# What is Campus Bridging and what is XSEDE doing about it?

**Craig A. Stewart**
Indiana University
2709 E 10th Street
Bloomington, IN 47408
+1 812-325-3452

stewart@iu.edu

**Richard Knepper**
Indiana University
2709 E 10th Street
Bloomington, IN 47408
+1 812-855-9574

rknepper@indiana.edu

**James Ferguson**
University of Tennessee, Knoxville
1 Bethel Valley Road, Bldg 5100
Oak Ridge, TN 37830
+1 865-241-3702

jwf@utk.edu

**Felix Bachmann**
Carnegie Mellon University
4500 5th Avenue
Pittsburgh, PA 15213
+1 412-268-6194

fb@sei.cmu.edu

**Ian Foster**
The University of Chicago and
Argonne National Laboratory
Argonne, IL 60439
+1 630-252-4619

foster@anl.gov

**Andrew Grimshaw**
University of Virginia
PO Box 400740
Charlottesville, VA 22904
+1 434-982-2204

grimshaw@virginia.edu

**Victor Hazlewood**
University of Tennessee, Knoxville
1 Bethel Valley Road, Bldg 5100
Oak Ridge, TN 37830
+1 865-241-1528

vhazlewo@utk.edu

**David Lifka**
Cornell University
512 Frank H. T. Rhodes Hall
Ithaca, NY 14853
+1 607-254-8621

lifka@cac.cornell.edu

## ABSTRACT

The term "campus bridging" was first used in the charge given to an NSF Advisory Committee for Cyberinfrastructure task force. That task force developed this description of campus bridging:

*"Campus bridging is the seamlessly integrated use of cyberinfrastructure operated by a scientist or engineer with other cyberinfrastructure on the scientist's campus, at other campuses, and at the regional, national, and international levels as if they were proximate to the scientist, and when working within the context of a Virtual Organization (VO) make the 'virtual' aspect of the organization irrelevant (or helpful) to the work of the VO."*

Campus bridging is more a viewpoint and a set of approaches to usability, software, and information concerns than a particular set of tools or software. We outline here several specific use cases that have been identified as priorities for XSEDE in the next four years. These priorities include documentation, deployment of software used entirely outside of XSEDE, and software that helps bridge from individual researcher to campus to XSEDE cyberinfrastructure. We also describe early pilot tests and means by which the user community may stay informed of campus bridging activities and participate in the implementation of Campus Bridging tools created by XSEDE. Metrics are still being developed, and will include (1) the number of campuses that adopt and use Campus Bridging tools developed by XSEDE and (2) the number of and extent to which XSEDE-developed Campus Bridging tools are adopted among other CI projects.

## Categories and Subject Descriptors

*D.2 SOFTWARE ENGINEERING*

## General Terms

Human Factors, Standardization, Design

## Keywords

Campus Bridging, cyberinfrastructure, usability, middleware

## 1. INTRODUCTION

H. Edward Seidel coined the term *campus bridging* in 2009, when he charged six task forces of the National Science Foundation (NSF) Advisory Committee for Cyberinfrastructure (ACCI). Five of the six task forces focused on well-known topics of importance in cyberinfrastructure (CI): cyberlearning and workforce development; data and visualization; grand challenges; high performance computing; and software [12]. The sixth task force—Campus Bridging—tackled a novel topic. Bridges from campus to national CI were needed because going from local to national facilities felt, to many researchers, like falling off a cliff. One goal of the Task Force on Campus Bridging was to suggest steps that the national cyberinfrastructure community and the NSF could take to develop a better, more comprehensive national CI including campus, state, regional, and federally funded CI facilities. One finding included in the ACCI Campus Bridging Task Force Final Report [24] is that the current state of CI software and organizational, security, and policy constraints caused many systems within the US to be used at less than their full capability and capacity. At the same time, a survey of researchers funded by the NSF as principal investigators but not TeraGrid users [6] found that more than 20% of respondents reported that they never had access to sufficient CI resources to do their research, and more than 70% indicated that at least some of

the time they lacked sufficient CI resources [30]. Taken together, these points clearly indicated a need for a better (and better integrated) national cyberinfrastructure. The NSF ACCI Task Force on Campus Bridging, members of the committee, and many members of the community produced several reports from 2009 to mid-2011 with findings and recommendations to create such a better national cyberinfrastructure [1, 10, 24, 29]. Part of the challenge was defining campus bridging, which the task force did as follows [24]:

*"Campus bridging is the seamlessly integrated use of cyberinfrastructure operated by a scientist or engineer with other cyberinfrastructure on the scientist's campus, at other campuses, and at the regional, national, and international levels as if they were proximate to the scientist, and when working within the context of a Virtual Organization (VO) make the 'virtual' aspect of the organization irrelevant (or helpful) to the work of the VO."*

More colloquially, the goal is to create "virtual proximity[1]"—everything from a modest local campus cluster to the largest systems should feel as easy to use as a peripheral attached directly to the user's computer. The value of creating virtual proximity in terms of interactions among members of a VO are clear, and a VO that feels "virtually close" while being physically dispersed can be important and beneficial in certain types of medical or civil emergency situations (e.g., Ellisman et al. [11]).

From a national standpoint, a key goal of campus bridging generally is to enable the creation of "comprehensive, integrated, sustainable, and secure cyberinfrastructure to accelerate research and education and new functional capabilities in computational and data-intensive science and engineering"—the core of the NSF Cyberinfrastructure Framework for 21st Century Science and Engineering (CIF21) [23]. This document goes on to say that "Individuals, teams and communities need to be able work together; likewise, instruments, facilities (including MREFCs), datasets, and cyber-services must be integrated from the group to campus to national scale."

The ACCI Task Force on Campus Bridging Final Report and NSF CIF21 vision are necessarily broad and general. Within the eXtreme Science and Engineering Discovery Environment (XSEDE), the Campus Bridging effort is more narrowly focused. The XSEDE Campus Bridging effort examines the XSEDE organization and user communities to identify and prioritize requirements and to help XSEDE put in place services that fulfill those requirements. Most importantly, XSEDE will work with campuses across the country to test, refine, and incorporate the Campus Bridging tools and resources that campuses and their populations find most useful.

Activities related to Campus Bridging are carried out by each and every one of the major organizational subunits of XSEDE. We view Campus Bridging activities as empowering CI users and administrators. Users (researchers, educators, and students) want to be able to use computing resources on and off campus with as little difficulty as possible. Information Technology (IT) administrators want to provide convenient access to both on-campus and off-campus resources. The mission of XSEDE Campus Bridging activities is enable researchers to bridge from personal workstations to labs to campuses to XSEDE, using any

or all in ways that meet the researcher's needs. The XSEDE Campus Bridging effort achieves this goal working within, across, and outside XSEDE to provide information, tools, and services to the XSEDE user community and beyond.

Communicating about campus bridging is different from communicating about supercomputers, data, or visualization systems. It is easy to point to a supercomputer and say, "That's a supercomputer." There is no "campus bridge" to display. Rather, Campus Bridging is an approach that informs a large portion of what XSEDE does and what we in the US cyberinfrastructure and research communities do. This aspect of Campus Bridging is reflected in this paper, which is more progress report and statement of plan than description of technical accomplishments. This focus seems fitting for inclusion in the inaugural XSEDE conference, as the accomplishment of NSF and XSEDE goals related to Campus Bridging require engagement with and coordinated action by the national research community.

## 2. XSEDE AND SYSTEMS ENGINEERING

XSEDE, funded by NSF Award 1053575 [37], is "the most advanced, powerful, and robust collection of integrated advanced digital resources and services in the world. It is a single virtual system that scientists can use to interactively share computing resources, data, and expertise" [39]. The national CI also includes entities called Service Providers (SPs), which may or may not have a relationship to XSEDE. For example, the National Institute for Computational Services (NICS) is an XSEDE Service Provider, and the Kraken supercomputer is one of the resources that NICS provides for use by the XSEDE user community [22]. The Service Provider Forum charter [40] defines three levels of Service Providers, as follows:

- *'Level 1' Service Providers meet all XSEDE integration requirements and will explicitly share digital services with the broader community of researchers supported by the XSEDE environment and infrastructure, with access to those digital services made available through the XSEDE allocation process.*
- *'Level 2' Service Providers make one or more digital services accessible via XSEDE services and interfaces, share one or more digital services with the XSEDE community along with the organization's local users, and meet a subset of the requirements for Level 1 Providers. Level 2 SPs have exposure to a broader audience, can leverage significant XSEDE-provided infrastructure and services in meeting their objectives, and have the option of using the XSEDE allocation process to provide access to the digital services they provide.* (IU's Quarry Virtual Machine hosting service [38] will likely be among the first formally identified Level 2 SP resources.)
- *'Level 3' Service Providers are the most loosely coupled within the XSEDE Federation; they will advertise the characteristics of one or more digital services via XSEDE mechanisms, might make those resources or services accessible via XSEDE compatible interfaces but need not make digital services available 'locally' via XSEDE services or interfaces, and are not required to share services with the XSEDE user community. Most critically, Level 3 SPs are an important way for making the community aware of rich resources that exist beyond the formal structure of the XSEDE program and the services provided by the Level 1 and 2 SPs.* (The Ocean Observatories Initiative [8] CI is likely to be an early Level 3 SP.)

This approach has the impact of creating a set of interface, interaction, and service level agreements between the SPs and

---

[1] Coinage in this context is due to Von Welch, Center for Applied Cybersecurity, Pervasive Technology Institute, IU

XSEDE, and among the SPs themselves. These agreements are defined in greater detail in the Service Provider Forum charter and will continue to evolve over time. We will take *carrying out a computational task* "within XSEDE" to mean carrying out a task at a Level 1, 2, or 3 SP and/or over a network controlled by XSEDE via operational responsibility or lease. The entities that are "bridged to" within XSEDE from campuses are the Level 1, 2, and 3 SPs. "Beyond XSEDE" will mean a computational task done at any other CI facility and across networks not controlled by XSEDE. "Beyond XSEDE" generally constitutes the starting points for things that someone might want to do related to Campus Bridging.

XSEDE has a clear focus on service evolution using a system-engineering approach based on community requirements [3]. XSEDE Principal Investigator, John Towns, has described XSEDE as "an organization that delivers a series of instantiations of services to the US research community." The evolution of idea to software tool or service delivered within one of those instantiations starts with the person or group that has an idea for a service that meets some community need. New ideas are then discussed with the XSEDE Architecture and Design Team. The Architecture and Design Team formalizes software and service concepts, which are passed to the Software Design & Implementation Team and then to the Operations Team. Once a software tool has passed a review by the Operations Team and all documentation and support materials are in place, that tool is ready for deployment. This process is a great way to build a plane, but one challenge facing XSEDE is the need to keep the "plane" flying while it evolves over time. XSEDE has had to do much during its first year that did not easily fit into any standard model of system engineering. Because there were few services related to Campus Bridging in operation at the inception of XSEDE (in spite of capabilities available in the TeraGrid Community Capability Kit [36]), Campus Bridging was viewed as an ideal area to develop using the full-scale XSEDE systems engineering methodology.

The systems-engineering approach to new services includes a requirements database and risk registry, but development of new Campus Bridging services was based primarily on the use cases linked to community requirements drawn from work related to the NSF ACCI Task Force on Campus Bridging [1, 10, 24, 29], other reports on this topic, and survey- and focus-group activities that informed the final XSEDE funding proposal to the NSF [30, 33]. These use cases have been discussed with and confirmed by the XSEDE Campus Champions [41].

# 3. CAMPUS BRIDGING USE CASES
The term "use case," as introduced by Jacobson et al. [19], is the building block that defines a system's functionality. According to Malan and Bredemeyer [21], "use cases capture who (actor) does what (interaction) with the system, for what purpose (goal), without dealing with system internals." Use case descriptions leading to software development are the first in a series of

- a document describing use cases;
- a document mapping use cases to requirements in the requirements database (this is a simple binary mapping; for each use case a "yes" or "no" flag indicates whether a particular requirement is required to enable that use case);
- a set of Level 3 decomposition documents, which include quality-attribute scenarios (describing successful execution of a

use case) [5], state diagrams, and connection diagrams in UML (Unified Modeling Language).

Campus Bridging use case descriptions and quality attribute documents have been completed [31, 32]. In the following sections, we summarize the use cases related to Campus Bridging that have been identified and formalized to date. Many use cases described have to do with the interaction of a user, XSEDE software and services, and services that may be provided by an SP. When relevant, we specify what SP levels are required to fulfill a particular use case. The numbering of these represents the order in time of when we expect to be able to make significant progress in implementing information resources and tools available to the user community. Many of the use cases described are difficult to achieve with current CI software; the challenges, technical and otherwise, are described in greater detail in [31, 32]. Implementing and hardening software tools to enable these use cases, and overcoming other barriers to adoption, will be one among the many important activities of XSEDE overall.

## 3.1 InCommon-based Authentication
**Goal**: *Simplify the authentication process and maintain appropriately high standards of security by adopting InCommon-based authentication mechanisms and SAML certificates for XSEDE resources*. This capability will allow users to access XSEDE resources by authenticating to their home institution so long as their home institution is an InCommon Federation member (via a third-party private provider otherwise). It will allow a user to work with a single identity, rather than managing multiple identities at multiple SPs. This functionality will likely be required of Level 1 and 2 resources, and will be optional but recommended for Level 3 resources. **Progress to date**: XSEDE is now a member of the InCommon Federation, and appropriate software tools (such as CILogon [7]) exist. Guidelines to use InCommon authentication are available online [16].

## 3.2 Information Dissemination
**Goal**: *Enable economies of scale in usability and training for XSEDE and campus resources through dissemination of information*. The XSEDE organization and associated SPs represent one of the largest coordinated sets of investments that the NSF will make in cyberinfrastructure facilities for some time. XSEDE thus has a tremendous opportunity to foster consistency and economies of scale in national research. This use case includes three fairly distinct types of activities:

- Make it easier for on-campus users to use XSEDE resources, and for everyone to create high-quality, reusable training materials. XSEDE can do this by adopting a consistent format for disseminating system information, and then disseminating a template for the administrators of campus CI facilities to use in providing information about those systems.
- Provide a mechanism by which instructors can upload instructional materials to a definitive (and vetted) repository, so that instructors can re-use materials and independent learners can educate themselves.
- Create a "ROCKS Roll" distribution that allows a campus-based systems administrator to install a cluster that includes the open source elements of a basic XSEDE cluster configuration using ROCKS [28].

**Progress to date**: The Texas Advanced Computing Center (TACC) and the Campus Bridging team have drafted a template for defining systems information [20].

One of the most ambitious elements of Campus Bridging planning so far is the idea of a "ROCKS Roll" distribution that would allow a systems administrator to build a cluster that closely matches the open source configuration of a basic XSEDE cluster. ROCKS is a widely used cluster-management tool; the authors know of no open source tool that documents a larger installed base of clusters. We do not expect that a systems administrator would take down a working cluster and rebuild it with ROCKS. However, a cluster administrator might use a ROCKS Roll to set up a new cluster if it allows the administrator to set up a system easily, leverage XSEDE documentation and training materials, and focus her/his own time on local issues such as network performance and file I/O and installing software of interest to local users. XSEDE has allocated funds for one person to work on this project full time during Program Year 2 (2012–2013). The Open Science Grid is adopting an RPM-based model for software distribution. We will monitor their experiences and evaluate our experiences with ROCKS as we develop strategies for tools that satisfy the functional requirement of creating more consistency at a national level by making it easier—when appropriate—for an administrator to set up a campus cluster in a way that is as similar as possible to XSEDE clusters.

## 3.3 Support for Interactive Computing

**Goal**: *Support long-running interactive graphical sessions*. One goal of XSEDE is to broaden the disciplines of inquiry and the usage modalities it supports, particularly interactive computing. A researcher performing a large data analysis or simulation task may want to keep an interactive graphical (e.g., X-Windows) session open for days at a time. **Progress to date**: XSEDE is in the early stages of investigating requirements and possible technical solutions.

## 3.4 Data Transfer

**Goal**: *Use of data resources from campus on XSEDE, or from XSEDE at a campus*. The most common variant of this use case is that a user wants to transfer data from a campus resource to an XSEDE Level 1 or 2 resource for analysis and visualization. Variants of this use case include: XSEDE maintains copies of community and reference data collections for efficiency and to support VOs, or a user identifies a data set to maintain synchronously on a campus resource and one or more XSEDE resources.

**Progress to date**: XSEDE is pursuing two complementary strategies at present. Users will be able to move and synchronize data using the Globus Online file transfer service and also directly access remote data using the Global Federated File System software and service. Enabling both services will likely be a requirement for Level 1 and 2 XSEDE Service Providers.

### 3.4.1 Globus Online
Globus Online enables file movement via a low-overhead implementation of software-as-a-service (SaaS) methods, akin to those used to deliver consumer and commercial IT services such as home movies and accounting. With SaaS, the web browser is the computer: a user points it at the right URL and has immediate access to a powerful interactive software capability. (A useful variant with Globus SaaS is that the user can point his or her SSH Secure Shell client to the right location and start scripting.)

The Globus Transfer [4, 12] SaaS data-movement service makes it trivial for users to move and synchronize data among desktop computers, campus HPC resources, and national facilities. The user registers with Globus Online, and can then immediately start using a web browser (or SSH client) to request, monitor, and control data transfers. Most national computing facilities, including XSEDE Service Providers, already run the necessary software required to transfer data to and from those facilities. To move data to/from their desktop, users can quickly and easily download and install the Globus Connect agent for Linux, Mac, or Windows. Globus Online removes all of the complexities of transferring large amounts of data, such as automatically tuning the use of the network for high performance, automatic recovery from common network and server failures, and navigating security domains and firewalls.

For administrators, Globus Transfer makes it easy to integrate a new resource, such as a campus HPC system, into the national cyberinfrastructure. Downloading and installing the Globus Connect Multi User (GCMU) software, which integrates GridFTP and MyProxy to enable convenient integration with campus storage and authentication services, is a 30-minute task. The resource is then accessible as a Globus Transfer endpoint, which can be assigned a meaningful symbolic name.

Experiences deploying Globus Transfer on campuses have been positive. From GCMU's release in October 2011 to April 2012, the number of deployments has grown to 24 on 15 campuses. At the University of Colorado Boulder, for example, Globus Online has been used to move more than 100 terabytes of data to and from their Janus HPC system. A recent paper [13] provides more details on how Globus services support Campus Bridging.

### 3.4.2 Global Federated File System
The Global Federated File System (GFFS) provides a single uniform mechanism to integrate and access resources whether those resources are at an NSF supercomputer center, on a campus, in an individual research lab, or on a home computer. GFFS is designed to enable end-user autonomy and control by virtue of running within user space on a computer (either personal computer or shared system) using FUSE (Filesystem in USErspace) [2]. This allows an end user the opportunity to create a geographically distributed file system without any requirement to have system privileges on any system participating in the distributed file system.

GFFS links resources into a global path-based name space so that they can be accessed from anywhere in a location-, migration-, and replication-transparent fashion. The path-based namespace can be mapped into the local operating system as a file system mount. The user can then use that mount like any other file system mount. He or she can run applications and shell scripts that create, read, update, and delete files in GFFS without any program modification; start and manage running jobs using the file system abstraction; and interact with the working directory of a running job without knowing even where the job is running.

For example, suppose that a research group leader keeps her sources and scripts in the directory */home/ME/sources* on a departmental file server. The researcher can then *"export"* */home/ME/sources* into GFFS and access it via the FUSE driver in scripts or at the command line from any of the service providers. Any changes made to the files, either at the researcher's home institution or at any XSEDE Service Provider, will be immediately visible to everyone authorized to access those files, including her own jobs and scripts running at her home institution and at XSEDE SPs. This is shown diagrammatically in Figure 1.
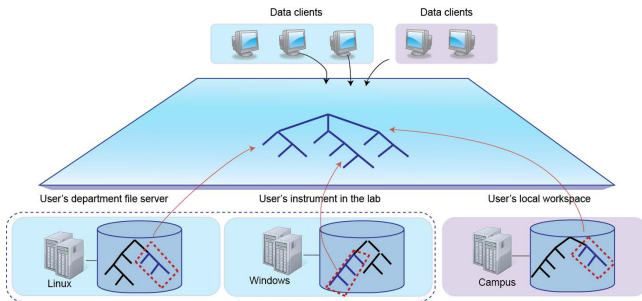
**Figure 1. A user exports data from department's file system, instrument, and from local home directory into the global namespace. The directories and files are then accessible from data clients (including their own campus computers) throughout GFFS. Directories and files are linked from the source location to the GFFS directory and user-specified name. This presents a unified view of the data across platforms, locations, domains, etc. The user controls the authorization policy.**

## 3.5 Distributed Workflows

**Goal**: *Support for distributed workflows spanning XSEDE and campus-based data, computational, and visualization resources.* This use case is the "holy grail" of campus bridging—to support complex scientific workflows that operate as if there were no boundaries between resources within and outside XSEDE.

The Linked Environments for Atmospheric Discovery—which allows on-the-fly composition of workflows to predict weather—is an example of a workflow using a remote data source, a campus-based computational system, XSEDE computational and storage systems, and a local visualization facility [14]. A diagram of this type of workflow spanning TeraGrid and non-TeraGrid resources is presented in Stewart et al. [34] and depicted graphically in Figure 2. To take advantage of the proliferation of
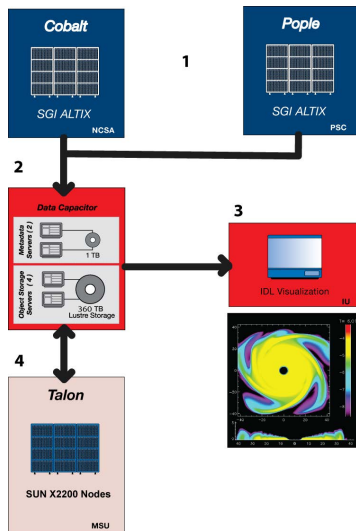


**Figure 2. An astronomer (3) views the results of simulations occurring at PSC's Pople and NCSA's Cobalt supercomputers (1) (1) as they are written to IU's Data Capacitor (2). Further analysis of the data that have been written takes place on MSU's Talon supercomputer (4). From Stewart et al [31].**

digital data sources that exist beyond XSEDE's electronic borders and to apply XSEDE's computational capabilities to the analysis

of data from these sources, it will be necessary to support this sort of complex workflow seamlessly in a way that takes advantage of many of the capabilities mentioned in prior use cases (including implementation of authentication systems and movement of data in ways that facilitate these workflows). While XSEDE may choose to focus on one or more workflow systems that it recommends and supports, these systems generally do not require support at the Service Providers, so there are no particular requirements for Level 1, 2 or 3 SPs. **Progress to date**: XSEDE continues to support several workflow systems that were supported by TeraGrid in years past.

## 3.6 Shared Computational Facilities

**Goal**: *Shared use of computational facilities mediated or facilitated by XSEDE.* This use case has a great deal to do with addressing the challenge of using well the nation's aggregate CI. A key issue here is facilitation of collaboration within a virtual organization (VO) without direct engagement with XSEDE. A VO may wish to use a set of computational facilities in concert as a shared virtual compute facility (SVCF). The VO wants to control access, usage policies, etc., but wishes to use tools provided by and certified by XSEDE and to leverage XSEDE credentials to operate this facility. Such an SVCF might consist of multiple compute clusters or might be a high-throughput computing (HTC) facility. SURAGrid [35] is a good example of a virtual organization sharing cluster facilities. The most longstanding examples of SVCFs are in HTC, based on Condor and Globus [25, 26]. We discuss below the largest and most well established—the Open Science Grid (OSG). Many campuses or groups of campuses operate campus-based Condor flocks, such as the Purdue-led DIAGRID [27].

The above examples involve individuals or VOs working together independently of XSEDE, leveraging trust relationships within the VO, and using software provided by XSEDE or by organizations partnered with XSEDE. Such scenarios have a direct benefit to XSEDE and the national research community even though they may not directly involve use of a resource at an XSEDE SP. The more effectively the US research community uses resources outside of XSEDE, the more XSEDE capacity and capability will be available to focus on computational tasks that require features not available from other facilities. The recognition that some SVCFs may want to operate essentially independently of XSEDE while using tools provided by XSEDE is important in helping XSEDE deliver tools that will meet the VOs' diverse organizational structures and needs.

A variant of this use case involves a VO working in collaboration with XSEDE, accepting jobs from users who have accounts and allocations on XSEDE. In such a case, the VO is helping XSEDE fulfill a commitment made through the allocation process to an XSEDE user, and is accepting the XSEDE credentials presented by such users. In turn, the operators of such a VO would want to be able to run their own jobs on XSEDE at some other time. This requires the ability to rationally relate the computing time on the VO's facility and the XSEDE facility where credits are 'recouped'—something that can be done reasonably well in many cases.

**Progress to date**: OSG is one of the first, most important, and most widely adopted solutions to the challenge of campus bridging today. OSG enables campus bridging in support of distributed high-throughput computing through Condor Flocking and resource-management overlays that are deployed by the OSG

GlideIn Factory service. This service is the current XSEDE solution to distribution of tools for the creation of SVCFs for high-throughput computing. Also, OSG has been working with Globus Online to make OSG and Globus Online services visible to campuses through a coordinated framework. The Campus Bridging team will continue to support the work of the Condor group and OSG, promoting OSG's efforts in this area. The creation of other sorts of *virtual clusters* is possible with tools currently being considered for deployment. Using the XSEDE CI tools will enable the creation of a virtual cluster—composed of multiple physical compute clusters—and a unified virtual job queue. Such an approach may well involve a loss of efficiency from a computational standpoint. However, the creation of such a virtual cluster and management of job throughput may result in more effective overall handling of computational work, in terms of supporting scientific research, than a set of isolated servers. OSG is now an XSEDE SP, and is being compensated in ways other than *cycle exchange* for handling jobs from XSEDE allocations. Many policy and accounting challenges remain to be solved before cycle exchange agreements can be implemented with other VOs. This would be a means by which XSEDE could greatly increase the effectiveness with which US aggregate CI resources are used.

## 3.7 Services On Demand

**Goal**: *Access to resources on a service-for-money basis ("on demand").* Bridging from XSEDE to private service offerings available from cloud providers such as Google, Microsoft, and Amazon, as well as Internet2 NET+ services. Cloud providers address particular and important computational requirements that are not necessarily good architectural matches for current XSEDE resources, and the overall distribution of computational capacity in the US no longer resembles a Branscomb pyramid [24]. **Progress to date**: Two "on demand" services will soon be identified as XSEDE SPs: Cornell's Red Cloud service [9] and the IU / Penguin "Penguin on Demand" service [17].

## 3.8 Trouble Ticket Systems

**Goal**: *Coordinated flow of trouble ticket information between campus and XSEDE IT staff.* This use case is different from those mentioned already, in that it is really a general use case and need for XSEDE as a whole. It is included here because it has particular relevance to Campus Bridging. Many of the examples of Campus Bridging described so far involve integrated use of XSEDE resources and campus-based resources. When from the user's perspective something fails, the failure may originate with the user, at a campus resource, or at an XSEDE resource. The user neither knows nor cares where the failure occurred; he or she just wants to get the work done. Providing excellent customer service will require the ability to update trouble tickets and transfer ownership of tickets between XSEDE and local trouble ticket systems. In that way, all of the people involved in solving a user's problem can communicate effectively with each other and with the user, to produce in the end a solved problem and a user satisfied.. **Progress to date**: XSEDE is making a decision on a trouble ticket system for XSEDE. The tool selected should include capabilities to exchange tickets among multiple ticket systems.

## 4. PILOTS AND EVALUATION

Campus Bridging is a new way of thinking about the broader mission of XSEDE and about a set of particular technology implementation goals. Campus Bridging activities thus operate within the systems-engineering processes of XSEDE, and Campus Bridging software will be put into use only after a successful Operational Readiness Review. Because many activities that XSEDE supports via Campus Bridging will often interact with CI resources beyond XSEDE, which have not gone through the quality assurance and operational reviews that will be typical of XSEDE, there is more uncertainty about how these tools will work in practice. As it was put more than a hundred years ago, "No plan of operations extends with certainty beyond the first contact with the enemy's main strength" [15]. In this case the enemy is the relatively fragile and complex state of much scientific middleware.

To test Campus Bridging technology being developed to address the use cases described here, XSEDE is engaged in four pilot implementations at US universities. The XSEDE Campus Bridging team received a total of 17 pilot project proposals. All proposals were of such high quality that XSEDE selected four projects to proceed as pilots based on the match between the proposed activities and the capabilities of the tools we have ready for testing. The pilot projects are

- Texas A&M University: using GFFS to move between users on campus and the Brazos file systems and to the TACC Ranger system.
- University of Kansas: using GFFS for cosmology, molecular modeling, and polar research to share data between KU, NCSA, Purdue, and Indiana University.
- University of Miami: using GFFS to simplify data transfer between Miami and XSEDE resources as well as data sharing within UM.
- City University of New York: using GFFS to facilitate researcher use of CUNY Center resources. Data will be shared via GFFS between the CUNY Center, College of Staten Island, Miami, and Delaware.

A report outlining outcomes and lessons learned from the pilot projects will be published when all four projects are completed.

## 5. COMMUNITY INVOLVEMENT

The XSEDE vision for Campus Bridging will be achieved only if we have significant, active engagement from the US research community. The need for engagement and the crosscutting nature of the activity within XSEDE are factors that have led to including Campus Bridging as part of the XSEDE Training, Education, and Outreach Services group. There are considerable interactions and ongoing collaboration with Campus Champions [41]. Thus, one way to be involved is to become a Campus Champion or to work with your campus's Champion if you already have one. There is also an online forum [43] monitored by Campus Champions; anyone with an XSEDE portal login is welcome to post questions, comments, or suggestions [42]. As information resources and tools related to Campus Bridging are released, there will be notifications through XSEDE communication channels including the portal and electronic newsletters.

As is the case with any new concept or phrase, the use of the term *Campus Bridging* within XSEDE is evolving. Throughout this discussion we have used the phrases "within XSEDE" and "beyond XSEDE" and with the use cases explained should return to that point. As mentioned earlier, one way to think of what computing "within XSEDE" means is "at a Level 1, 2, or 3 resource as defined in the Service Provider Forum Charter" [40]. Taking this view, a Shared Virtual Compute Facility operated under a cycle-exchange agreement with XSEDE would be

consistent with such a SVCF being a Level 3 SP, and this would in some sense be "within XSEDE." But a SVCF might be operated entirely independently of XSEDE, clearly "beyond" XSEDE, and still yet be of great benefit to the overall XSEDE mission and to the US research community. In general, it probably leads to less confusion to talk of resources with the Level 1, 2, and 3 terminology defined in the SP Forum document, or to talk in specific terms such as "Shared Virtual Compute Facility," than to talk about a "bridged campus" or a "bridged cluster." In the same sense, we know what it means when we say an individual is a Campus Champion. It probably does not make sense to talk about a "Campus Bridger" in the same way. Some observers have commented that "researcher bridging" might be a better term than "campus bridging." While a good observation, the NSF ACCI Task Force on Campus Bridging has already created significant name recognition for this term, and trying to change it is likely to create more confusion rather than less. If campus, state, regional, and national CI resources are well connected, then there is a bridge from campus to national resources—easy for the researcher to use and aiding US global competitiveness.

# 6. EVALUATION AND METRICS

XSEDE will measure progress of its efforts in Campus Bridging overall and each of the use cases described here. We plan to work closely with the OSG in developing and refining metrics, as OSG has the longest history of delivering campus bridging services of any SP affiliated with XSEDE. Metrics are still being developed, and will include

- The number of campuses that adopt and use Campus Bridging tools developed by XSEDE.
- The number of campus resources made more readily accessible to the national community as a result of implementation of Campus Bridging tools.
- The number of classes taught, and number of students taught, on campuses using XSEDE-generated class materials.
- The number and extent to which XSEDE-developed Campus Bridging tools are adopted among other CI projects.

Overall, the most critical long-term metric is whether XSEDE Campus Bridging efforts increase the value of US CI in meeting the needs of US researchers. This metric will be evaluated through annual surveys of NSF-funded PIs (and select PIs funded by other federal agencies) and XSEDE users. We will continually evaluate community needs and requests, being cognizant of related activities such as Internet2 NET+ [18], which supports higher-education access to general-purpose tools such as box.net.

# 7. PLAN AND TIMELINE

The use cases described in this paper constitute the initial key goals to address the recommendations of the ACCI report and the needs of the XSEDE community. The system description templates now available [20] constitute the first XSEDE deliverables related to Campus Bridging.. The distribution of a ROCKS Rolls for a 'generic XSEDE-like cluster' will be one of the Program Year 2 deliverables. The XSEDE groups involved in Campus Bridging will develop a detailed plan to address as many use cases as possible and route that plan through the XSEDE governance processes, so that we can publicize a four-year plan for Campus Bridging deliverables.

# 8. CONCLUSION

Campus bridging is a new concept, with a set of recommendations to the NSF and nation codified in a report issued just over a year ago. XSEDE efforts in the general area of Campus Bridging are designed to aid US research effectiveness, accelerate discovery, and meet NSF goals set out in the XD solicitation and in CIF21. There are many activities that one could label as Campus Bridging. In this paper, we have set out eight use cases representing high-priority needs of XSEDE users and the national science, technology, engineering, and mathematics research and education community. XSEDE is funded for at least another four years as of this writing. If in those four years XSEDE and the US research community are able to work together and implement tools that enable the use cases identified in this document, we shall have done a great deal to aid US innovation and education.

# 9. ACKNOWLEDGMENTS

# 10. REFERENCES

1. Cyberinfrastructure Software Sustainability and Reusability Workshop Final Report, 2010. [Accessed 13 June 2012]; Available from: http://hdl.handle.net/2022/6701 or https://www.createspace.com/3506064.

2. FUSE: Filesystem in Userspace. [Accessed 22 Apr 2011]; Available from: http://fuse.sourceforge.net/

3. Akli, L. and Crane, G. The Dawn of XSEDE: An Overview, 2011. [Accessed 22 Apr 2011]; Available from: www.suragrid.org/docs/TG-to-**XSEDE**-SURA.pdf

4. Allen, B., Bresnahan, J., Childers, L., Foster, I., Kandaswamy, G., Kettimuthu, R., Kordas, J., Link, M., Martin, S., Pickett, K. and Tuecke, S. Software as a Service for Data Scientists. *Communications of the ACM*, 55(2):81–88, 2012.

5. Bass, L., Clements, P. and Kazman, R. *Software Architecture in Practice*. Addison-Wesley, 2003.

6. Catlett, C. et al.. TeraGrid: Analysis of Organization, System Architecture, and Middleware Enabling New Types of Applications. In Grandinetti, L. ed. *High Performance Computing and Grids in Action*, pp. 225–249. IOS Press, 2007.

7. CILogon Service. Home page. [Accessed 31 Jan 2011]; Available from: https://cilogon.org.

8. Consortium for Ocean Leadership. Ocean Observatories Initiative. [Accessed 31 Jan 2011]; Available from: http://www.oceanleadership.org/programs-and-partnerships/ocean-observing/ooi/.

9. Cornell University Center for Advanced Computing. Red Cloud. [Accessed 23 May 2012]; Available from: http://www.cac.cornell.edu/redcloud/.

10. Dreher, P., Agarwala, V., Ahalt, S.C., Almes, G., Fratkin, S., Hauser, T., Odegard, J., Pepin, J. and Stewart, C.A. Developing a Coherent Cyberinfrastructure from Local Campuses to National Facilities: Challenges and Strategies. *EDUCAUSE*, 2009.

11. Ellisman, M., Brady, M., Hart, D., Lin, F.-P., Müller, M. and Smarr, L. The Emerging Role of Biogrids. *Communications of the ACM–Bioinformatics*, 47(11):52–57, 2004.

12. Foster, I. Globus Online: Accelerating and Democratizing Science through Cloud-Based Services. *IEEE Internet Computing,* (May/June):70–73, 2011.

13. Foster, I. Kettimuthu, R., Martin, S., Tuecke, S., Hauser, T., Milroy, D., Palen, B. and Braden, J. Campus Bridging Made Easy via Globus Services. XSEDE 2012, Chicago, IL, USA, 2012.

14. Gannon, D., Plale, B., Marru, S., Kandaswamy, G., Simmhan, Y. and Shirasuna, S. Dynamic, Adaptive Workflows for Mesoscale Meteorology. In Taylor, I.J., Deelman, E., Gannon, D.B. and Shields, M. eds. *Workflows for eScience* (Chapter 10), Springer Verlag, 2007.

15. Hughes, D.J. *Moltke on the Art of War: Selected Writings*. Random House Ballantine, 1993.

16. InCommon Federation. A Roadmap for Using NSF Cyberinfrastructure with InCommon. [Accessed 13 June 2012]; Available from http://www.incommonfederation.org/cyberroadmap.html.

17. Indiana University Pervasive Technology Institute. Rockhopper. [Accessed 23 May 2012]; Available from: http://pti.iu.edu/ci/systems/rockhopper.

18. Internet2. Internet2 NET+. [Accessed 12 Feb 2012]; Available from: http://www.internet2.edu/netplus/.

19. Jacobson, I., Christerson, M., Jonsson, P. and Overgaard, G. *Object-Oriented Software Engineering: A Use Case Driven Approach*. Addison-Wesley, 1992.

20. Lindsey, S. and Dahan, M. 2012. Digital service user guides. Accessed 13 June 2012]; Available from: https://portal.xsede.org/knowledge-base/-/kb/document/bccy

21. Malan, R. and Bredemeyer, D. Functional Requirements and Use Cases, 2001. [Accessed 13 June 2012]; Available from: http://www.bredemeyer.com/pdf_files/functreq.pdf.

22. National Institute of Computational Sciences. Kraken. [Accessed 31 Jan 2011]; Available from: http://www.nics.tennessee.edu/computing-resources/kraken.

23. National Science Foundation. Cyberinfrastructure Framework for 21st Century Science and Engineering (CIF21) 2011 [Accessed 17 Nov 2011]; Available from: http://www.nsf.gov/funding/pgm_summ.jsp?pims_id=504730.

24. NSF Advisory Committee for Cyberinfrastructure Task Force on Campus Bridging. Final Report, 2011. [Accessed 13 June 2012]; Available from: http://www.nsf.gov/od/oci/taskforces/TaskForceReport_CampusBridging.pdf or http://pti.iu.edu/campusbridging/.

25. Open Science Grid. Home page. [Accessed 31 Jan 2011]; Available from: http://www.opensciencegrid.org/.

26. Pordes, R., Petravick, D., Kramer, B., Olson, D., Livny, M., Roy, A., Avery, P., Blackburn, K., Wenaus, T., Würthwein, F., Foster, I., Gardner, R., Wilde, M., Blatecky, A., McGee, J. and Quick, R. The Open Science Grid. Scientific Discovery through Advanced Computing (SciDAC) Conference, 2007.

27. Purdue University. DiaGrid. [Accessed 1 May 2012]; Available from: https://diagrid.org/.

28. Rocks Clusters. Home page. [Accessed 31 Jan 2011]; Available from: http://www.rocksclusters.org/.

29. Stewart, C.A. Progress Report on Implementation of Recommendations from the Indiana University Cyberinfrastructure Research Taskforce, 2010. [Accessed 13 June 2012]; Available from: http://hdl.handle.net/2022/8974.

30. Stewart, C.A., Katz, D.S., Hart, D.L., Lantrip, D., McCaulay, D.S. and Moore, R.L. Technical Report: Survey of Cyberinfrastructure Needs and Interests of NSF-Funded Principal Investigators, 2011. [Accessed 13 June 2012]; Available from: http://hdl.handle.net/2022/9917.

31. Stewart, C.A., Knepper, R., Grimshaw, A., Foster, I., Bachmann, F., Lifka, D., Riedel, M. and Tuecke, S. Campus Bridging Use Case Quality Attribute Scenarios, 2012. [Accessed 13 June 2012]; Available from: http://hdl.handle.net/2022/14476.

32. Stewart, C.A., Knepper, R., Grimshaw, A., Foster, I., Bachmann, F., Lifka, D., Riedel, M. and Tuecke, S. XSEDE Campus Bridging Use Cases, 2012. [Accessed 13 June 2012]; Available from: http://hdl.handle.net/2022/14475.

33. Stewart, C.A., McCaulay, D.S., Jul, S. and Moore, R.L. Technical Report: TeraGrid eXtreme Digital Campus Bridging Requirements Elicitation Meeting, 2011. [Accessed 13 June 2012]; Available from: http://hdl.handle.net/2022/13026.

34. Stewart, C.A., Simms, S., Plale, B., Link, M., Hancock, D. and Fox, G. What is Cyberinfrastructure? SIGUCCS 2010, Norfolk, VA, 2010.

35. SURAgrid. About SURAgrid. [Accessed 8 Apr 2011]; Available from: http://www.sura.org/programs/sura_grid.html.

36. TeraGrid. TeraGrid Community Capability Registration Guidelines. [Accessed 1 May 2012]; Available from: http://dms.teragrid.org/mediawiki/index.php?title=TeraGrid_Community_Capability_Registration_Guidelines.

37. Towns, J., Roskies, R., Boisseau, J., Kovatch, P. and Andrews, P. XSEDE: eXtreme Science and Engineering Discovery Environment. National Science Foundation, 2011.

38. XSEDE. IU Quarry User Guide. [Accessed 13 June 2012]; Available from https://www.xsede.org/iu-quarry.

39. XSEDE. Overview. [Accessed 23 May 2012]; Available from: https://www.xsede.org/overview.

40. XSEDE. XD Service Providers Forum: Charter, Membership, and Governance, 2012. [Accessed 13 June 2012]; Available from: https://www.xsede.org/documents/10157/281380/SPF_Definition_v10.1_120228.pdf.

41. XSEDE. XSEDE Campus Champions. [Accessed 20 Feb 2012]; Available from: https://www.xsede.org/campus-champions.

42. XSEDE. XSEDE User Portal. [Accessed 23 May 2012]; Available from: https://portal.xsede.org/.

43. XSEDE. XSEDE User Portal: Forums. [Accessed 23 May 2012]; Available from: https://portal.xsede.org/forums.