Beyond Campus Bridging: A retrospective of Cyberinfrastructure Integration Efforts

J. Eric Coulter
jecoulte@iu.edu
Science Gateways Research Center
Indiana University
Bloomington, IN, USA

Rich Knepper
rich.knepper@cornell.edu
Center for Advanced Computing
Cornell University
Ithaca, NY, USA

Resa Reynolds
resa.reynolds@cornell.edu
Center for Advanced Computing
Cornell University
Ithaca, NY, USA

Jodie Sprouse
jhs43@cornell.edu
Center for Advanced Computing
Cornell University
Ithaca, NY, USA

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

ACM Reference Format:
J. Eric Coulter, Rich Knepper, Resa Reynolds, Jodie Sprouse, and Stephen Bird. 2019. Beyond Campus Bridging: A retrospective of Cyberinfrastructure Integration Efforts. In *Practice and Experience in Advanced Research Computing (PEARC '19), July 28-August 1, 2019, Chicago, IL, USA.* ACM, New York, NY, USA, **??** pages. https://doi.org/10.1145/3332186.3333151

1 INTRODUCTION

The XSEDE Cyberinfrastructure Resource Integration team (XCRI) was originally known as the "Campus Bridging" team for XSEDE, incorporating the recommendations of the NSF ACCI task force for Campus Bridging [**?**], and charged to provide easier means for researchers to adapt to using computational resources at the scale of national supercomputing centers. When the XSEDE project entered its second iteration, it was decided that the Campus Bridging name was too broad and unclear, and the Cyberinfrastructure Resource Integration name was coined[1]. Based on input from the community and the NSF, XCRI now works to extend the leveragability of the aggregate national cyberinfrastructure through the XSEDE Service Provider Forum, campuses, and regional partnerships in order to maximize support for research initiatives. A significant portion of XCRI's activities revolve around developing and implementing toolkits that support campus level cyberinfrastructure initiatives. These toolkits are created with the intention of allowing campus system administrators to easily create an environment that incorporates a number of features of national supercomputing resources. The rationale behind this type of toolkit is that most researchers begin by running on local systems at the campus level, and making the transition to large-scale supercomputing systems is difficult and time-consuming. At the same time, campus system administrators are interested in using software used at national sites that makes it easier to allocate resources, manage and monitor systems, and produce metrics. By providing these toolkits, XCRI attempts to facilitate more effective management of the local resource and easier transition to national resources when researchers are ready to take advantage of XSEDE systems.

XCRI also works with campuses to implement these toolkits on local resources, resulting in a growing community of cyberinfrastructure administrators with resources at their disposal that enable computational research activities easily translated to XSEDE resources. Working with this community, XCRI has identified a number of improvements to its toolkits as well as its own processes, as well as some gaps where improvements can be made in both the kinds of software and the types of services provided. In the paper below, we detail the efforts, responses of the community, and lessons learned in carrying out the XCRI mission activities. We discuss the activities involved in establishing a multi-institutional cyberinfrastructure team, reaching out to the community in order to build interest and find interested schools, developing the toolkits and services which the XCRI provides, and establishing metrics in order to gauge productivity and identify goals for XCRI activities.

2 PUBLICITY AND OUTREACH

A significant portion of any such activity involves informing the community of available services and clearly explaining the types of services on offer. In order to garner interest in XCRI activities, considerable time was spent in the initial years of the program presenting the ideas behind XCRI at community events, soliciting requirements that would shape the choices made by XCRI in their toolkits and other offerings, and preparing documentation and similar materials that would make clear the XCRI mission and activities. Over the last 8 years of Campus Bridging and XCRI activities representing both the original XSEDE project and its renewal, the team has made dozens of presentations, presented 5 tutorials, and over 20 papers or posters representing activities and services available to the user community, in addition to participating in the broader cyberinfrastructure community via the Campus Champions group and the OpenHPC [**?**] project technical leadership. The XCRI team

[1] Just in time for everyone to start calling the team "the Campus Bridging group".

PEARC '19, July 28-August 1, 2019, Chicago, IL, USA

ACM ISBN 978-1-4503-7227-5/19/07…$15.00
https://doi.org/10.1145/3332186.3333151

has also formed a strong collaboration with the Science Gateways Research Center[**?**], which provides easy cyberinfrastructure access to a huge variety of domain science projects.There has been substantial "return on investment" for these activities in the form of new collaborations, site visit partners, and a growing group of XCRI adopters. The most effective forms of outreach have been in-person tutorials, in which the team is able to walk users through the implementation process and discuss choices and conditions for particular details.

3 SERVICES AND TOOLKITS

3.1 Toolkits

The XCRI offerings include toolkits for a variety of needs that a campus resource provider might encounter. The primary toolkit is the so-called XSEDE-Compatible Basic Cluster[**? ? ?**] (XCBC), the implementation of which typically involves a site visit from an XCRI field engineer. The XSEDE National Integration Toolkit (XNIT) offers a repository of commonly used scientific applications, is intended to ease the burden of ongoing system administration, and is easily installed with little assistance from XCRI team members. The Cluster Monitoring Toolkit is typically used alongside the XCBC, and enables easy setup and configuration of common tracking tools for HPC systems. The XCRI Virtual Cluster toolkit[**?**] allows users to create bespoke HPC-style systems on the Jetstream[**?**] cloud, for experimental, educational, or research purposes. The XCBC, Cluster Monitoring, and Virtual Cluster toolkits have all benefited from the use of Ansible, a commonly used daemon-less configuration management tool. In the full paper, we will report more detailed useage, descriptions, and ease of updating the various toolkits.

3.1.1 XCBC. The XCBC is designed to provide an easy entry point for system administrators either unfamiliar with HPC systems, or without adequate time to easily implement such a system on their own. The OpenHPC component provides HPC-oriented software offerings on top of the basic operating system installation. The aims of the OpenHPC project are well-described in [**?**], providing a stable basic operating system with a minimum of duplicate packages that would not be supported by other linux distributions, and a selection of HPC software that meets a broad range of scientific computing needs. The XCRI team found that the OpenHPC project aligned well with their goals for propagating best practices from scientific computing centers to institutions which would otherwise need to do considerable investigation on their own.

3.1.2 XNIT. The XNIT is implemented as a standard yum repository. For more details, see[**?**]. There have been approximately 100 site subscribers over the XNIT lifetime, based on our collected metrics, but XCRI is in the process of transitioning to an easier to maintain tool for easing the pain of installing scientific software.

3.1.3 Cluster Monitoring Toolkit. The Cluster Monitoring toolkit allows an HPC administrator to quickly install and configure OpenXDMoD[**?**] and Ganglia[**?**]. OpenXDMoD allows for tracking of jobs at a very granular level, and subsequently creating reports on cluster useage, which is often invaluable to resource operators in justifying and procuring funding. Ganglia allows for real-time monitoring of the current state of the entire system, giving an at-a-glance view of its health.

3.1.4 Virtual Clusters. The Virtual Cluster toolkit is implemented as a series of scripts that enable one to quickly create an elastic HPC environment on the Jetstream cloud, which allows for near-maximally efficient use of limited allocations, in that resources are used only while computational jobs are actually running. This frees users from the burden of creating and destroying virtual machines manually.

Site	Nodes	Users	TFLOPS
Brandeis University	145	100	200*
South Dakota State University	70	10-100	200*
Slippery Rock University	10	10	10
Doane University	33	1.6	10
University of Cincinnati	36	101	10-100

Table 1: This table shows recent site visits in which the XCRI team implemented the latest version of the XCBC. As can be seen in the table, metric choice is extremely important in presenting an accurate picture of a site. *Indicates inclusion of GPU nodes in the theoretical TFLOPs estimate.

3.2 Site Visits

XCRI has embarked on 5 site visits in which XCBC resources were implemented, separate from those done during the Campus Bridging years, which are well described in previous publications[**?**].

In this section, we will discuss in more detail the results and lessons learned from our partner site interactions over the to-date lifetime of the XCRI team.

4 ORGANIZATIONAL CONSIDERATIONS

XCRI has maintained these efforts with fewer than 3.5 full time staff equivalents (FTEs), with responsibilities spread across 6 actual humans. While fragmentary effort makes sustained focus on a particular project difficult, the multi-project nature of a team aimed at so diverse a user community is a good fit for the typical grant-funded worker. It is also quite helpful to have connections to experienced professionals at funded CI providers, who often are able to provide deep advice on specific technical topics to collaborators which would otherwise be beyond the ability of the XCRI team, such as in the case of long-term user management or filesystem performance.

Any discussion of organizational details would be incomplete without mention of how metrics are tracked to ensure activities are in alignment with the core mission. During the initial Campus Bridging phase, one of the prime metrics tracked was theoretical TFLOPs enabled across various XCBC or XNIT host sites, which turned out to be unreliable in two ways: one, TFLOPs was not necessarily correlated with number of users, and two, was not really valuable to the goal of engaging CI professionals and researchers with the national research computing landscape. While hardware numbers are easy to track and fun to throw around, the team realized it was perhaps more important to track an estimated number of *users* supported by partner institutions. This is not to imply that

Metric	Description
Total TFLOPs	Theoretical Rmax TFLOPs at sites using XCRI Toolkits
Number of Users	Users supported at XCBC or XNIT host institutions
Partnership Interactions	Significant interactions by XCRI team members with outside institutions
Toolkit Updates	Number of significant updates to software toolkits
Toolkit Releases	Number of new toolkits released

Table 2: Illustration of some of the core metrics tracked by the XCRI team.

institutions without a certain number of users would be turned away, but that such human factors are more relevant to the actual mission. Additionally, after a number of shorter consultations in which sites received some form of aid without implementing local XCBC resources, or who simply wished for brief discussions about their existing resources, XCRI began tracking the actual number of interactions with interested parties, which again, proved to be more useful in showing the amount of value provided by the team. For the "User Satisfaction" metric, data is gathered from any workshops, tutorials, or site visits undertaken in the reporting period.

4.1 Maintenance

One important feature of such software-organizing projects must be a plan for deprecation of outdated software, for whatever definition of outdated is relevant. For example, in the case of scheduler choice, the XCRI team migrated to using the SLURM[**?**] scheduler rather than SGE, due to obvious signs of community shift, such as the percentage of HPC centers using a particular scheduler[**?**]. Even well-established technologies rise and fall, and it is best to be prepared, in order to more rapidly transition to workable solutions. During the long interim between the update from the Rocks cluster management software[**? ?**] supporting CentOS 6 to CentOS 7, it became clear that a transition was needed, but this took the better part of a year, due to the complexity of changing technologies after adoption. Unlike deprecation plans, XCRI has wisely budgeted for and tracked maintenance of their existing toolkits, which have largely remained in fettle over the project lifetime after initial creation.

4.2 Rolling out new Tools

Since the transition to OpenHPC, the XCRI team has implemented best practices for new toolkit creation based on those used by the XSEDE Requirments Analysis and Capability Delivery (RACD) team[**?**]. While the team keeps in touch with the community in a variety of ways discussed elsewhere (Sec. **??**), soliciting feedback about user needs is particularly difficult when it comes to tooling. Oftentimes, CI admins-to-be are not fully aware of the needs of their researchers, or aware of what tools will help, while established CI admins are happy with their environments, or enmeshed in issues of user control/engagement. When new technologies appear in the CI landscape, or appear to be gaining mindspace, discussion is naturally undertaken within the team as to whether or not there is a need that could be filled by XCRI in that area. If possible, the team will attempt to solicit feedback on such thoughts from the community, either by mailing list, discussion during a conference presentation, or by casually contacting campus collaborators. While the XCRI team is not funded for the development of new software tools, part of our mandate is to ease the installation and configuration of existing tools, which often necessitates a certain level of development-like activity, often similar to that described as "DevOps" in the commercial world. In such activities, access to testing hardware is a necessity, for which the Jetstream[**?**] resource has been particularly invaluable. Access to a fully configurable cloud has almost (but not quite) obviated the need for XCRI to maintain our own hardware resources. In some cases, such as helping sites implement GPU-powered computing resources, it is still necessary to have access to real silicon-and-metal hardware.

During this discussion, the word "community" is often thrown around without due consideration. In the case of XCRI, the community is a bit nebulous. We work with the weedy fringe of the research computing forest, not quite the endless meadow of research workstations, but also not the deciduous giants that make up the large computing centers, and attract researchers with the greatest need and computational expertise (pollinators). At the fringes of this forest, XCRI functions as a forester, both nurturing and planting new growth in the form of campus resource providers. At some point, such institutions either grow beyond the need for our help, or fail to realize sufficient usage of their resources. Given the "uninitiated" nature of our user base, "keeping up with" the community requires paying attention to a variety of different groups, and making regular interactions in order to display our availability to new entrants. Additionally, this process yields insight into the current state of practice at larger institutions and more established CI providers. As mentioned in section **??**, having XCRI staff situated at a variety of home institutions is critical for deep access to institutional CI provider knowledge.

5 PARTNERSHIPS

In all of XCRIs activities, providing lasting value to partner institutions in the form of improved access to research computing infrastructure has been the prime driver, but the relationship is not at all a one-way street. Every interaction with a partner site provides feedback in some form or another, and ongoing interactions highlight areas of greatest need. For example, four different sites have requested aid in setting up compute nodes with GPU hardware, after initial site visits from the XCRI team, in cases where the hardware was not available during the visit. Sites are always encouraged to contact XCRI if they run into any problems, though they are also pushed towards the ever-helpful Campus Champions community as well. Several have become XSEDE Level 3 service providers, further strengthening their connection to the national landscape. Thanks to the lean nature of the XCRI team, these partners have been absolutely invaluable in spreading word of our services, resulting in further contacts with interested groups. We would especially like to thank Chet Langin (Southern Illinois University), Jason Wells (Bentley University), Nitin Sukhija (Slippery Rock University), Gi

Vania (University of Texas at Dallas), and Adam Erck (Doane University), in no particular order, for their continued collaboration and efforts in promoting our work to the larger community.

6 CONCLUSION

In short, XCRI has used a wide variety of strategies to succeed in facilitating cyberinfrastructure awareness and implementation at resource-constrained institutions across a large geographical area. Providing software toolkits to ease the burden of computational resource providers continues to reduce barriers to entry in the national research computing scene. While large national computing centers provide an invaluable resource to a certain class of researcher, investment in small and mid-scale resources is a must for continuing to provide both a pipeline of large-scale users and a continuum of resources for new researchers.

It is clear from recent NSF solicitations that cyber-teams similar to XCRI should continue to play an important role in the US national CI landscape. These lessons learned from our efforts will ideally help strengthen future programs aimed at similar ends, in whatever form those take. The organizational format of XCRI has been effective in meeting and exceeding various goals in pursuit of increasing the accessibility of both national and local research computing resources across the US, with a small team.

ACKNOWLEDGMENTS

This work was funded through the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation grant number ACI-1548562. Additionally, the authors would like to thank George Turner of Indiana University for his invaluable support during the visit to the University of Cincinnati, for calling us in to begin with, and for continuing to provide support without hesitation.