Campus Bridging in the Age of COVID-19

J. Eric Coulter
j.eric@gatech.edu
Georgia Institute of Technology
Atlanta, Georgia, USA

Stephen Bird
stebird@iu.edu
Indiana University
Bloomington, IN, USA

Resa Reynolds
resa.reynolds@cornell.edu
Cornell University
Ithaca, New York, USA

Winona Snapp-Childs
winona@iu.edu
Indiana University
Bloomington, IN, USA

Tabitha Samuel
tsamuel@utk.edu
University of Tennessee
Oak Ridge, TN, USA

Richard J. Knepper
rich.knepper@cornell.edu
Cornell University
Ithaca, New York, USA

ABSTRACT

Since 2015, the XSEDE Cyberinfrastructure Resource Integration (XCRI) team has engaged in a series of on-site visits of various types with institutions wishing to implement research computing resources in a way that gains from the lessons learned by larger XSEDE Resource Providers. The team originally developed a variety of flexible toolkits for such institutions, with the intent that they could be picked up at will and used to bootstrap local research computing programs by interested folks. In practice, the real value provided by the XCRI has proven to be the deep side-by-side work, mentorship, and focus on relationship building that comes along with carrying out a site visit. While the in-person aspect of these visits ceased of necessity beginning in 2020, activities continued in a distributed fashion, which proved to be a boon for several reasons. This paper provides an overview of the site visit process, and an exploration of the deep benefits of these remote collaborations, which are an efficient way to spread RCD knowledge to under-resourced institutions.

CCS CONCEPTS

• **Social and professional topics** → **Professional topics**; • **General and reference**;

KEYWORDS

HPC, XSEDE, ACCESS, Campus Bridging, Mentorship

ACM Reference Format:
J. Eric Coulter, Stephen Bird, Resa Reynolds, Winona Snapp-Childs, Tabitha Samuel, and Richard J. Knepper. 2023. Campus Bridging in the Age of COVID-19. In *Practice and Experience in Advanced Research Computing (PEARC '23), July 23–27, 2023, Portland, OR, USA.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3569951.3597604

1 INTRODUCTION

The initial practice of "Campus Bridging" derived from the Advisory Committee for Cyberinfrastructure (ACCI) Task Force for Campus Bridging [7]. The report of the ACCI Task Force noted that overall demand for computational resources by researchers largely outstripped the availability of resources and provided a number of recommendations to stimulate the sharing and utilization of resources. These recommendations included federation of campus resources, providing generalized software and workflow support, and sharing experiences of the national supercomputing centers and other universities with burgeoning cyberinfrastructure resources to campuses wishing to implement their own resources and build technical expertise. At the beginning of XSEDE 1 in 2011, a Campus Bridging team was formed that would provide software toolkits to make software management simpler and to provide guidance to campuses implementing their own resources that could be incorporated into the XSEDE Federation. At the beginning of XSEDE 2 in 2016, the team was renamed as the XSEDE Cyberinfrastructure Resource Integration (XCRI) but continued to carry out "Campus Bridging" activities.

2 TOOL DEVELOPMENT

The initial focus of the Campus Bridging project was to provide toolkits and documentation to make it possible for research computing administrators and facilitators to implement XSEDE-like resources at a local level. The goal of this was two-fold - one, to create a path for institutions to become L3 service providers[13] within XSEDE, and two, to aid the growth of the peripheral research computing ecosystem by making it easier for under-resourced (in time, money, or expertise) institutions to start and support local initiatives. While nominally the goal was to create a path towards becoming an L3 service provider, the unique constraints around XCRI engagements made that end goal difficult to achieve. Typically the institutions we were able to engage with only had some form of donated hardware, sufficient to support a few local research groups at a small scale before scaling out to larger XSEDE resources, or for classroom use. That being said, it became clear that the larger goal was being strongly supported in ways beyond the mere creation of local clusters. The deep side-by-side engagement model proved to be highly effective at transferring knowledge of HPC administration *and the larger research computing ecosystem* to under-resourced institutions. This often resulted in career moves for individuals leading those projects (with the US RCD field), representing greater flow in the jobs pipeline, increased funding to those institutions representing a move out of being "under"-resourced, or seeing leaders at those sites begin serving the community in larger roles at various RCD conferences and committees. These gains are impossible to

PEARC '23, July 23–27, 2023, Portland, OR, USA

ACM ISBN 978-1-4503-9985-2/23/07.
https://doi.org/10.1145/3569951.3597604

measure in clean metrics, so we hope to convey some of them in this wide-ranging experience paper.

2.1 Artifacts

The XCRI team developed a set of tools intended to smooth the path for campuses to implement resources on their own, and to effectively use XSEDE resources. The rationale behind the creation of these toolkits was to create environments for researchers that replicate what they might experience as users of national centers, allowing novice researchers to learn locally and work nationally. As the team worked with more sites, lessons learned from those interactions were built into the toolkits, to be shared with future and prior sites. A second-order effect of these toolkits was to provide scaffolding for new-to-RCD IT professionals to implement resources and learn the tools even with limited means.

The primary examples of these, are the The XSEDE Compatible Basic Cluster[6] (XCBC) and the The Jetstream Virtual Cluster[4] (JS-VC) toolkits. The XCBC is based on Ansible[14] and OpenHPC[15], enables a simple cluster build similar to those found on XSEDE resources The JS-VC toolkit allows one to build virtual clusters on the Jetstream[16] cloud, and is now in production on Jetstream 2[8]. Other toolkits include materials for running cluster-build workshops[18], container tutorials[20], and deploying big-data resources[19].

3 SITE VISITS

In-person site visits often had a typical flow, with minor variations. After initial contact, the XCRI team would schedule a series of meetings to explore the needs at the institution, set expectations for how XCRI would be able to help. Early representative site visits are described in great detail in prior works[3, 5]. In previous writings about the XCRI activities we also discuss the "catalytic action" that site visit activities provide to draw institutional attention to the activities and gain local support [2].

Site visits depended on three main requirements - that the institution had access to hardware, specifically needed an HPC-style system, and was agreeable to using the software stack that XCRI provided. After that point, actual planning began. This typically involved 1-3 logistics meetings, planning to ready the hardware for software deployment, travel planning, network access and checklist of cluster build activities. Consultation with local researchers was included when possible, to tailor the new environment to their needs. The site visit travel itself also involved planning on the part of XCRI. These trips typically encompassed 5 business days (40-50 "working" hours) but also consumed an additional 100 or so hours of "staff time away from home" - these hours are included in Fig. 1 for comparison with remote "visits". In many cases, there were subsequent remote followups (and occasional return trips), as new hardware was added, or new additional software was desired. Previous ROI investigations[17] have included explicit monetary costs (salaries and travel budgets) but not focussed on the time-investment at the level of specific operations, or gone to the level of including "second-order" effects on the sites visited.

XCRI also worked in fully remote mode with several sites. While initially due to the distributed nature of the team, this was useful when travel restrictions due to COVID-19 fell in 2020. In the ideal case for this mode, the team carried out initial consultation, then a few planning sessions followed by an "initial visit" with a week or two of intensive remote sessions, followed by regular working meetings of 2-4 hours through the course of the project. Despite the confounding factors of lack of in-person time and hands-on access to hardware, this proved to be a successful model for engagement in many ways.

For one, the total time spent working with a site is actually much less, as in Fig. 1, mainly due to the removal of travel time and logistics planning. Additionally, it is *much faster* to begin actual work. Hardware management was handled by the site in all cases, with occasional in-the-data-center zoom calls with the XCRI team. There was less schedule friction and no energy/time/money lost to travel. The "catalytic action" produced by planning an initial "visit" still resulted, which included the work of building the base cluster before adding on things like monitoring, portals, and scientific software (which the site may or may not have desired). The longer term nature of remote engagements also resulted in effectively spending more time with site personnel, which increased 2nd-order effects of familiarization with the larger RCD community and relationship building with XCRI staff.

3.1 University of Cincinnati

The first remote site visit occurred in late 2018, when XCRI worked with Jane Combs, director of the University of Cincinnati Advanced Research Computing Center to stand up a new central HPC resource[10]. Due to the distributed nature of the team and travel scheduling difficulties, one team member went on-site for the build, while another assisted remotely on a short 2-day visit. A second, longer visit was conducted, in which the entire system was brought online, including installation of XDMoD and configuration of the Spack software management system. The initial remote build was an excellent test run for the fully remote method.

3.2 Langston University

Our first fully-remote engagement began during the fall of 2019, when Dr. Franklin Fondjo of Langston University reached out to the XCRI team, wishing to develop a new branch of the XCBC toolkit, to support the use of the xCAT[1] provisioning software, rather than Warewulf[21]. This engagement took the form of roughly bi-weekly 2-3 hours meetings, in which Dr. Fondjo and worked with XCRI staff to update the Ansible playbooks which make up the XCBC toolkit, and deploy them on hardware local to Langston (the Lucille cluster), used for high-energy physics and computer science research. In addition to use of xCAT, XCRI also helped configure two different portals for Lucille - JupyterHub and RStudio Server, in addition to a list of scientific software needed at Langston. Dr. Fondjo assisted in the development effort of these as well, which were folded back into both the XCBC and the Jetstream Virtual Cluster toolkits.

This was the second time XCRI worked with a partner to expand the toolkit[5], and was successful despite the remote format. While the initial xCAT build finished in late 2020, by mid 2021, more hardware arrived, and Dr. Fondjo wished to migrate back to the Warewulf software, as more sites in the region and adjacent ecosystem were using it. This resulted in a *second* round of remote

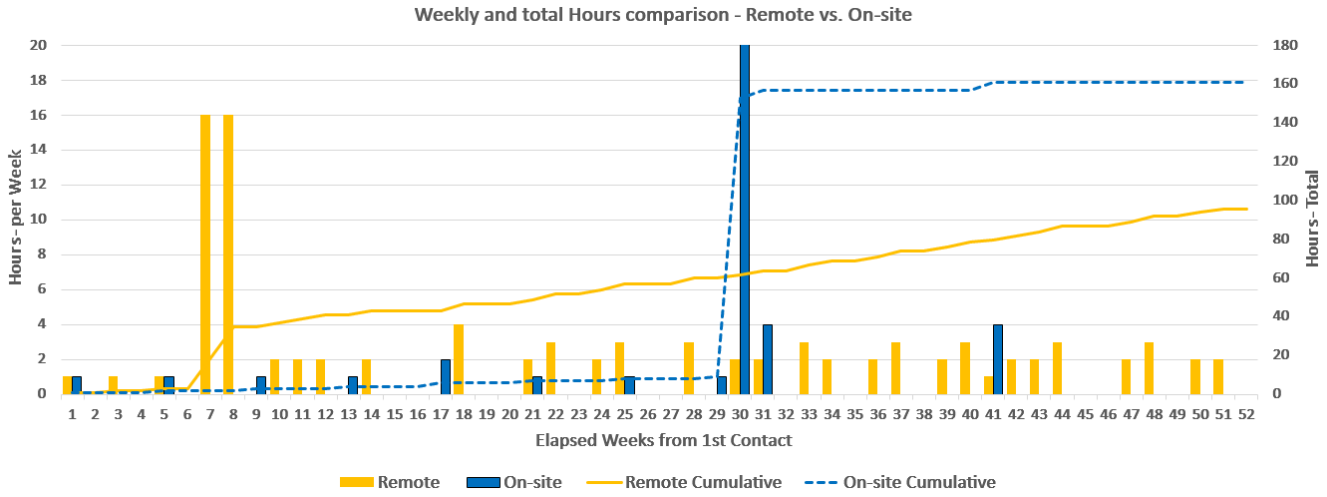

Figure 1: Comparison of hours worked over the course of a year while engaging with a site. This does NOT reflect specific data, but rather the typical forms of both types of interaction.

visits, in a bi-weekly meeting format, where the Lucille cluster was fully rebuilt to incorporate the new hardware, cluster management software, research portals, and scientific software.

3.3 University of Central Oklahoma (UCO)

While engaged with Langston University, the XCRI team also received a request to work with Dr. Evan Lemley and a student administrator at the University of Central Oklahoma, to rebuild their local HPC system, known as Buddy. After initial consultation, Dr. Lemley wanted to implement the basic XCBC toolkit with the addition of an Open OnDemand[9] gateway. The teams set a hybrid schedule, with the initial 5-day site visit spread across two weeks in order to reduce "zoom fatigue", and to accommodate complex schedules at the beginning of the semester. With the flexibility afforded by remote work, the visit was able to begin in earnest much more quickly than if a full trip was necessary. Once the base cluster build was complete, work turned to development effort, where the student admin and XCRI staff met bi-weekly to advance the goal of adding Open OnDemand to the XCRI toolkit. That effort took place the following spring from the initial build, due to scheduling requirements. Meeting bi-weekly From late February, 2021 to May, 2021, the team updated the XCBC toolkit with a branch allowing Open OnDemand to be fully installed alongside the base XCBC software stack. Not only did this empower UCO with a performant graphical interface for students using the cluster, it allowed the XCRI team to offer that same capability to future sites. The final build of "Buddy" consisted of 37 nodes, including 2 GPU and Intel Xeon Phi nodes[11].

3.4 Southwestern Oklahoma University (SWOSU)

The final remote "visit" took place with Southwestern Oklahoma State University (SWOSU). This engagement took place from Spring of 2021 to Summer of 2022, covering multiple cluster rebuilds and student administrators. Initially, Dr. Jeremy Evert of SWOSU contacted XCRI, asking for assistance building a local HPC resource with donated hardware, for teaching purposes, and to help a student admin gain HPC skills. The first build of "Duke" consisted of only 4 nodes, each with 32 cores and 256 GB of RAM, which were formerly large Data Transfer Nodes for use in the One OCII[12] network. The final build contained some 120 nodes of donated hardware, which Dr. Evert learned about thanks to the deeper involvement with the US Research Computing ecosystem gained from working with the XCRI team. The initial "visit" was carried out after just two or three initial consulting sessions, As usual, the build happened over 5 days spread out over two weeks, with semi-regular followups throughout the Fall. Given the need for student admin education, the XCRI team focused strongly on ensuring the student admin had hands-on as much as possible.

During the following spring, the XCRI team worked with the SWOSU admin to implement an Open OnDemand portal for "Duke". However, the team encountered package conflicts with the latest OpenHPC release, and spent several sessions working together to resolve these, modelling good collaboration with open-source software projects. Near the end of that spring, a new student admin came on-board, and the project shifted to assisting with the training and transition of the new student admin. Through the following summer, the XCRI team worked with the new admin to train them in HPC administration, and saw them engage deeply with the community, participating in both PEARC and SC, while keeping the cluster running for use by courses during the Fall semester. At this point, the remote engagement closed with the end of funding to XSEDE.

3.5 Outcomes

There is a wide spread in outcomes from the sites the XCRI team has worked with. In some cases, there was little to no feedback from the site after the visit, while in others, long-term collaboration emerged and led to greater engagement with the wider community. Outcomes range from "completely unknown" to "we know the

system supported several graduate students in their studies" to "the new RC program is thriving, continuing to bring in new funding and hardware". It is largely through effort and buy-in from the local sites, of course, that the latter outcomes are possible, and these are also typically cases where engagement with the XCRI team is more extended. The "2nd-order" network effects of knowing professionals in the RCD field become quite large when viewed at the level of a small institution or career.

3.5.1 Career Outcomes. In many cases, site visit representatives and local administrators went on to enjoy greater prosperity and opportunity in their careers shortly after working with XCRI. In some cases, there is direct attribution possible, while in others the water is murkier. In five known cases, site personnel went on to either more permanent employment in the RCD field or made significant upward career moves within the RCD field after gaining experience and exposure to the community by working with the XCRI team. In one case, a site was able to access and successfully use $400,000 in donated hardware thanks working with the team. Several other site visit contributors have continued to take on significant volunteer roles within the community, serving as members of the OpenHPC[15] Technical Steering Committee, the Campus Champions network, NSF review panels, Supercomputing student programs, PEARC conference committees, and the Campus Research Computing Consortium (CaRCC).

4 CONCLUSIONS

The activities of the team formerly known as Campus Bridging reveal evidence for the diverse ways that boosting research computing capabilities, and more importantly, knowledge, at smaller institutions can have ripple effects well beyond the boundaries of those institutions. In multiple instances, this has resulted in career improvements for students or beginning RCD staff, without touching on the number of local students or researchers who were impacted by having local access to a an XSEDE-like HPC resource. The effects of this work have been transformative in many cases, and point to an effective method of energizing the RCD pipeline and ecosystem across the US. While difficult to scale, this effort was carried out with only 2.5 FTE for the last 6 years, in part due to how effective long-term remote mentorship is at up-skilling individuals supporting smaller institutions.

ACKNOWLEDGMENTS

This work was funded by NSF Award #1548562. The XCRI Team thanks their collaborators: Ms. Jane Combs, Dr. Evan Lemley, Dr. Franklin Fondjo, and Dr. Jeremy Evert for their work and support in representing us to the community.

REFERENCES

[1] IBM Corporation. 2015. xCAT: Extreme Cluster/Cloud Administration Toolkit. https://xcat.org/.

[2] Eric Coulter, Jeremy Fischer, Barbara Hallock, Richard Knepper, Dave Lifka, JP Navarro, Marlon Pierce, and Craig Stewart. 2016. Cyberinfrastructure As a Platform to Facilitate Effective Collaboration Between Institutions and Support Collaboratories. In *Proceedings of the 2016 ACM on SIGUCCS Annual Conference* (Denver, Colorado, USA) *(SIGUCCS '16)*. ACM, New York, NY, USA, 37–42. https://doi.org/10.1145/2974927.2974962

[3] Eric Coulter, Jeremy Fischer, Barbara Hallock, Richard Knepper, and Craig Stewart. 2016. Implementation of Simple XSEDE-Like Clusters: Science Enabled and Lessons Learned. In *Proceedings of the XSEDE16 Conference on Diversity, Big Data, and Science at Scale* (Miami, USA) *(XSEDE16)*. ACM, New York, NY, USA, Article 10, 8 pages. https://doi.org/10.1145/2949550.2949570

[4] J. Eric Coulter, Eroma Abeysinghe, Sudhakar Pamidighantam, and Marlon Pierce. 2019. Virtual Clusters in the Jetstream Cloud: A Story of Elasticized HPC. In *Proceedings of the Humans in the Loop: Enabling and Facilitating Research on Cloud Computing* (Chicago, IL, USA) *(HARC '19)*. Association for Computing Machinery, New York, NY, USA, Article 8, 6 pages. https://doi.org/10.1145/3355738.3355752

[5] J. Eric Coulter, Rich Knepper, Resa Reynolds, Jodie Sprouse, and Stephen Bird. 2019. Beyond Campus Bridging: A Retrospective of Cyberinfrastructure Integration Efforts. In *Proceedings of the Practice and Experience in Advanced Research Computing on Rise of the Machines (Learning)* (Chicago, IL, USA) *(PEARC '19)*. Association for Computing Machinery, New York, NY, USA, Article 93, 4 pages. https://doi.org/10.1145/3332186.3333151

[6] Jeremy Fischer, J. Eric Coulter, Barbara Hallock, and Richard Knepper. 2015. XCBC & XNIT using the LittleFe and Limulus HPC200. https://scholarworks.iu.edu/dspace/handle/2022/20321 XSEDE 2015 Tutorial.

[7] National Science Foundation. 2012. NSF Advisory Committee for Cyberinfrastructure Task Force on Campus Bridging. Final Report, 2011. Retrieved June 25, 2020 from https://www.nsf.gov/cise/oac/taskforces/TaskForceReport_CampusBridging.pdf

[8] David Y. Hancock, Jeremy Fischer, John Michael Lowe, Winona Snapp-Childs, Marlon Pierce, Suresh Marru, J. Eric Coulter, Matthew Vaughn, Brian Beck, Nirav Merchant, Edwin Skidmore, and Gwen Jacobs. 2021. Jetstream2: Accelerating Cloud Computing via Jetstream. In *Practice and Experience in Advanced Research Computing* (Boston, MA, USA) *(PEARC '21)*. Association for Computing Machinery, New York, NY, USA, Article 11, 8 pages. https://doi.org/10.1145/3437359.3465565

[9] Dave Hudak, Doug Johnson, Alan Chalker, Jeremy Nicklas, Eric Franz, Trey Dockendorf, and Brian L. McMichael. 2018. Open OnDemand: A web-based client portal for HPC centers. *Journal of Open Source Software* 3, 25 (2018), 622. https://doi.org/10.21105/joss.00622

[10] IU UITS News. 2019. XSEDE personnel help University of Cincinnati create Advanced Research Computing Cluster. 'https://itnews.iu.edu/articles/2019/XSEDE%20personnel%20help%20University%20of%20Cincinnati%20create%20Advanced%20Research%20Computing%20Cluster%20.php'

[11] IU UITS News. 2021. XSEDE and University of Central Oklahoma (UCO) work together to rebuild UCO's high performance computing system, Buddy. https://itnews.iu.edu/articles/2021/XSEDE-and-University-of-Central-Oklahoma-work-together-to-rebuild-UCOs-high-performance-computing-system-Buddy-.php.

[12] One OCII. 2008. The Oklahoma Cynerinfrastructure Initiative. https://www.oneocii.okepscor.org/about/.

[13] XSEDE Project. 2020. Service Providers - XSEDE. https://web.archive.org/web/20220819181630/https://www.xsede.org/ecosystem/service-providers.

[14] RedHat. 2018. Ansible Configuration Management Software. http://www.ansible.com/. [Online; accessed 17-August-2018; Red Hat Inc.].

[15] Karl W. Schulz, C. Reese Baird, David Brayford, Yiannis Georgiou, Gregory M. Kurtzer, Derek Simmel, Thomas Sterling, Nirmala Sundararajan, and Eric Van Hensbergen. 2016. Cluster Computing with OpenHPC. http://hdl.handle.net/2022/21082.

[16] Craig A. Stewart, Timothy M. Cockerill, Ian Foster, David Hancock, Nirav Merchant, Edwin Skidmore, Daniel Stanzione, James Taylor, Steven Tuecke, George Turner, Matthew Vaughn, and Niall I. Gaffney. 2015. Jetstream: A Self-provisioned, Scalable Science and Engineering Cloud Environment. In *Proceedings of the 2015 XSEDE Conference: Scientific Advancements Enabled by Enhanced Cyberinfrastructure* (St. Louis, Missouri) *(XSEDE '15)*. ACM, New York, NY, USA, Article 29, 8 pages. https://doi.org/10.1145/2792745.2792774

[17] Craig A. Stewart, David Y. Hancock, Julie Wernert, Thomas Furlani, David Lifka, Alan Sill, Nicholas Berente, Donald F. McMullen, Thomas Cheatham, Amy Apon, Ron Payne, and Shawn D. Slavin. 2019. Assessment of Financial Returns on Investments in Cyberinfrastructure Facilities: A Survey of Current Methods. In *Proceedings of the Practice and Experience in Advanced Research Computing on Rise of the Machines (Learning)* (Chicago, IL, USA) *(PEARC '19)*. Association for Computing Machinery, New York, NY, USA, Article 33, 8 pages. https://doi.org/10.1145/3332186.3332228

[18] XSEDE XCRI Team. 2019. XCRI Cluster Build Tutorial. https://github.com/XSEDE/CRI_JS_Cluster_Tutorial.

[19] XSEDE XCRI Team. 2020. XCRI BigTop Stack. https://github.com/XSEDE/CRI_Bigtop_Stack.

[20] XSEDE XCRI Team. 2020. XCRI Container Tutorial Materials. https://github.com/XSEDE/Container_Tutorial/.

[21] Warewulf. 2018. Warewulf OS Management Toolkit. http://warewulf.lbl.gov/trac/wiki/About. [Online; accessed 26-Sept-2018; Berkely Lab].

Received 21 April 2023