
217-300-1384
jfr@illinois.edu

Andrew H. Sherman
Yale University
160 St. Ronan Street
New Haven Connecticut 06511
203-436-9171
andrew.sherman@yale.edu

Brian Stengel
University of Pittsburgh
4200 Fifth Ave.
Pittsburgh Pennsylvania 15260
412-624-8688
bstengel@pitt.edu

Dan Voss
University of Kansas
2335 Irving Hill Road
Lawrence Kansas 66045
785-864-7245
dan.voss@ku.edu

ABSTRACT

An Advanced Cyberinfrastructure Research and Education Facilitator (ACI-REF) works directly with researchers to advance the computing- and data-intensive aspects of their research, helping them to make effective use of Cyberinfrastructure (CI). The University of Oklahoma (OU) is leading a national "virtual residency" program to prepare ACI-REFs to provide CI facilitation to the diverse populations of Science, Technology, Engineering and Mathematics (STEM) researchers that they serve. Until recently, CI Facilitators have had no education or training program; the Virtual Residency program addresses this national need by providing: (1) training, specifically (a) summer workshops and (b) third party training opportunity alerts; (2) a community of CI Facilitators, enabled by (c) a biweekly conference call and (d) a mailing list.

XSEDE16, July 17 - 21, 2016, ,
ACM 978-1-4503-4755-6/16/07.
http://dx.doi.org/10.1145/2949550.2949584

Categories and Subject Descriptors
K.4.m [**Computers and Society**]: Miscellaneous.

General Terms
Management, Performance

Keywords
Cyberinfrastructure, workforce development, train the trainer

1. INTRODUCTION
An Advanced Cyberinfrastructure Research and Education Facilitator (ACI-REF) works directly with researchers to address the computing- and data-intensive aspects of their research, by helping them to make optimal use of Cyberinfrastructure (CI). Via a National Science Foundation (NSF) Campus Cyberinfrastructure - Infrastructure, Innovation and Engineering (CC*IIE) grant (ACI-1440783, "A Model for Advanced Cyberinfrastructure Research and Education Facilitators," 9/15/2014-9/14/2016, $400,000), the University of Oklahoma (OU), specifically the OU Supercomputing Center for Education & Research (OSCER), is leading a national "Virtual Residency" program [1] to more efficiently and effectively prepare ACI-REFs to provide CI facilitation to diverse populations of Science, Technology, Engineering, and Mathematics (STEM) researchers, addressing a crucial national need. As a National Academies report [2] states:

> *Sustainable and effective cyberinfrastructure depends critically on the skills and expertise … of committed and well-trained advanced computing professionals. … The report of the NSF Task Force on Cyberlearning and Workforce Development [3] … recognizes the need to train … the workforce that supports advanced computing ….*

Currently, those hired into CI Facilitator positions may have strong backgrounds in a STEM discipline, often not in Computer Science, but many aren't and haven't been faculty. Some have little or no research experience, especially if they have been Information Technology (IT) professionals instead of academics. And, even when they have strong research experience in their own disciplines, they may have little or no exposure to research in other disciplines – a potential problem, because the CI Facilitator role requires addressing needs of users across the full spectrum of computing- and data-intensive research institutionwide, including not only STEM disciplines, but at some institutions humanities and/or arts as well. Until recently, there have been few if any education or training opportunities for CI Facilitators, despite the fact that they are crucial enablers of many investigations. While there have been course and workshop opportunities covering specific techniques and technologies (e.g., parallel computing), these are no substitute for training on how to facilitate Computational and Data Enabled Science & Engineering (CDS&E) research. OU's Virtual Residency program is addressing this national need, by providing training opportunities (specifically summer workshops and CI technology training opportunity alerts), and by community building via biweekly conference calls and a mailing list.

As institutions continue to develop campus CI and related capabilities, it is imperative that they also develop knowledge and social acumen in the practice of CI leadership and engagement. Organizations charged with supporting STEM research need to be trusted sources for CI expertise, best practices, and functional support for research-focused CI at the institutional level. These organizations can also contribute to the growing national community of CI Engineers and CI Facilitators. Institutional CI organizations are called on to address CI-focused challenges that CI Facilitators and CI Engineers are well positioned to ameliorate:

- **At some institutions, CI, and the language used to describe and discuss it, can be unfamiliar to some of the researchers who could benefit from it.** Such familiarity is more commonplace within the network engineering and High Performance Computing (HPC) communities, but professionals in those communities may have difficulty in communicating effectively with researchers. CI Facilitators and CI Engineers can establish consistency in descriptions of CI on campus and can ensure that CI services are defined and described in ways that are meaningful to a broad spectrum of investigators.
- **CI is sometimes conflated with HPC, instead of being perceived as a superset that includes HPC.** At some institutions, HPC can dominate the research CI conversation. While many computational solutions involve advanced computing (either locally or at regional or national centers), a substantial fraction of computing- and data-intensive problems originate at the desktop or workgroup level. For institutional CI centers, modest to medium scale CI can be a crucial component of the overall computational environment needed for successful research. Despite this, however, at some institutions, HPC (and the people who visibly support it) is the primary or sole subject when discussing CI. This can lead to confusion and missed opportunities.
- **Bridging the gap between research and CI expertise is crucial.** Some organizational models for supporting institution-scale research computing include both researcher-facing and system-facing staff. Traditional IT teams may include system-facing professionals who, even if aware of available CI resources, can lack the breadth of knowledge, soft skills, and experience needed to effectively facilitate research. On the other hand, postdoctoral researchers, graduate students, undergraduates and others may be tasked with the scientific computing needs of research groups or laboratories and may be unaware of the full range of institutional CI, let alone of regional and/or national CI. Even those who are aware of such resources may lack the CI expertise to help researchers exploit them. Skilled CI Engineers and CI Facilitators can bridge the gaps between researchers and the on-campus organizations (both IT and non-IT) charged with supporting them.
- **Central IT organizations are not consistently recognized as CI experts.** At institutions where the central IT organization is focused principally on administrative and teaching needs rather than on research needs, IT staff may not have – or be recognized for having – CI expertise. This can be addressed by, for example, staffing a CI Facilitator position, or by IT staff being active in CI facilitation.
- **Multidisciplinary research centers may have a great deal to gain from ACI-REF/CI Engineer engagement.** Such centers may include large, diverse groups of faculty, staff, and students focused on broad research goals involving numerous partnerships, both on- and off-campus. They may span multiple departments, buildings, and campuses, and often encompass substantial laboratory and IT infrastructure and staff. Traditional IT organizations may find it challenging to engage with such centers, because IT staff rarely have the breadth of knowledge to appreciate the full scope, scale and diversity of computational needs or to communicate effectively about CI across the full spectrum. Working with central IT staff, however, CI Facilitators and CI Engineers can catalyze effective information exchange as well as development and deployment of appropriate solutions, leading to new opportunities for collaboration and funding.

1.1 National Need

In July 2015, the US White House released an executive order for the National Strategic Computing Initiative [4], under which the NSF was charged with, among other components, workforce development. Thus, the timing of the Virtual Residency's 2015 workshop (May 31 - June 6 2015) was fortuitous. Ultimately, a key question in workforce development is: How does the national and worldwide CI community quickly enable CI professionals who are responsible for rapidly making productive their institutions' computing- and data-intensive research? The Virtual Residency is designed to be a valuable component of the solution to this challenge.

The need for CI Facilitator training is acute. In the run-up to OU's submission of their NSF CC*IIE proposal, which is providing initial (2 year) funding for the Virtual Residency program, 33 institutions in 23 states and territories expressed interest in the Virtual Residency workshops (more than could realistically be funded on the CC*IIE budget), including 3 Minority Serving Institutions (MSIs), 19 institutions in 13 EPSCoR jurisdictions, and 7 non-doctoral institutions. Of the 49 institutions that submitted applications to participate in the 2015 workshop, 24 institutions were in 14 EPSCoR jurisdictions, 8 were MSIs and 6 were non-doctoral; of the 38 institutions that participated as the 2015 cohort, 19 institutions were in 12 EPSCoR jurisdictions, 5 were MSIs and 5 were non-doctoral. For the 2016 workshop, so far an additional another 36 institutions have applied, including institutions in 13 additional states and 3 other countries, among them 14 new institutions in 9 EPSCoR jurisdictions, 6 new Minority Serving Institutions, and 9 new non-doctoral institutions.

To date, as far as the authors are aware, there have been no other programs to develop the CI Facilitator workforce at the national scale.

1.2 Relationships to Other Efforts

1.2.1 Clemson-led ACI-REF Project

In 2014, a team of six institutions, led by Clemson University, were awarded an NSF Strategic Technologies for Cyberinfrastructure (STCI) grant focused on ACI-REFs (ACI-1341935, "Advanced Cyberinfrastructure - Research and Educational Facilitation: Campus-Based Computational Research Support," 3/1/2014-2/28/2017, $5,295,574) [5]. The original ACI-REF proposal included 13 institutions; the six that were funded on this first ACI-REF grant were designated "Phase 1," and the remaining seven, including OU, were designated "Phase 2," with an aspiration to fund the Phase 2 institutions later by other means. In the original ACI-REF proposal, OU's role included not only ACI-REFs, but also an EPSCoR-facing role and a Virtual Residency program. In fact, OU was the only institution to be all of (a) ACI-REF Phase 2, (b) in an EPSCoR jurisdiction, and (c) leading an active NSF Campus Cyberinfrastructure - Network Infrastructure and Engineering (CC-NIE) grant (ACI-1341028, "OneOklahoma Friction Free Network," 10/1/2013-9/30/2015 with a no cost extension to 9/30/2016, $499,961, PI H. Neeman [6]).

Of note in the context of the Virtual Residency program is that only one of the six ACI-REF Phase 1 institutions was represented in the 2015 Virtual Residency cohort. This demonstrates that the Virtual Residency program is both (a) designed to meet the needs of the national community as a whole, not primarily the Clemson-led ACI-REF project, and (b) already successful in impacting institutions nationwide.

1.2.2 XSEDE Campus Champions/Engagement

A significant factor in recruiting for the 2015 Virtual Residency cohort was the XSEDE Campus Champions [7] mailing list. Of the 38 institutions that participated in the 2015 workshop, 29 (76%) have Campus Champions (and of the 49 institutions that submitted one or more applications, 38, or 78%, have Champions). CC*IIE PI Neeman of OU and his counterpart at Oklahoma State University, co-author D. Brunson, have been appointed joint co-managers of the XSEDE Campus Engagement program, officially starting July 1 2016, which includes the Campus Champions.

1.2.3 Coalition for Academic Scientific Computation

Of the 38 institutions that participated in the 2015 workshop, 16 (42%) are members of the Coalition for Academic Scientific Computation (CASC) [8], well above the fraction of academic institutions with Carnegie Doctoral classification [9] that are CASC members: there are 82 CASC academic institution members, compared to all 329 academic institutions that have Carnegie Doctoral classifications (approximately 25% of Doctoral institutions are CASC academic members at this writing).

1.2.4 Linux Clusters Institute

CC*IIE PI Neeman is on the Linux Clusters Institute (LCI) [10] steering committee and hosted their most recent workshop.

1.2.5 International HPC Training Consortium

CC*IIE PI Neeman has served on the International HPC Training Consortium committee since it was founded in 2014 [11].

1.2.6 NSF Advisory Committee for Cyberinfrastructure

At this writing, CC*IIE PI Neeman is in his final year as a member of the NSF Advisory Committee for Cyberinfrastructure (ACCI) [12]; in 2016, he was appointed to lead ACCI's Working Group on Learning and Workforce Development.

1.2.7 Internet2 High Performance Research Computing Program Advisory Group

CC*IIE PI Neeman has been serving on the Internet2 High Performance Research Computing Program Advisory Group [13] since it was founded in 2014.

1.2.8 NSF Big Data Hubs Workforce Development

The NSF, via their Big Data Innovation Hubs (BD Hubs) program, has established four regional Big Data Hubs. Their mission is to form public-private partnerships among industry, academia, government, and nonprofits, in order to support the National Big Data Research and Development Initiative's charge to "solve some of the Nation's most pressing R&D challenges related to extracting knowledge and insights from large, complex collections of digital data." [14, 15] The BD Hubs solicitation explicitly directed proposers to incorporate workforce development as a key cross-cutting component within their organizational structures. Suggested workforce development efforts included the creation of new Data Science curricula and the hosting of events such as educational workshops and hackathon competitions [15]. To reinforce this role, an additional round of funding was provided to each BD Hub in April 2016 by the Computing Community Consortium, to promote early career education and opportunities [16], This funding was to be used to host workshops, hackathons and data-related competitions, to provide for the development of lecture series, and to allow for conference travel, short-term industry internships, and visits to national CI facilities for early career researchers [17, 18, 19, 20, 21, 22, 23].

2. VIRTUAL RESIDENCY WORKSHOP

2.1 2015 Participating Institution Details

2.1.1 Institution Types

Considering Carnegie classification [9], of the 38 institutions that participated in the 2015 Virtual Residency workshop, 29 (76%) were Doctoral institutions (19 Doctoral - Highest Research, 9 Doctoral - Higher Research, 1 Doctoral - Modest Research); 3 (8%) were Masters - Larger; 4 (11%) were Bachelors (3 Bachelors - Arts & Sciences, 1 Bachelors - Diverse Fields).

2.1.2 Institutional CI Resources

Of the 38 institutions that participated in the 2015 Virtual Residency workshop, 34 (89%) have institutional CI resources.

2.1.3 CI Grants

Of the 38 institutions that participated in the 2015 Virtual Residency workshop, 22 (89%) had CI grants (e.g., NSF Campus Cyberinfrastructure, NSF Major Research Instrumentation).

2.2 Onsite and Remote Attendance

Because of high demand, the 2015 Virtual Residency workshop was provided both live onsite and live via videoconferencing. OU has extensive experience with this approach, including: (1) OU's "Supercomputing in Plain English" workshops [24, 25, 26, 27, 28, 29]; (2) hosting Virtual School for Computational Science & Engineering [30] events 2012-15; (3) National Computational Science Institute/SC11 [31,32] summer workshops in 2011 held jointly via videoconferencing across timezones at (a) the University of Washington and Idaho State University [33], (b) Oklahoma State University and Washington & Lee University, and (c) OU and Polytechnic University of Puerto Rico [34], all co-coordinated by CC*IIE PI Neeman; (a) and (c) were also co-taught by OU. While the remote approach provides less total value than onsite, it is viable for those who cannot attend onsite. For the 2015 Virtual Residency workshop, 28 participated onsite and 22 live via videoconferencing.

2.3 2015 Workshop Agenda

The 2015 Virtual Residency workshop used (with permission) the structure the National Computational Science Institute's [31] weeklong workshops, several of which Neeman has co-taught.

2.3.1 Sunday

- Welcome and Virtual Residency Overview

This session introduced the Virtual Residency, including what an ACI-REF is, some background on the Clemson-led ACI-REF project, the NSF's CC*IIE program and the Campus CI Engineer subprogram, OU's Campus CI Engineer grant and its Virtual Residency component, information on what the Virtual Residency workshop would cover (and what it wouldn't cover, especially CDS&E technical content), and finally a focus on the strong national need for CI Facilitators and how the Virtual Residency addresses that need.

- Introduction to Research Cyberinfrastructure Consulting

This session described the distinction between "lore" (storytelling) vs "data" (technical content) and touched on OU's history of CI facilitation. It discussed the core goal and the typical components of CI facilitation (and what components typically aren't included).

- How to Give a CI Tour

This session, which took place at OU IT's primary data center, was both a CI tour and a meta-discussion of how to conduct such a tour.

2.3.2 Monday

- Early AM: Plenary: Effective Communication: How to Talk to Researchers about Their Research

The purpose of this session was to lay out, explicitly, how challenging it can be to communicate across discipline boundaries and career path boundaries. The session started with how researchers in each discipline use language in ways that are well understood within the discipline but opaque or confusing to those outside the discipline. The key example was, "Is Oxygen a Metal?" This segment focused on how astronomers, whose research encompasses (in some cases) the entire universe over its entire lifetime, view the components of visible matter in the universe as (1) hydrogen (approximately 75% of all baryonic mass in the universe [35]), (2) helium (approximately 24% of total elemental mass in the universe [36]) and (3) everything else combined – and, given that planets are typically made primarily of substances like iron, category (3) mostly consists of actual metals, so astronomers use the word "metals" to mean "neither hydrogen nor helium." While this is fundamentally different from how everyone else uses that word, in light of the information provided, the usage makes perfect sense in context – but would be opaque to anyone outside the discipline. Follow-on examples included: (a) consider the thought experiment of a conversation about "projection" among a mathematician, a psychologist and a movie producer; (b) consider a high energy physicist studying the smallest known particles, in which case, to what extent is gravity important? – but then consider cosmology. This session also covers issues such as: (i) effective things to say to researchers to get them interested in collaborating or in using institutional resources; (ii) how to find and recruit researchers at an institution to use institutional CI.

- Breakout: CDS&E Track
  - Mid AM: Deploying Community Codes

This session focused on the technical aspects of community code deployment.

  - Early PM: Real Users' CDS&E Research

An OU Computer Science faculty member, Prof. Amy McGovern, whose research focuses on data mining of meteorology data, presented not only on her research content but specifically on the CI aspects of that research.

- Breakout: Science DMZ Track
  - Mid AM: OpenFlow - Lecture

This session taught the concepts of a Science DMZ, CDS&E research, Software Defined Networking (SDN), and OpenFlow, all primarily by analogy; the session presented traditional industry networks as a vault holding private information such as intellectual property, and academic enterprise networks as a prison holding private information (for example, financial information). By contrast, CDS&E research was likened to a velociraptor: "agile, moves rapidly, more effective in groups, consumes all available resources." But releasing a velociraptor inside a prison (CDS&E research in an academic enterprise network) is likely to be non-ideal. A Science DMZ was depicted as an interstate: "no stoplights, and dedicated on and off ramps," and also like an onion, having multiple layers: the Science DMZ has "dedicated paths, transfer nodes, perfSONAR." SDN itself is like a body with multiple brains scattered throughout. It has "[a] controller, network switches, OpenFlow, northbound APIs, [and] east-west monitoring."

  - Early PM: OpenFlow - Lab

This was a hands-on opportunity to work with OpenFlow.

- Mid PM: Plenary: Cyberinfrastructure User Support

Originally designed by Mehmet (Memo) Belgin of Georgia Tech and used with his permission, this session focused on: (1) CI user expectations, categorization, and commonalities; (2) policies, politics, conflict, and personality management; (3) outreach and education; (4) lessons learned. It discussed CI vs enterprise IT; expectations of faculty vs students; needs, expectations, and approaches for novices vs intermediates vs advanced users.

2.3.3 Tuesday

- Very Early AM: Plenary: Project Guidelines

The projects were 5 minute presentations at the end of the week by each Virtual Resident (or team of Virtual Residents). This session gave an overview of expectations for content, structure and length.

- Early AM: Plenary: Faculty: Tenure, Promotion, Reward System (guest presenter Prof. Bruce Mason, OU)

The purpose of this session was to explain the thought processes of faculty, as a significant CI constituency. The session focused on tenure and promotion as a wellspring of faculty incentives that substantially shape faculty behavior and priorities. Included in the discussion were the various kinds of faculty (both tenure track and non-tenure track), the tenure process, faculty productivity and the reward system, the role of teaching, and abrogation of tenure.

- Breakout: CDS&E Track
  - Mid AM: Benchmarking & Tuning

This session included: the importance of readable code; resource utilization; optimization approaches; the storage hierarchy; determining system performance; profiling; performance counters.

  - Early PM: Real Users' CDS&E Research

A team of OU Civil Engineering & Environmental Science faculty members, Prof. Randy Kolar and Prof. Kendra Dresback, presented their research on coastal simulation via the finite element method using the ADCIRC package [37]; they are part of the national ADCIRC development team.

  - Mid PM: "Speed Dating:" Real Users' CDS&E Research – Practicing Intake Interviews

Originally, this session was intended simply to be additional presentations by OU researchers, but a last minute change created a fundamentally new capability. First, a presentation was provided (the tail end of Monday's "Effective Communication" plenary) on conducting intake interviews with researchers, and then the CDS&E Virtual Residents were split into three subgroups, to match the three research teams that were attending. Each research team was assigned to a subgroup of Virtual Residents for a period of only 15 minutes, during which an interview was conducted and feedback was provided to the research team. At the end of each 15 minute segment, the research teams rotated to the next subgroup of Virtual Residents. Thus, each subgroup of Virtual Residents had three opportunities to practice intake interviewing techniques, and each research team got feedback from all three subgroups.

- Breakout: Science DMZ Track
  - Mid AM: Exploring OpenDaylight - Lecture

This section focused on SDN practicalities, including: Why SDN?; SDN architecture characteristics; SDN overview; OpenDaylight; the layers of OpenDaylight; tools and paradigms; RESTful APIs.

  - Early PM: Exploring OpenDaylight - Lab
  - Mid PM: Real Users' High Bandwidth Research

This featured Dr. Horst Severini of OU High Energy Physics.

2.3.4 Wednesday

- Early AM: Plenary: Using Videoconferencing and Collaboration Technologies for Consulting (guest presenters J.E.B. Sheriff and S. Patrick Calhoun)

This session featured guest presentations by J.E.B. Sheriff, OU IT's conferencing lead, and S. Patrick Calhoun of OSCER. Content focused on the practicalities of using conferencing technologies to facilitate computing- and data-intensive research, with a core emphasis given to finding the right technology (or mix of technologies) to serve each researcher's needs.

- Mid AM: Plenary: Writing Grant Proposals (guest presenter Linda Mason, Oklahoma State Regents for Higher Education)

Some CI Facilitators and CI Engineers are called upon to participate in, or even lead, development of grant proposals to fund their institutions' CI resources. This session provided a basic introduction to grant proposal writing. Topics included: types of grants, locating grants, NSF CI grant programs, planning a grant, writing the proposal, proposal review, and managing a grant.

2.3.5 Thursday

- Early AM: The Shifting Landscape of CI Funding Opportunities (guest presenter Dr. J. Barr von Oehsen, then at Clemson University)

The purpose of this session was to introduce CI funding opportunities that would directly serve the needs of researchers. focusing on: CI programs like NSF Campus Cyberinfrastructure, NSF Major Research Instrumentation (MRI), NSF CISE Research Infrastructure (CRI), DOD Defense University Research Instrumentation Program (DURIP), NIH Shared Instrumentation Grant; campus CI plans; history and evolution of CI programs; recent NSF national-scale CI resource grants (Jetstream, Comet, Louisiana State University's SuperMIC, Stanford University's X-GPU, CloudLab) and the Clemson-led ACI-REF Phase 1 grant.

- CDS&E Track
  - Mid AM: Finding and Provisioning Remote Resources (guest presenters from XSEDE and Open Science Grid)

The purpose of this session was to introduce external resources that researchers can exploit. The session focused on national resources such as XSEDE (represented by Jeff Pummill) and the Open Science Grid (represented by Elizabeth Prout and Rob Gardner). OSG topics included: the OSG; high throughput computing; submitting and tracking jobs on OSG; OSG storage; OSG training and the OSG User School; Software Carpentry.

  - Early PM: "Speed Dating:" Real Users' CDS&E Research – Practicing Intake Interviews

See Tuesday.

  - Mid PM: CDS&E Catch-up

Several sessions ran long and were unable to complete all materials, so this timeslot was used to finish some of the leftover topics.

- Science DMZ Track
  - Mid AM: The Software in SDN - Lecture
  - Early PM: The Software in SDN - Lab
  - Mid PM: Real Users' High Bandwidth Research

This session featured Gerry Creager of OU's Cooperative Institute for Mesoscale Meteorology Studies.

2.3.6 Friday

- Early AM: Plenary: So You Want to Write a CI Proposal

This session focused on the practicalities of writing a grant proposal for CI, with primary emphasis on the NSF's Major Research Instrumentation (MRI) and Campus Cyberinfrastructure programs. The session was structured around components of an MRI proposal as a launch point for discussing practical aspects of developing such proposals. Specific topics included: vision; domain STEM research projects, including core issues to be addressed (current, pending and planned funding; number of faculty, staff, postdocs, graduate students and undergraduates on each domain STEM research project who will use the resource to be acquired; the innovative/transformative aspects of the research; broader impacts); amount of resource consumption expected (and how that was calculated); for computing proposals, whether each project's software has been benchmarked and/or optimized for the target platform; results from prior NSF support; instrument description (rationale, purchase process, current similar instruments, instrument role); research impact; broader impacts (integration of research and education; underrepresented populations; dissemination; technology transfer); management plan (facility; operations labor; apportioning; decision making procedure; advisory committees; timeline and milestones; sustainability plan; cost share vs institutional commitment; CI plan).

- Mid AM: Plenary Panel: Stories from the Trenches

This was an informal discussion about practical experiences.

- Early-Mid PM: Project preparation time
- Late PM: Project presentations from early departers

2.3.7 Saturday

Saturday was spent on Virtual Resident project presentations exclusively.

3. HYPOTHETICAL SCENARIOS

One of the goals of the Virtual Residency program is that Virtual Residents can apply their experiences in the Virtual Residency to their CI facilitation roles at their home institutions. The following are hypothetical example scenarios of applying the concepts that would be encountered in the Virtual Residency program to effect concrete change at a Virtual Resident's home institution.

3.1 Building Campus Relationships

Following best practice guidance provided at the workshop, a Virtual Resident returns to their home institution and begins searching departmental websites for researchers who may have an interest in, or a need for, that institution's CI resources. Using this method, the Virtual Resident builds an e-mail list of CI research stakeholders, which can be used as a resource for composing proposals and/or publications, as well as for generating reports. Additionally, each week, the Virtual Resident uses this information to contact a handful of researchers to initiate individual intake and fact-finding meetings. Via this method, the Virtual Resident becomes familiar with the CI needs of their home institution's research community, and builds a group of campus CI advocates who will be helpful in making the case for expansion of CI capabilities.

3.2 Consulting

A Virtual Resident holds regular walk-in consulting hours in an easily accessible location on campus. A faculty investigator communicates an urgent need to access a Hadoop system so that they can process their large, unwieldy dataset in time for a publication deadline. Based on experience with the speed dating exercise, the Virtual Resident works to understand the needs of the faculty member at a more fundamental level, knowing that sometimes researchers will hear of a CI technique or technology that sounds promising, but they may not be familiar with the suite of options or the details of using the technique they have identified. By asking questions to better understand the nature of the researcher's workflow and where the key challenges lie, the Virtual Resident determines that Hadoop is not the most appropriate solution. In this case, the consultant recommends simpler scripting and Unix text parsing tools, which would be more flexible with respect to evolving needs, and would allow using a larger and better supported batch HPC system, instead of a small experimental Hadoop system. The consultant follows up with the investigator's graduate students, sharing ideas and examples to help advance towards implementing this approach. The students successfully implement the solution, and the Virtual Resident checks in with the research team periodically to ensure their needs continue to be met.

3.3 Identifying Resources

A researcher speaks with a colleague about an exciting new project, but also shares their frustration with long wait times on the campus condominium cluster and frequent notices from the administrators that they are exceeding their storage quota. Their colleague recommends speaking with the campus CI Facilitator that the colleague has been working with on another project, and makes an introduction. The CI Facilitator determines that the condominium cluster will not – and indeed no existing campus resource can – meet the needs of this project over the long term. Believing that this project is indicative of future needs across a large user community, the CI Facilitator initiates discussions on campus to submit a MRI proposal to the NSF to better support both processing and long term, quick-access data. A coalition of support on campus and example use cases are readily available, thanks to the relationships built between the CI Facilitator, researchers, and administrators.

4. FUTURE WORK

Planning for the 2016 Virtual Residency workshop, scheduled for August 7-13 2016, is well underway at this writing, and applications are already being received. The 2016 workshop will feature an overlap of three cohorts: the original 2015 Virtual Residency workshop cohort, ACI-REFs from the Clemson-led NSF ACI-REF Phase 1 project, and a new 2016 cohort. The 2015 and Phase 1 cohorts are designing the curriculum and will lead and teach parts of it. At this writing, 72 applications have been received from 55 institutions (16 from the 2015 cohort, of which one is also from a Phase 1 institution; 5 from 3 additional Phase 1 institutions in 3 states; 51 in the new 2016 cohort). In addition to the 2015 cohort, another 36 institutions have applied, including institutions in 13 additional states and 3 other countries, among them 14 new institutions in 9 EPSCoR jurisdictions, 6 new Minority Serving Institutions, and 9 new non-doctoral institutions.

A significant challenge is how to fund the Virtual Residency after the 2016 workshop. The team is pursuing all of (a) additional external funds via a standalone grant proposal; (b) additional external funds as part of a follow-on ACI-REF grant; (c) CI vendor sponsorships. Ultimately, the Virtual Residency will need to be self-sustaining, but the goal is to ramp down external funding, instead of a sudden cutoff.

The team has also considered the option of charging a registration fee. While this may prove to be necessary in the future, it would be preferable not to have to charge such a fee, especially to attendees from minority serving, non-doctoral and EPSCoR institutions.

5. ACKNOWLEDGMENTS

Portions of this material are based upon work supported by the National Science Foundation under the following grant: Grant No. ACI-1440783, "A Model for Advanced Cyberinfrastructure Research and Education Facilitators." Portions of the text of this paper were originally developed for various grant proposals and grant reports, and are used with permission of the Principal Investigators, including some of this paper's authors. The Cyberinfrastructure Support session was originally designed by Mehmet (Memo) Belgin of Georgia Tech and was used with his permission. Several of the sessions were presented by guest lecturers (described above), to whom the authors are deeply grateful.

6. REFERENCES

[1] University of Oklahoma Advanced Cyberinfrastructure Research and Education Facilitators Virtual Residency 2015 workshop webpage. http://www.oscer.ou.edu/acirefvirtres2015/

[2] Committee on Future Directions for NSF Advanced Computing Infrastructure to Support U.S. Science in 2017-2020; Computer Science and Telecommunications Board; Division on Engineering and Physical Sciences; National Academies of Sciences, Engineering, and Medicine, 2016: *Future Directions for NSF Advanced Computing Infrastructure to Support U.S. Science and Engineering in 2017-2020*. The National Academies Press. DOI: http://dx.doi.org/10.17226/21886.

[3] National Science Foundation, Advisory Committee for Cyberinfrastructure, 2011: *Task Force on Cyberlearning and Workforce Development Final Report,* https://www.nsf.gov/cise/aci/taskforces/FrontCyberLearning.pdf.

[4] US White House, 2015: "Executive Order -- Creating a National Strategic Computing Initiative." https://www.whitehouse.gov/the-press-office/2015/07/29/executive-order-creating-national-strategic-computing-initiative

[5] Advanced Cyberinfrastructure - Research and Education Facilitators website. https://aciref.rc.fas.harvard.edu/

[6] H. Neeman, D. Akin, J. Alexander, D. Brunson, S. P. Calhoun, J. Deaton, F. Fondjo Fotou, B. George, D. Gentis, Z. Gray, E. Huebsch, G. Louthan, M. Runion, J. Snow and B. Zimmerman, 2014: "The OneOklahoma Friction Free Network: Towards a Multi-Institutional Science DMZ in an EPSCoR State." *Proceedings of the 2014 Annual Conference on Extreme Science and Engineering Discovery Environment* (XSEDE '14). ACM, article 49. DOI: 10.1145/2616498.2616542.

[7] XSEDE Campus Champions webpage. https://www.xsede.org/campus-champions

[8] Coalition for Academic Scientific Computation website. http://www.casc.org/

[9] Carnegie Classifications website. http://carnegieclassifications.iu.edu/

[10] Linux Clusters Institute website. http://www.linuxclustersinstitute.org/

[11] International HPC Training Consortium. https://sites.google.com/a/lbl.gov/hpc-training-best-practices/home

[12] National Science Foundation Advisory Committee for Cyberinfrastructure webpage. https://www.nsf.gov/cise/aci/advisory.jsp

[13] Internet2 High Performance Research Computing Program Advisory Group press release, 2014. http://www.internet2.edu/news/detail/6039/

[14] National Big Data Research Initiative - https://www.whitehouse.gov/sites/default/files/microsites/ostp/big_data_press_release_final_2.pdf

[15] NSF BDHubs Solicitation. http://www.nsf.gov/pubs/2015/nsf15562/nsf15562.htm

[16] CCC Big Data Hubs Announcement. http://www.cccblog.org/2016/04/11/ccc-announces-industry-academic-program-with-the-big-data-regional-hubs/

[17] Computing Community Consortium – Big Data Regional Hubs Industry-Academic Collaboration. http://cra.org/ccc/leadership-development/big-data-regional-hubs/

[18] Midwest Big Data Hub – Seed Funding for Early Career Researchers. http://midwestbigdatahub.org/funding/call-for-seed-funding-proposals/

[19] Midwest Big Data Hub - UND Early Career Big Data Summit. http://library.und.edu/digital/bigdatasummit/

[20] Northeast Big Data Hub – Catalyzing Industry-Academic Partnerships. http://northeastbdhub.dsi.columbia.edu/CCCapplication.htm

[21] South Big Data Hub – Data Start: The Southern Startup Internship Program in Data Science. http://www.southbdhub.org/datastart.html

[22] South Big Data Hub – Program to Empower Partnerships with Industry (PEPI). http://www.southbdhub.org/pepi.html

[23] West Big Data Hub – Collaboratory Faire v1.0, Workshop on Data Hackathon Best Practices, Tools of the (Data Journalism) Trade: Workshop on Crafting a Compelling Data Science Narrative. http://westbigdatahub.org/2016/04/15/registration-now-open-for-wbdih-all-hands-may-12/

[24] H. Neeman, H. Severini, D. Wu and K. Kantardjieff, 2010: "Teaching High Performance Computing via Videoconferencing." *ACM Inroads*, 1(1), March 2010, pp. 67-71.

[25] H. Neeman, H. Severini and D. Wu, 2008: "Supercomputing in Plain English: Teaching Cyberinfrastructure to Computing Novices." *inroads: SIGCSE Bulletin*, 40(2), 27-30, 2008.

[26] H. Neeman, H. Severini, D. Wu and K. Kantardjieff, 2008: "Teaching Supercomputing via Videoconferencing." *Proceedings of TeraGrid 2008*.

[27] H. Neeman, L. Lee, J. Mullen, G. Newman, 2006: "Analogies for Teaching Parallel Computing to Inexperienced Programmers." *inroads: SIGCSE Bulletin*, 38 (4), 64-67.

[28] H. Neeman, J. Mullen, L. Lee and G. Newman, "Supercomputing in Plain English: Teaching High Performance Computing to Inexperienced Programmers." *Proceedings of the Third LCI International Conference on Linux Clusters: The HPC Revolution 2002*.

[29] OSCER education webpage. http://www.oscer.ou.edu/education/

[30] Virtual School for Computational Science and Engineering website. http://www.vscse.org/

[31] National Computational Science Institute website. http://computationalscience.org/

[32] SC11 conference website. http://sc11.supercomputing.org/

[33] National Computational Science Institute/SC11 Introduction to Parallel Programming and Cluster Computing webpage. http://www.oscer.ou.edu/ncsiworkshop2011intropar_logistics.php

[34] National Computational Science Institute/SC11 Intermediate Parallel Programming and Cluster Computing webpage. http://www.oscer.ou.edu/ncsiworkshop2011intmdpar_logistics.php

[35] Wikipedia page on Hydrogen. http://en.wikipedia.org/wiki/Hydrogen

[36] Wikipedia page on Helium. http://en.wikipedia.org/wiki/Hydrogen

[37] ADCIRC website. http://adcirc.org/