



# The Rapidly Evolving Research Computing Professional Workforce

**Henry Neeman**

**Director, OU Supercomputing Center for Education & Research**

**Associate Professor, Gallogly College of Engineering**

**Adjunct Associate Professor, School of Computer Science**

**University of Oklahoma**

**NITRD HEC IWG, Thursday August 20 2020**



**INFORMATION TECHNOLOGY**  
*The UNIVERSITY of OKLAHOMA*

# Program Notes

- My career has been in academia, so that'll color how I talk about this stuff.
- I have limited experience with government and non-profit research institutes, and even less with industry.
  - But I'm happy to learn!
- I love questions!
  - I'd rather you interrupt me and I run out of time before I run out of slides, than I drone on and on while your eyes glaze over ....



INFORMATION TECHNOLOGY  
The UNIVERSITY of OKLAHOMA

Research Comp Workforce  
NITRD HEC IWG, Thu Aug 20 2020





# Cyberinfrastructure Professionals

Research Comp Workforce  
NITRD HEC IWG, Thu Aug 20 2020

# What is a CI Professional?

- Cyberinfrastructure (CI) professionals are producers of CI, not consumers of CI.
  - Some CI professionals do both, as different aspects of their position.



Research Comp Workforce  
NITRD HEC IWG, Thu Aug 20 2020



# CI Professional Roles: The Facings

- Researcher-facing (e.g., CI Facilitators)
- Systems-facing (e.g., sysadmin, network engineer)
- Sponsor/Stakeholder-facing (e.g., HPC center director)
- Software-facing (e.g., Research Software Engineer)
- Data-facing (e.g., Research Data Librarian)



INFORMATION TECHNOLOGY  
The UNIVERSITY of OKLAHOMA

Research Comp Workforce  
NITRD HEC IWG, Thu Aug 20 2020



# What is a Cyberinfrastructure Facilitator?

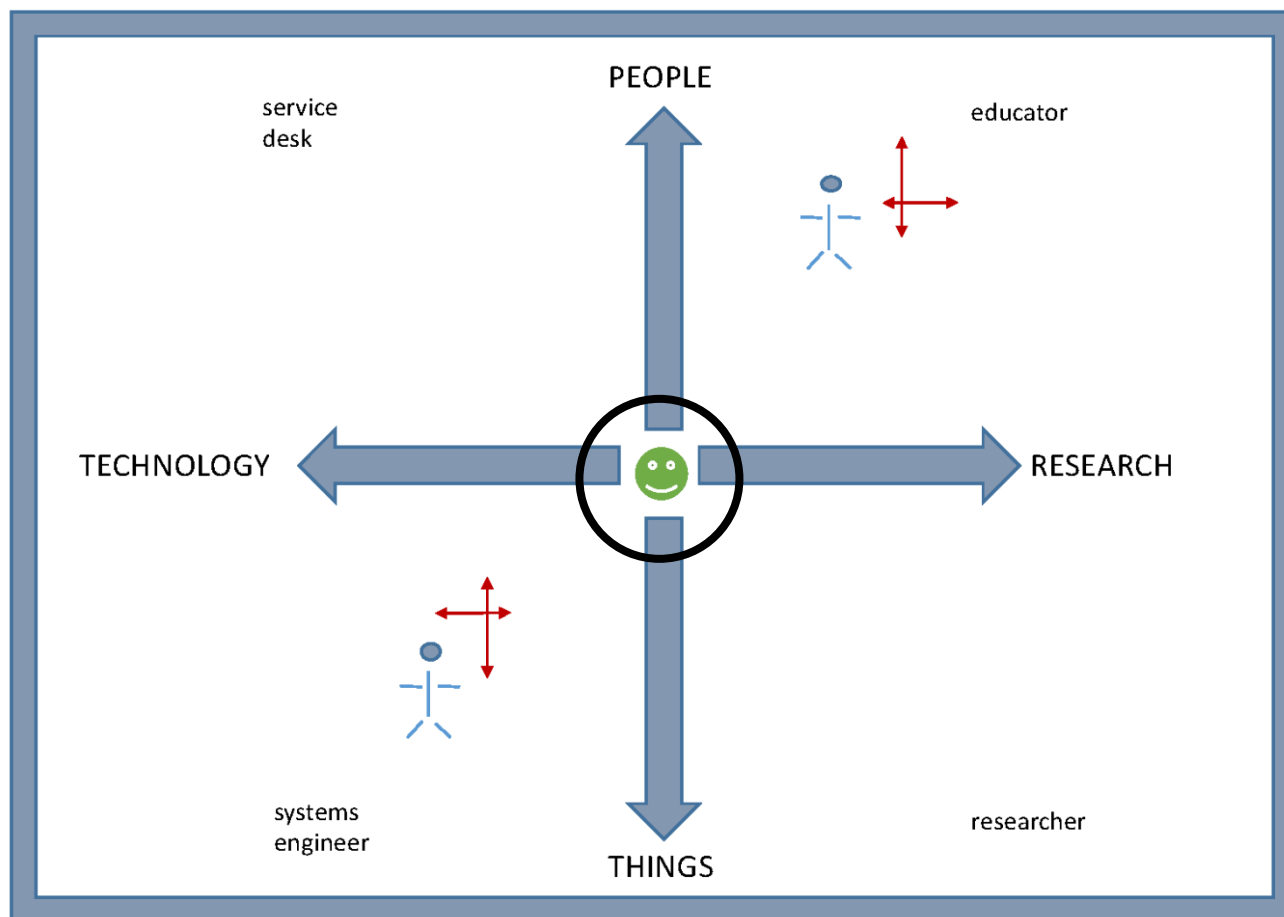
- Work with users – researchers and educators – to help them improve their research and/or education productivity and aspirations via advanced Cyberinfrastructure (CI).
- Typically, one or a few CI Facilitators have responsibility for an entire institution, or even multiple institutions.
- At some institutions, CI Facilitation is part time; at others, it's full time.
- Some CI Facilitators are:
  - faculty or former faculty;
  - postdocs or former postdocs;
  - research staff or former research staff;
  - IT professionals;
  - graduate or undergraduate students.



Research Comp Workforce  
NITRD HEC IWG, Thu Aug 20 2020



# CI Facilitators: What Qualities?



Neeman/Cuff 2016

# CI Facilitator Groups

- Campus Champions (started by TeraGrid, continued by XSEDE)
- Advanced Cyberinfrastructure Research & Education Facilitators (ACI-REF): completed – led by Clemson U
- Virtual Residency (training)
- Campus Research Computing Consortium (CaRCC): follow-on to Clemson-led ACI-REF project
  - Also includes the other “facings”



Research Comp Workforce  
NITRD HEC IWG, Thu Aug 20 2020





# Campus Champions

A national community of practice,  
facilitating computing-intensive and  
data-intensive research and education



Every US state plus  
4 US territories;  
every EPSCoR  
jurisdiction



<http://www.xsede.org/community-engagement/campus-champions>



# Campus Champions

Our community of over 700 Campus Champions, at over 300 institutions, **promotes and facilitates the effective participation** of a diverse national community of academic and not-for-profit institutions in the **application of advanced digital resources** and services to accelerate discovery, enhance education, and foster scholarly achievement.



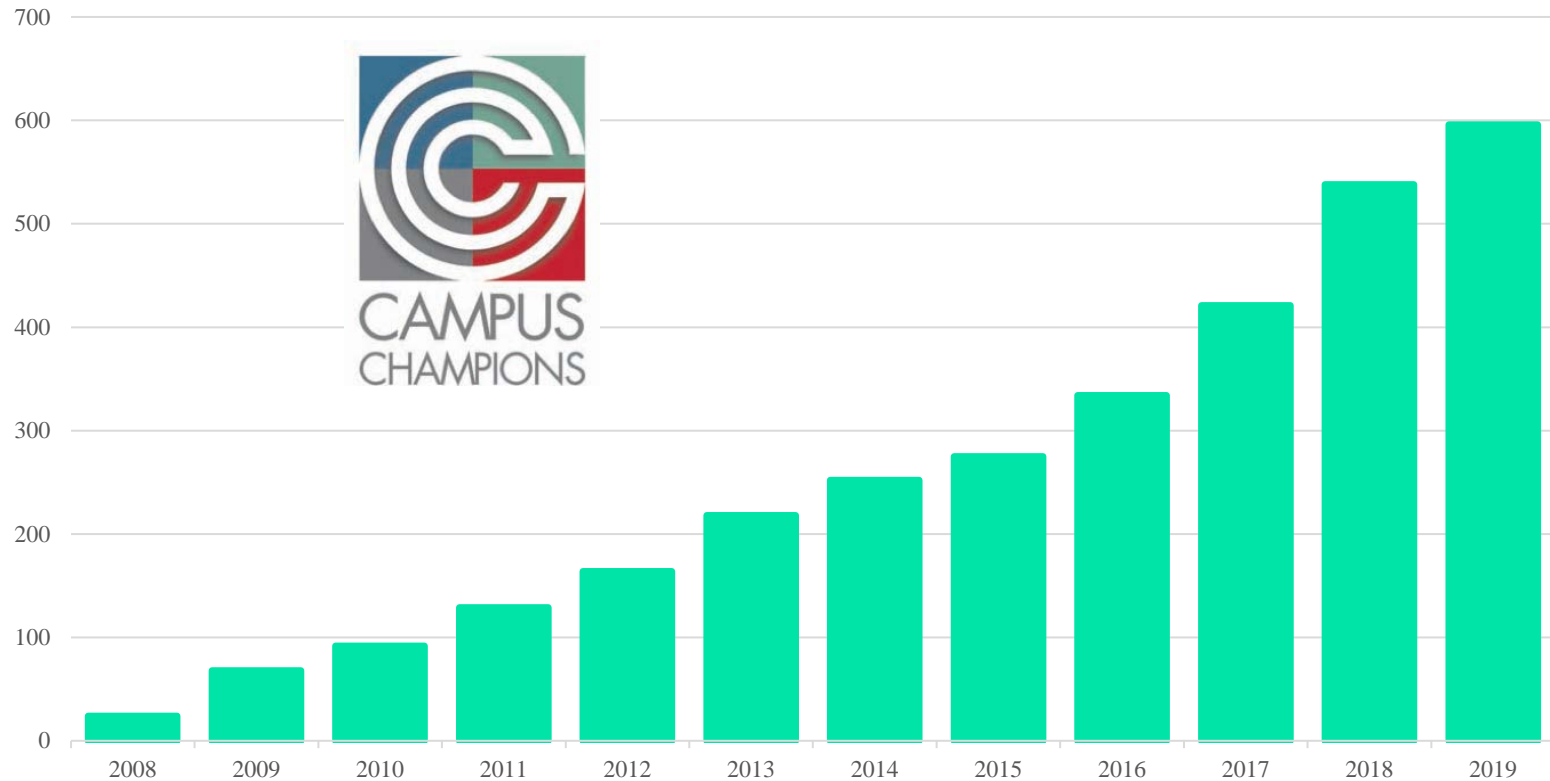
The **700+** Campus Champions – include CI organization leaders, faculty, researchers, students, as well as research-enabling professionals and systems professionals.



<http://www.xsede.org/community-engagement/campus-champions/>



# An Emerging Profession



# Virtual Residency



- We created a program to teach people how to be research computing facilitators, and ultimately to be institutional CI leaders.
- Workshops: Introductory 2015, 2016, 2017; Introductory/Intermediate 2019, Intermediate/Advanced 2018, 2020
  - Planning for the 2021 workshop starts soon
- Regular conference calls
- Grant Proposal Writing Apprenticeship (2017-18/18-19/19-20)
- Paper Writing Apprenticeship (2018-19/19-20: PEARC'19, PEARC'20 papers)
- Funded through a variety of NSF grants: CC\*, CaRCC RCN, XSEDE, CI Leadership Academy (workshop grant)
- Attendees: 924 from 370 institutions (74% of CC institutions)

<http://www.oscer.ou.edu/virtualresidency/>



INFORMATION TECHNOLOGY  
The UNIVERSITY of OKLAHOMA

Research Comp Workforce  
NITRD HEC IWG, Thu Aug 20 2020



# Campus Research Computing Consortium

An organization of dedicated CI professionals developing, advocating for, and advancing campus CI and associated professions.

Current focus areas include:

- People Network – year-round virtual conference
- Connecting the broader CI ecosystem
- Professionalization and workforce development
- Developing a common Capabilities Model for Research Computing and Data
- <https://carcc.org/>



INFORMATION TECHNOLOGY  
The UNIVERSITY of OKLAHOMA

Research Comp Workforce  
NITRD HEC IWG, Thu Aug 20 2020



# CaRCC People Network Tracks

- **Research-facing**: Outreach, education/training, consulting/facilitation, leasing collaborations, etc.
- **Systems-facing**: Systems planning, engineering, security, optimization, middleware, virtualization, and cloud, etc.
- **Data-facing**: Data management, publishing/sharing, data science, data visualization, data workflows, data transfer and networks, etc.
- **Emerging Centers**: Support for smaller or developing research computing and data centers.
- **Software-Facing**: Software development, portability, installation, optimization, support, etc. (Launching at a future date)
- **Sponsor/Stakeholder-facing**: Project and personnel management, service evaluation, funding/finance, and running a research computing and data team, etc. (TBD in collaboration with CASC.)





# The CI Professional Ecosystem

- Campus Champions
- Campus Research Computing Consortium (CaRCC)
- Coalition for Academic Scientific Computation
- CyberAmbassadors
- Cyberinfrastructure Engineers
- EDUCAUSE
- Linux Clusters Institute
- Science Gateways Community Institute
- SIGHPC Education Chapter
- Software & Data Carpentry (The Carpentries)
- Trusted Cyberinfrastructure
- United Kingdom Research Software Engineer Association
- United States Research Software Engineer Association
- United States Research Software Sustainability Institute

This list is always **incomplete!**



Research Comp Workforce  
NITRD HEC IWG, Thu Aug 20 2020





# Cyberinfrastructure Facilitators



INFORMATION TECHNOLOGY  
*The UNIVERSITY of OKLAHOMA*





# More Institutions Have On-Premise CI

Fraction of national universities (US News rankings) that have **central on-premises CI resources**:

- 49 of the Top 50 (98%) of the US News rankings;
- 95 of the Top 100 (95%);
- 132 of the Top 150 (88%);
- 159 of the Top 200 (80%).

Fraction of R1 and R2 institutions:

- 130 of 131 R1 (99%);
- 82 of 135 R2 (61%).

R1: “Doctoral Universities: Very High Research Activity”

R2: “Doctoral Universities: High Research Activity”

<http://carnegieclassifications.iu.edu/>

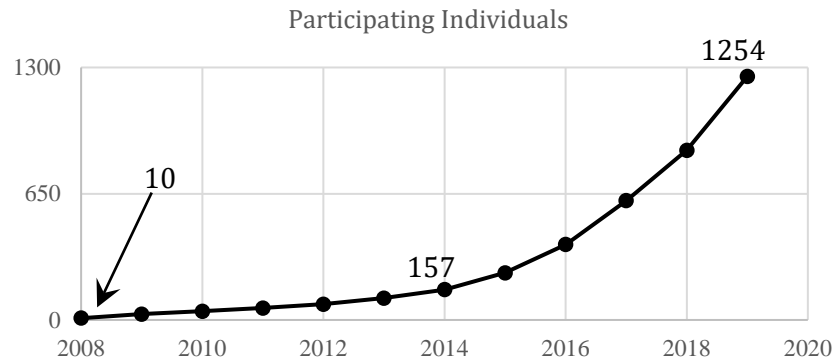
<https://www.usnews.com/best-colleges/rankings/national-universities>



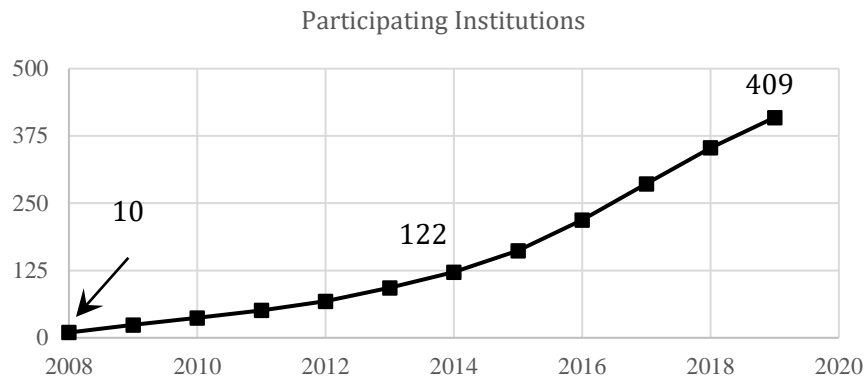
Research Comp Workforce  
NITRD HEC IWG, Thu Aug 20 2020



# More CI Facilitators, More Institutions



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



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

# Why is Helping Researchers Hard?

- **Ubiquity**: Within any discipline, a greater proportion of researchers do computing-intensive and/or data-intensive research.
- **Applicability**: More disciplines do computing-intensive and/or data-intensive research.
- **System Complexity**: The storage hierarchy is getting deeper, and parallelism is getting more hybrid.
- **Conceptual Distance**: The mental gap from handheld computing to command line/Linux/batch/remote/shared.

But we still only have one hour before they lose interest!



INFORMATION TECHNOLOGY  
The UNIVERSITY of OKLAHOMA

Research Comp Workforce  
NITRD HEC IWG, Thu Aug 20 2020





# Training for Cyberinfrastructure Professionals



INFORMATION TECHNOLOGY  
The UNIVERSITY of OKLAHOMA



# How Are CI Professionals Trained?

- Researcher-facing (e.g., CI Facilitators): Virtual Residency
- Systems-facing (e.g., sysadmin, network engineer):  
Linux Clusters Institute
- Sponsor/Stakeholder-facing (e.g., HPC center director):  
CI Leadership Academy (Feb 201)
- Software-facing (e.g., Research Software Engineer):  
Typically, a degree in a related domain discipline, plus ???
- Data-facing (e.g., Research Data Librarian):  
Library Science plus ???



# Linux Clusters Institute

- Teach HPC system administration
- 2-3 workshops per year (intro, intmd, adv)
- Designed and taught by HPC sysadmins
- Mix of lectures and hands-on labs
- Intro: Target audience is Linux sysadmins with zero HPC
- Intmd: Assumes Intro or equivalent background



Research Comp Workforce  
NITRD HEC IWG, Thu Aug 20 2020



# Example LCI Topics

- Intro to HPC
- Intro to MPI and OpenMP
- Designing a Cluster
- Storage (parallel filesystems, e.g., Lustre, Ceph, ZFS, Spectrum Scale/GPFS)
- Cluster Software Stack
- Configuration Mgmt
- Scheduling and Resource Mgmt (e.g., Slurm, PBS)
- Monitoring
- HPC User Support
- Networking
- Account Mgmt
- Modules
- User Software Mgmt
- Limiting Resources: Cgroups, Containers, Processor Affinity
- Node Health Check
- Science DMZ Design
- Infiniband
- OpenStack

# CI Leadership Academy

- NSF workshop grant
- Held in Feb 2019
- Senior/national CI leaders as mentors for institutional/emerging CI leaders
- Perspectives panels by:
  - (Former) funding agency representatives
  - (Former) industry representatives
  - (Current) computing/data-intensive researchers
- Lots of discussion, storytelling, idea swapping



Research Comp Workforce  
NITRD HEC IWG, Thu Aug 20 2020







# Virtual Residency Program

<http://www.oscer.ou.edu/virtualresidency/>



INFORMATION TECHNOLOGY  
The UNIVERSITY of OKLAHOMA



CaRCC  
Center for Advanced Research in Computing & Communications

# Virtual Residency Program: What?

We teach pre-service and in-service Cyberinfrastructure Facilitators how to do (or do better) Cyberinfrastructure Facilitation.



INFORMATION TECHNOLOGY  
The UNIVERSITY of OKLAHOMA

Research Comp Workforce  
NITRD HEC IWG, Thu Aug 20 2020



# Virtual Residency Program: How?

- Annual weeklong summer workshop (2015-2020)
  - 2019: Sun June 2 - Fri June 7
  - U California System has run its own targeted workshop based on our introductory workshop, in 2017, 2018 and 2019.
- Workshop planning videoconference calls
- Annual meeting at the SC supercomputing conference
- 2017-18, 18-19, 19-20: Grant Proposal Writing Apprenticeship
- 2018-19, 19-20: Paper Writing Apprenticeship

Before the Virtual Residency Program (VRP),  
**no one had ever been dumb enough to try to teach this stuff.**



Research Comp Workforce  
NITRD HEC IWG, Thu Aug 20 2020



# Virtual Residency Program: Why?

- CI Facilitators have strong experience within their discipline (often non-CS).
- Most CI Facilitators and CI Engineers haven't been faculty.
- Sometimes little or no research experience (especially for IT staff who have an enterprise IT background).
- Even if strong research background, typically little or no experience with research outside their own discipline.
- When we started the Virtual Residency in 2015, there were no local, regional or national programs to teach people how to be a Cyberinfrastructure Facilitator.
- In the olden days, you could take your time learning how to do this – but not anymore ....



# Virtual Residency: Who?

2015-present: 924 people from 370 institutions in 50 US states & 3 US territories, plus 11 other countries:

- 56 institutions (15%) are Minority Serving Institutions;
- 94 institutions (25%) are non-PhD-granting institutions;
- 101 institutions (27%) in 27 of 28 (96%) EPSCoR jurisdictions;
- 241 institutions (65%) are Campus Champion institutions (74% of Campus Champions' 327 institutions).
- 188 institutions (51%) are Campus Cyberinfrastructure Consortium institutions (83% of CaRCC's 226 institutions).

This is for **ALL** Virtual Residency activities, including:

- workshops (including mini-workshops by/for U California);
- workshop planning calls;
- the Grant Proposal Writing Apprenticeship;
- the Paper Writing Apprenticeship.



Research Comp Workforce  
NITRD HEC IWG, Thu Aug 20 2020



# Virtual Residency: Who? (cont'd)

2015-present: 924 people from 370 institutions  
in 50 US states & 3 US territories, plus 11 other countries :

## US News Rankings (Nat'l Univs)

- all of the Top 10 (100%);
- 23 of the Top 25 (88%);
- 45 of the Top 50 (85%);
- 65 of the Top 75 (86%);
- 83 of the Top 100 (81%);
- 116 of the Top 150 (76%);
- 137 of the Top 200 (68%).

<https://www.usnews.com/best-colleges/rankings/national-universities>

## Carnegie Classifications

- 121 of 131 R1s (92%);
- 54 of 82 R2s (66%)  
that **DO** have on-premise CI;
- 21 of 53 R2s (40%)  
that **DON'T** have on-premise CI.

<http://carnegieclassifications.iu.edu/>



INFORMATION TECHNOLOGY  
The UNIVERSITY of OKLAHOMA

Research Comp Workforce  
NITRD HEC IWG, Thu Aug 20 2020



# What Do We Cover?

- How to work with researchers who are using CI.
  - How to talk to them.
  - How to help them.
- How to contribute to, and ultimately lead, CI initiatives at your institution.
  - Some already us knew how to do this, so our job was to help the rest.
- Computational Science & Engineering Track
  - Get some practice working with researchers.
- Science DMZ Track (2015-16)
  - How to manage a Science DMZ.
- **Introductory**: Foundational information about researchers (especially faculty); developing skills to work with them (to jumpstart CI Facilitators).
- **Intermediate**: Build on Intro, to include more in-depth understanding to support researchers and their activities.
- **Advanced**: Institutional CI leadership.



# What Aren't We Trying to Do?

- We **AREN'T** trying to cover a lot of ~~technical content~~.
  - People can learn that from other sources.
- Instead, the goal is to teach the **PROFESSION** of CI facilitation.



Research Comp Workforce  
NITRD HEC IWG, Thu Aug 20 2020





# What's Our Hidden Agenda?

- The real goal is to prepare for an upcoming transition to:
  - more need for this kind of skilled workforce, but
  - fewer people who know how to do it, with
  - no mechanism to prepare a sufficiently large cohort.
- Some of the participants already knew how to do this.
  - But it took a very long time to learn on their own.
  - To keep up with demand, the community needs us to streamline the process so that new facilitators can become fully productive quickly.
- These are the CI leaders of tomorrow.
  - We need to get to scalability!



Research Comp Workforce  
NITRD HEC IWG, Thu Aug 20 2020



# Virtual Residency Workshops

- Introductory/Intermediate/Advanced
- One week each summer (2015-20, 2021 about to be planned)
- Onsite in person, remote via videoconferencing, and asynchronously via video recordings
- Curriculum designed by the community
- Introductory topics as talks and activities; intermediate and advanced topics are mostly panels.
- <http://www.oscer.ou.edu/virtualresidency2019/>



INFORMATION TECHNOLOGY  
The UNIVERSITY of OKLAHOMA

Research Comp Workforce  
NITRD HEC IWG, Thu Aug 20 2020



# Intro/Intmd Workshop Themes

- How to Understand and Work with Real Researchers
- Technical Content
- The Cyberinfrastructure Milieu
- Grant Proposal Writing
  - Shifted to the Grant Proposal Writing Apprenticeship.



INFORMATION TECHNOLOGY  
The UNIVERSITY of OKLAHOMA

Research Comp Workforce  
NITRD HEC IWG, Thu Aug 20 2020



# Intmd/Adv Workshop Themes 2018/2020

- CI Leadership
  - Perspectives about CI from CIOs & VPRs\*
- CI Outreach
- CI Expertise
- CI Communication
  - Working Effectively with Vendors\*
- CI Budgeting
  - The CI Funding Landscape: Funding Agency Perspectives\*

\* CI Leadership Academy session (2020)



Research Comp Workforce  
NITRD HEC IWG, Thu Aug 20 2020



# Virtual Residency Conf Calls

- In 2015, the Virtual Residency workshop was designed by OU IT, most of the speakers were from OU IT, and the rest were invited by OU IT.
- The Virtual Residency conference calls were originally meant to cover additional content throughout the fall and spring.
- But, with the rise of the Campus Cyberinfrastructure Consortium (CaRCC) and its researcher-facing calls, that functionality was no longer needed in the Virtual Residency Program (VRP).
- Starting with the 2016 Virtual Residency workshop, we used the Virtual Residency conference calls to design each year's workshop as a community, and to recruit speakers.
  - Starting in 2016, most of the speakers have been either Virtual Residents themselves, or recommended by Virtual Residents.



# How Are the Sessions Picked?

- In the Virtual Residency (workshop planning) conference calls, we work together on developing the curriculum and identifying speakers/panelists/moderators.
- Example: For the 2020 intermediate/advanced workshop:
  - Survey #1: Prioritize sessions from 2018 intermediate/advanced workshop, plus a box to suggest new topics.
  - Survey #2: Prioritize the new topic suggestions (some came from suggestions during the conference calls).
- We used the top 9 prioritized session topics from Survey #1, and the top 9 prioritized session topics from Survey #2, plus Intro to the Virtual Residency and Stories from the Trenches.
- We begged and pleaded for speakers (for months!).

# Apprenticeships

	Grant Proposal Writing	Paper Writing
<b>Established</b>	2017-18	2018-19
<b>Submitted</b>	NSF CyberTraining 2018, 2019, 2020 (declined)	PEARC'19, PEARC'20 (both published)
<b>Participants</b>	157 distinct (17% of VRP): 2017-18: 65; 2018-19: 86; 2019-20: 73	128 distinct (14% of VRP): 2018-19: 79; 2019-20: 81
<b>Institutions</b>	138 (37%)	112 (30%)
<b>States</b>	43 US states + 3 US territories + 2 other countries	38 US states + 2 US territories + 2 other countries
<b>Minority Serving</b>	24 (43% of VRP MSIs)	18 (32% of VRP MSIs)
<b>Non-PhD-granting</b>	37 (39% of VRP non-PhD)	25 (27% of VRP non-PhD)
<b>EPSCoR</b>	51 (50% of VRP EPSCoR) in 22 of 28 EPSCoR jurisdictions	39 (39% of VRP EPSCoR) in 18 of 28 EPSCoR juris.
<b>2020-21 Season</b>	Starting soon!	Starting soon!

Because of space limits on page 1 of the ACM format, only 15 of the 79 (19%) in 2018-19 and 15 of 81 (19%) in 2019-20 got to be listed as authors of each of those papers.



# Virtual Residency Workshop 2020

- Intermediate/Advanced
- Mon June 1 - Fri June 5 2020
- Remote via videoconferencing, and asynchronously via video recordings
- Curriculum designed by the community
- <http://www.oscer.ou.edu/virtualresidency2020/>
- **NEW!** Evaluation by XSEDE evaluation team
- **NEW!** Sessions based on CI Leadership Academy workshop
- We plan to submit a paper about it next year.



Research Comp Workforce  
NITRD HEC IWG, Thu Aug 20 2020





# Workshop 2020 Demographics

## ■ Gender

- VRP Women: 30%
  - US Population: 51% (VRP = 59% of US population)
  - All Computing/IT Occupations: 26% (VRP = 115% of CS/IT)
  - SC15-17: 13-14% (VRP = 200+% of recent SC conferences)

## ■ Race/Ethnicity

- VRP Underrepresented Minorities: 21%
  - US Population: 34% (VRP = 62% of US population)
  - All Computing/IT Occupations: 10% (VRP = 200+% of CS/IT)

<https://www.census.gov/quickfacts/fact/table/US/PST045219>

<https://www.bls.gov/cps/cpsaat11.htm>

<http://sc16.supercomputing.org/diversity/index.html>

<https://sc20.supercomputing.org/attend/inclusivity/demographics/>



Research Comp Workforce  
NITRD HEC IWG, Thu Aug 20 2020



# Does the Virtual Residency Work? #1

- The XSEDE evaluation team (Lorna Rivera, Lizanne DeStefano) have done an evaluation of the 2020 workshop.
- Sessions were rated 3.90 - 4.42 on a 1 - 5 scale.
- Effect on underrepresented populations
  - **Underrepresented Minorities**
    - **Experience**: Underrepresented minorities rated their experience as 5% **MORE SUCCESSFUL** than non-URMs rated it (4.76 vs 4.52).
    - **Sessions**: Underrepresented minorities rated 2 sessions 12% **HIGHER** than non-URMs rated them (4.71 vs 4.20, 4.71 vs 4.19).
    - **Google Docs**: Underrepresented minorities rated the Google Docs 13% **MORE USEFUL** than non-URMs rated them (4.58 vs 4.07)
    - No other statistically significant differences found.
  - **Women**
    - **Sessions**: Women rated 1 session 10% **LOWER** than men (3.85 vs 4.27).
    - No other statistically significant differences found.

# Does the Virtual Residency Work? #2

## INDIRECT MEASURE

Of the 330 institutions that started participating in the Virtual Residency Program before the 2020 workshop (and therefore could have participated in multiple activities):

- Multiple VRP activities: 257 institutions (78%)
- Multiple **types** of VRP activities: 237 institutions (72%)
  - workshops;
  - workshop planning calls;
  - grant proposal writing apprenticeship;
  - paper writing apprenticeship.

If it wasn't valuable to them, would they keep coming back?



INFORMATION TECHNOLOGY  
The UNIVERSITY of OKLAHOMA

Research Comp Workforce  
NITRD HEC IWG, Thu Aug 20 2020



# Acknowledgements

Portions of this material are based upon work supported by the National Science Foundation under the following grants:

- Grant No. ACI-1440783, “A Model for Advanced Cyberinfrastructure Research and Education Facilitators”
- Grant No. ACI-1548562, “XSEDE 2.0: Integrating, Enabling and Enhancing National Cyberinfrastructure with Expanding Community Involvement”
- Grant No. ACI-1620695, “RCN: Advancing Research and Education Through a National Network of Campus Cyberinfrastructure, Infrastructures – The CaRC Consortium”
- Grant No. ACI-1649475, “Cyberinfrastructure Leadership Academy”



INFORMATION TECHNOLOGY  
The UNIVERSITY of OKLAHOMA

<http://www.oscer.ou.edu/virtualresidency/>  
Research Comp Workforce  
NITRD HEC IWG, Thu Aug 20 2020



# Thanks for your attention!



## Questions?

[hneeman@ou.edu](mailto:hneeman@ou.edu)

*"Any opinions, findings, conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the Networking and Information Technology Research and Development Program."*

The Networking and Information Technology Research and Development  
(NITRD) Program

**Mailing Address:** NCO/NITRD, 2415 Eisenhower Avenue, Alexandria, VA 22314

**Physical Address:** 490 L'Enfant Plaza SW, Suite 8001, Washington, DC 20024, USA Tel: 202-459-9674,  
Fax: 202-459-9673, Email: [nco@nitrd.gov](mailto:nco@nitrd.gov), Website: <https://www.nitrd.gov>

