Cyberinfrastructure as a platform to facilitate effective collaboration between institutions and support collaboratories

Eric Coulter
Indiana University Pervasive
Technology Institute
2709 E 10th Street
Bloomington, IN 47408
(812) 856-3250
jecoulte@iu.edu

Jeremy Fischer
IUPTI
2709 E 10th Street
Bloomington, IN 47408
(812) 856-0992
jeremy@iu.edu

Barbara Hallock
IUPTI
2709 E 10th Street
Bloomington, IN 47408
(812) 856-2364
bahalloc@iu.edu

Richard Knepper
IUPTI
2709 E 10th Street
Bloomington, IN 47408
(812) 855-9574
rknepper@iu.edu

Dave Lifka
Cornell University
512 Frank H. T. Rhodes Hall
Ithaca, NY 14853
(607) 254-8621
lifka@cornell.edu

JP Navarro
Argonne National Laboratory
Mathematics and Computer Science
Division, TCS Building 240
Lemont, IL 60439
(630) 252-1233
navarro@mcs.anl.gov

Marlon Pierce
Science Gateways Research Center /
IUPTI
2709 E 10th Street
Bloomington, IN 47408
(812) 856-1212
marpierc@iu.edu

Craig Stewart
IUPTI
2709 E 10th Street
Bloomington, IN 47408
(812) 855-4240
stewart@iu.edu

ABSTRACT

Researchers, scientists, engineers, granting agencies, and increasingly complex research problems have given rise to the scientific "collaboratory"—large organizations that span many institutions, with individual members working together to explore a particular phenomenon. These organizations require computational resources in order to support analyses and to provide platforms where the collaborators can interact. The XSEDE Community Infrastructure (XCI) group assists campuses in using their own resources and promotes the sharing of those resources in order to create collaboratories improving use of the nation's collective cyberinfrastructure. Currently XCI provides toolkits and training, and collaborates with organizations such as ACI-REF, XSEDE Campus Champions, and the Open Science Grid to identify tools and best practices that support the community. This paper discusses the progress in and barriers to developing a robust collaborative environment where computational resources can be shared.

Categories and Subject Descriptors

• **Information systems~Enterprise applications.**

General Terms

Design, Economics, Reliability, Human Factors, Standardization.

Keywords

XCI, XSEDE, Collaboratories, ACI-REF, Open Science Grid, Computation Institute, Globus, Cluster Computing, Collaboration, Big Data, Jetstream, Science Gateways

1. INTRODUCTION

Collaboration is increasingly important to the way that academic work gets done. Nowhere is this more pronounced than in the sciences, where the complexity of certain instruments and the scarcity of funding necessitate that multiple organizations collaborate in order to make the most effective use of the resources available. Additionally, greater transparency in research is driving the need for increased access to the data on which scientific discoveries are predicated, and the amount and variety

of such data are growing exponentially as computational capabilities and the complexity of the questions researchers seek to answer increase. Collaboration between institutions also fosters greater understanding, as organizations share insights.

Common tools such as email and file sharing meet many of the basic requirements for effective collaboration. Given the ubiquity of these technologies and their limited capabilities for advanced application, we instead focus this paper on the needs of scientists that go beyond what these technologies can provide. These needs are encompassed in the term cyberinfrastructure, a term popularized by the National Science Foundation (NSF) and now commonly defined in a fashion similar to the following [1]:

> *Cyberinfrastructure consists of computing systems, data storage systems, advanced instruments and data repositories, visualization environments, and people, all linked together by software and high performance networks to improve research productivity and enable breakthroughs not otherwise possible.*

The NSF has prioritized the development of cyberinfrastructure to facilitate scientific collaboration among universities [2] and make collaboratories as effective and innovative as possible.

In this paper we briefly describe several major U,S, cyberinfrastructure projects that aim to aid the operation of collaboratories. We then discuss current efforts designed to improve the effectiveness of cyberinfrastructure services in the support of collaboration and the formation of collaboratories. We discuss the impact of these innovations on collaborative science as well as some of the needs for future development in order to support these activities more fully.

2. COLLABORATORIES

A collaboratory is a specific type of virtual organization—that is, a set of people with a common goal who are connected by a shared cyberinfrastructure need [3]—in which that specific need centers on the facilitation and completion of scientific experimentation. Scientific and technological advances in recent history have meant that scientific work is increasingly being done by groups of researchers in geographically disparate locations. Hence, some expansion on the traditional notion of the "laboratory" becomes necessary [4].

The collaboratory as a concept is a blending of "collaboration" and "laboratory"—"a center without walls, in which researchers can perform their research without regard to physical location-interacting with colleagues, accessing instrumentation, sharing data and computational resources, and accessing information in digital libraries" [3]. The benefits of the collaboratory are multiple and varied, depending on the goals and resources of the organizations participating. Most important, as scientific advances increasingly depend on large teams of experts with diverse backgrounds and expertise, supporting these scientific teams with collaboratories becomes increasingly important. This is not meant to imply that there no brilliant individual investigators are doing important and groundbreaking work; but most new insights today are the result of large collaborative teams.

A number of technologies have been key to the formation of collaboratories; perhaps the most fundamental ones are electronic mail and networking. Since these technologies are well established and unchanging, however, much of the body of the work presented here focuses instead on other resources that are being actively developed by the organizations described.

2.1 Examples of Collaboratories
Two of the largest cyberinfrastructure projects supporting collaboratories are the Open Science Grid and the eXtreme Science and Engineering Discovery Environment (XSEDE). These two cyberinfrastructure projects, and how they support the operation of collaboratories, are described below.

2.1.1 Large Hadron Collider and Open Science Grid
Perhaps the foremost example in modern science of collaborative work among multiple groups brought together by the need to share expensive instruments is the Large Hadron Collider (LHC). The world's largest machine, the LHC is a 27 km ring with four separate particle detectors situated at points around the circle: ATLAS, CMS, ALICE, and LHCb, each with its own set of tasks [5]. A notable result of this massive collaborative effort of over thirty years has been the detection of a particle consistent with the Higgs boson predicted by the Standard Model, which resulted in the joint award of the 2013 Nobel Prize in Physics to François Englert and Peter Higgs [6]. Because of the one-of-a-kind nature of the LHC and the importance of reproducibility in determining the viability of experimental results, the LHC also fits the "data" component of a collaboratory—with estimates in the range of 25 GB/s during run 2 [7]. The Open Science Grid (OSG) [8] is a computational, data analysis grid specifically designed to analyze the data created by the LHC. The OSG grew out of a set of other grid research projects related to physics experiments. It has proved particularly well suited for analyzing data from the LHC for several reasons. First and foremost the nature of the LHC data and the design of the OSG made it possible to ship data from centralized data repositories to distributed computational resources over long-haul networks in a way that well matched the amount of data shipped to a remote computing environment and the amount of computing to be done, managing a ratio of the two that was appropriate for long-haul networks of typical performance characteristics.

The OSG organizational model is based specifically on sharing resources. These resources include compute, core operational infrastructures, and networks of resource and user communities. OSG functions as a set of virtual organizations, or collaboratories, that tself-organize their support for collaborative activities. This philosophical approach has proven reliable, as the OSG has scaled to many institutions and user disciplines. The core OSG VO administers and supports smaller resource and user bases that do not have effort to maintain a full VO infrastructure. The number and activity of VOs working in other disciplines is now expanding as well, with a total of 89 VOs currently making active use of the OSG [9, 10].

2.1.2 The eXtreme Science and Engineering Discovery Environment (XSEDE)
XSEDE supports more than a dozen supercomputers and high-end visualization and data analysis resources across the country [11]. Funded by the National Science Foundation (NSF) as the successor to a previous project known as the TeraGrid, XSEDE could be considered a collaboratory in the business of supporting other collaboratories. Work on the bleeding edge of science has become increasingly computationally intensive as technology has improved; indeed, high-performance computing (HPC) is one of the few areas where supply seems to always increase demand, because the greater the capability and capacity of nationally accessible CI for research, the larger the data analyses that

become possible and the more realistic become massive computer simulations of physical, astronomical, and biological phenomena.

XSEDE provides access to cyberinfrastructure in support of collaboratories, by integrating software and services that enable the infrastructure integration and that lower the bar for users to request access to and then use the integrated infrastructure. The ECSS (Extended Collaborative Support Services) unit of XSEDE brings together cyberinfrastructure experts with a wide variety of expertise to provide support for researchers, who can request ECSS services along with their allocation requests [12].

A particularly important aspect of XSEDE is the Campus Champions [13]. As their web page states, "The Campus Champions program supports campus representatives as a local source of knowledge about high-performance and high-throughput computing and other digital services, opportunities and resources. This knowledge and assistance empowers campus researchers, educators, and students to advance scientific discovery." Campus Champions provide essential human linkages connecting XSEDE as a resource provider and "community collaborator" and individuals on campuses across the United States.

Given the relatively high cost of acquiring and running the resources made available through and supported by XSEDE, one can easily understand the value of the collaborative approach and the value that XSEDE offers to collaboratories.

3. FACILITATING THE CREATION AND EFFICACY OF COLLABORATORIES

In 2011 a taskforce of the NSF Advisory Committee on Cyberinfrastructure concluded that "the current state of cyberinfrastructure software and current levels of expert support for use of cyberinfrastructure create barriers in use of the many and varied campus and national cyberinfrastructure facilities. These barriers prevent the US open science and engineering research community from using the existing, open US cyberinfrastructure as effectively and efficiently as possible" [1].

In retrospect, the chair of that committee—coauthor Craig A. Stewart—believes that this task force missed an important opportunity to communicate information about cyberinfrastructure and human resources available to enhance the efficacy of collaboratories.

Current social science understanding of technology adoption suggests adoption decisions are driven by the following [14]:

- Performance expectancy (perceived value)
- Effort expectancy (perceived ease of use)
- Social influence
- Facilitating conditions (including knowledge of a technology and the belief that end users will find it accessible)

Social influences and facilitating conditions largely involve informing people about what technology choices are available for adoption—an essential prerequisite to determining whether a particular bit of technology offers high value relative to any potential difficulties of adoption.

3.1 The Computation Institute and Globus

The Computation Institute (CI) (affiliated with the University of Chicago and Argonne National Laboratory) has developed over many years a set of tools designed to facilitate collaboration. The Globus suite of tools has evolved to be been of great use in facilitating collaboration between geographically disparate sites.

Some of the most popular Globus tools supporting collaborators are the following:

Globus Transfer. Globus Transfer is an easy-to use, fast, secure client that allows users to move large amounts of data without having to attend to the entire process. Instead, Globus Transfer operates under a "set it and forget it" paradigm: the user simply starts the file transfer and waits for an email stating that the transfer has finished. The Globus Transfer software uses a graphical user interface, rather than requiring users to learn to operate in a command line paradigm [15].

File sharing. Globus offers a file-sharing solution that allows users to configure shared endpoints in order to share large files with others securely and without having to move the data to a new location—something that can save significant time in the case of large datasets. Any storage system that is configured as a Globus endpoint can easily be configured to use the file-sharing capabilities. In addition to the cost savings and security advantages of being able to determine which files are and are not sharable, another advantage of the Globus model of file sharing is that it does not require collaborators from other institutions to have login access to the storage system in order to access files [16].

Data publication. Globus Data Publication allows organizations to publish datasets; metadata are stored in the cloud, while datasets are stored on campus, institutional, or group resources. Datasets may be curated into collections, and the Globus Data Publication platform enables access controls based on both user and group membership. The community that publishes the data can define their own workflows and licensing terms; published data is then searchable via the web interface. Once published, data may be transferred to a Globus endpoint for further processing [17].

3.2 XSEDE Community Infrastructure (XCI)

The XCI team grew out of the XSEDE organization and the 2011 report of the NSF ACCI Task Force on Campus Bridging [2]. During the first five years of XSEDE, what was then called the XSEDE Campus Bridging team focused primarily on "bridging" the gap between researchers and cyberinfrastructure. Created as a response to findings that even well-resourced researchers had difficulty completing the journey to results, the XCI team takes a number of approaches to making collaboration over a distance easier.

The XSEDE National Interoperability Toolkit is a software repository provided by the XCI group in order to facilitate a network of XSEDE-like collaboratories. Aimed at campuses or institutions that host their own computational resources, the XNIT offers the easy installation of a variety of scientific software in the same configuration as on the larger XSEDE Service Providers. This allows institutions with their own already-running computational resources to easily help researchers scale up to the resources available within XSEDE, to help researchers be certain they are using a tested version of their software, and to remove some of the burden on local system administrators. The XCI team works to foster communication among the communities using (or interested in) this resource, through the use of quarterly webinars, regular appearances at conferences, and communication via the XSEDE Campus Champions mailing list.

The XSEDE-Compatible Basic Cluster is a second toolkit championed by the XCI group, aimed at institutions that want their own computational resources but do not necessarily have the local expertise or time to build their own from scratch. The XCBC consists of the Rocks cluster management software (developed at

the San Diego Supercomputer Center, for use on the large XSEDE resources hosted there), and an additional package of scientific software, similar to that found in the XNIT, that has been selected to work within the Rocks system. The XCBC has been found to be valuable particularly at resource-constrained institutions, where local system administrators may be unfamiliar with HPC, underfunded, and lacking in time for building a new resource without external help [18]. There is a small but steadily growing (at about 4 sites per year) community of XCBC-hosting institutions, and XCI is working to enable collaboration among them, by hosting conference sessions and fostering communication between site hosts. XCBC and XNIT provide a common technology basis that supports collaboration among institutions in conducting scientific research with the support of cyberinfrastructure.

XSEDE documents user infrastructure access and use requirements as formal use cases. These use case include most notably the ability for users to create and management use profiles information, link and use federated authentication identities to access infrastructure, login to distributed resources, move data between resources, execute remote jobs, and discover available infrastructure. XSEDE then identifies and evaluates community software/service components that can satisfy these use cases and integrates those components into the distributed infrastructure. Notable data capabilities integrated by XSEDE include the Globus reliable and high-performance data transfer service, the Globus data sharing service, and the Globus Connect data transfer endpoint, which facilitate data transfer and sharing between high-end HPC, visualization, storage systems, campus systems, and individual researcher personal systems.

3.3 Jetstream

Jetstream is the NSF's first production cloud for general-purpose science and engineering research and education. Jetstream, which began early operations in February 2016, is based on the OpenStack cloud environment software with a menu-driven interface to enable users to easily select a precomposed virtual machine (VM) to perform discipline-specific analysis. Jetstream uses the Atmosphere user interface developed as part of CyVerse (formerly iPlant), providing a low barrier to use by practicing scientists, engineers, educators, and students, and Globus services from the University of Chicago for seamless integration into the national cyberinfrastructure fabric. The team implementing Jetstream has as their primary mission extending the reach of the NSF's eXtreme Digital (XD) program to researchers, educators, and research students who have not previously used NSF XD program resources, including those in communities and institutions that traditionally lack significant cyberinfrastructure resources [19].

Jetstream was designed by using a formal architectural process based on definition of use cases and The Open Group Architecture Framework architectural processes. The design was based on a needs analysis of communities of researchers and research educators, taking into consideration NSF's goal of supporting the entire spectrum of NSF-funded communities. The strategy in defining the architecture, interface, and overall design of Jetstream was to identify communities that do not currently make extensive use of XD/XSEDE resources, document and understand their needs and work patterns, and architect a computing resource to meet those needs. One key aspect of this strategy was to collaborate directly with major research software projects, data resources, and groups with common needs, in order to create conditions that facilitate adoption of Jetstream and engage their

community leaders as partners to leverage their community contacts and influence [20].

Jetstream specifically promotes collaboration by enabling researchers to share VM images that run on Jetstream or any other cloud system that runs on the open source OpenStack software. Jetstream makes available to all users of Jetstream a library of what we call "featured" VMs. These provide commonly needed functionality, and we guarantee that they will work properly or we will fix them promptly. There is also a category of VMs that are contributed by one research team for use by others. An individual researcher can use any of these contributed VMs and then store versions of these VMs with, for example, data and scripts, in an individual researcher's private VM library. Jetstream thus provides a means by which researchers can distribute scientific software so that other researchers may discover and use it without downloading it to a local system. In addition, the team operating Jetstream offers researchers the opportunity to upload a VM image to Indiana University's digital repository (scholarworks.iu.edu) and associate a digital object identifier with that VM. In this way a researcher can make available a VM containing the entire lifecycle of their research—input data, scripts, all of the programs used to analyze the data, and output data—in a way that allows other researchers to discover that VM through citations in a paper. This provides an important mechanism for sharing code, sharing analyses, and supporting collaboratories. Furthermore, like the Open Science Grid, the Jetstream project relies explicitly on communities of practice, collaboratories, and virtual organizations to support the members of those virtual groups [19].

3.4 ACI-REF

The NSF-funded Advanced Cyberinfrastructure – Research and Educational Facilitation (ACI-REF) project [21] provides a community of cyberinfrastructure facilitators who can assist researchers in the "long tail" of scientific inquiry in order to increase the effectiveness of campus and national cyberinfrastructure in support of research. These campus facilitators work directly with researchers in order to assist in harnessing resources on their own campuses as well as in the national arena. The program provides training to help facilitators provide support and education to researchers at their institutions, including providing insight about the integral concerns of these researchers.

The members of ACI-REF's network of facilitators are able to leverage each other's skills and allow for specialization in domain areas. While many of the ACI-REF facilitators are skilled in the traditional cyberinfrastructure areas (HPC in support of high-energy physics, chemistry, astronomy, and materials science), facilitators also can provide support for research tools such as geographic information systems. Facilitators also provide an agnostic guide to resources available, whether they be XSEDE, Open Science Grid, or resources provided by campus or regional arrangements. In this way, the ACI-REF program is able to provide guidance to the overall cyberinfrastructure environment and engage deeply in the domain of research in order to identify the best ways to meet researchers' computational needs.

The ACI-REF environment is collaborative in that facilitators share technical information and tactics for working with researchers, It is a joint enterprise of research facilitators rather than being individual efforts focused on local campuses. This collaborative platform creates a template for researchers in their own work, sharing tools and techniques either through the facilitator network or via their own networks of contacts.

3.5 Science Gateways

Science gateways are software-based cyberinfrastructure that provide user-facing services, user interfaces, and APIs for virtual organizations [22] [23]. Gateways were originally considered to be browser-based Web portals for scientists to access remote computing resources through Grid middleware, but the software systems used to build gateways have matured to be considered first class cyberinfrastructure middleware in their own right. Projects such as Apache Airavata [24], HUBzero [25], CyVerse Agave [26], Galaxy [27], and others provide software and platform services for building gateways. Science gateways have also matured into production services for XSEDE, with the number of science gateway-based users surpassing traditional command-line users every quarterly reporting period since December 2013.

Despite notable successes, science gateways arguably have potential for much greater growth. Identifying the communities and organizations that would benefit from gateways implies two additional challenges: gateways must be easier to build and maintain technically, and communities operating gateways should plan for sustainability. Two organizations provide services to assist with these challenges. The XSEDE Extended Collaborative Support Services [28] effort's Science Gateway program provides gateway building experts distributed across XSEDE partner institutions who are available to provide supporting for building and integrating science gateways with XSEDE. The NSF-funded Science Gateway Institute, which began operations in August 2016, goes further, providing support for developers to use more diverse resources than just XSEDE, to provide "incubator" services that help gateway providers plan how to sustain their efforts, and a software collaborator that helps gateways find the software that they need.

Looking forward, there is a natural partnership between science gateways and the other components described in this section. Gateways are user support "force multipliers" that can help campus infrastructure providers using XCI tools to scalably support larger user communities than would be possible using one-on-one interactions.

4. CONCLUSIONS

Science has become largely a team undertaking, and the need for people and computational resources—in short, cyberinfrastructure—to support these teams is growing. In this paper we have discussed a number of organizations that are making efforts to bridge the gap between researchers and the tools they require in order to get work done. We also have discussed challenges of large-scale academic computation that such organizations aim to reduce or resolve. The tools and facilities described in this paper have already proved useful in supporting collaboratories and collaboration among institutions, and effort is ongoing to continue improving the computational landscape. We hope that by highlighting tools that are designed to aid collaboratories and collaboration between institutions, we will help members of SIGUCCS and the higher education community discover, evaluate, and adopt new cyberinfrastructure tools to aid collaboration in research and in education.

REFERENCES

[1] Stewart, C.A., Simms, S., Plale, B., Link, M., Hancock, D., and Fox, G. What Is Cyberinfrastructure? In: Proceedings of SIGUCCS 2010. (Norfolk, VA, 24-27 Oct, 2010). Available from: http://portal.acm.org/citation.cfm?doid=1878335.1878347.

[2] ACCI Task Force On Campus Bridging. National Science Foundation Advisory Committee for Cyberinfrastructure Task Force on Campus Bridging Final Report (2011).doi=http://hdl.handle.net/2022/13210.

[3] A. Mowshowitz. 1997. Virtual Organization. Commun. ACM 40, 9 (September 1997), 30-37. doi=http://dx.doi.org/10.1145/260750.260759.

[4] T. A. Finholt. Collaboratories. Annual Review of Information Science and Technology (ARIST), 36 (2002): 73-107.

[5] The Large Hadron Collider. http://home.cern/topics/large-hadron-collider.

[6] The Nobel Prize in Physics 2013. http://www.nobelprize.org/nobel_prizes/physics/laureates/2013/.

[7] Processing: What to Record? http://home.cern/about/computing/processing-what-record.

[8] Open Science Grid. http://opensciencegrid.org.

[9] Virtual Organization Summary. http://myosg.grid.iu.edu/vosummary?all_vos=on&active=on&active_value=1.

[10] Galison, P., Hevly, B. W. (1992). Big Science: The Growth of Large-Scale Research. Stanford University Press.

[11] What We Do. https://www.xsede.org/what-we-do.

[12] Extended Collaboration and Support Services. https://www.xsede.org/ecss.

[13] Campus Champions. https://www.xsede.org/campus-champions.

[14] Venkatesh, V., Morris, M. G., Davis, F. D., Davis, G. B. User Acceptance of Information Technology: Toward a Unified View. MIS Quarterly, 2003. 27(3): 425-478.

[15] How it Works. https://www.globus.org/how-it-works.

[16] Data Sharing. https://www.globus.org/data-sharing.

[17] Data Publication with Globus. https://www.globus.org/data-publication.

[18] Coulter, E., Fischer, J., Hallock, B., Knepper, R., Stewart, C. A. Implementation of Simple XSEDE-Like Clusters: Science Enabled and Lessons Learned. Proceedings of the XSEDE16 Conference on Diversity, Big Data, and Science at Scale. Article No. 10. Available from http://dl.acm.org/citation.cfm?id=2949570&CFID=837206729&CFTOKEN=83076021.

[19] Fischer, J., Tuecke, S., Foster, I., Stewart, C. A. Jetstream: A Distributed Cloud Infrastructure for Underresourced Higher Education Communities. In: Proceedings of the 1st

Workshop on the Science of Cyberinfrastructure: Research, Experience, Applications and Models (SCREAM '15). ACM, New York, NY, 2015, pp. 53-61. DOI=http://dx.doi.org/10.1145/2753524.2753530.

[20] Stewart, C. A., Cockerill, T. M., Foster, I., Hancock, D. Y., Merchant, N., Skidmore, E., Stanzione, D., Taylor, J., Tuecke, S., Turner, S., Vaughn, M., Gaffney, N. I. 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, July 26-30, 2015.

[21] Neeman, H., Bergstrom, A., Brunson, D., Ganote, C., Gray, Z., Guilfoos, B., Kalescky, R., Lemley, E., Moore, B. G., Ramadugu, S. K., Romanella, A., Rush, J., Sherman, A. H., Stengel, B., Voss, D. The Advanced Cyberinfrastructure Research and Education Facilitators Virtual Residency: Toward a National Cyberinfrastructure Workforce. Proceedings of the XSEDE16 Conference on Diversity, Big Data, and Science at Scale. Article No. 57. http://dl.acm.org/citation.cfm?id=2949584&CFID=837206729&CFTOKEN=83076021

[22] Wilkins-Diehr, Nancy, Dennis Gannon, Gerhard Klimeck, Scott Oster, and Sudhakar Pamidighantam. "TeraGrid science gateways and their impact on science." *Computer* 41, no. 11 (2008): 32-41.

[23] Lawrence, Katherine A., Michael Zentner, Nancy Wilkins-Diehr, Julie A. Wernert, Marlon Pierce, Suresh Marru, and Scott Michael. "Science gateways today and tomorrow: positive perspectives of nearly 5000 members of the research community." *Concurrency and Computation: Practice and Experience* 27, no. 16 (2015): 4252-4268.

[24] Pierce, Marlon E., Suresh Marru, Lahiru Gunathilake, Don Kushan Wijeratne, Raminder Singh, Chathuri Wimalasena, Shameera Ratnayaka, and Sudhakar Pamidighantam. "Apache Airavata: design and directions of a science gateway framework." *Concurrency and Computation: Practice and Experience* 27, no. 16 (2015): 4282-4291.

[25] McLennan, Michael, and Rick Kennell. "HUBzero: a platform for dissemination and collaboration in computational science and engineering."*Computing in Science & Engineering* 12, no. 2 (2010): 48-53.

[26] Dooley, Rion, Matthew Vaughn, Dan Stanzione, Steve Terry, and Edwin Skidmore. "Software-as-a-service: the iPlant foundation API." In *5th IEEE Workshop on Many-Task Computing on Grids and Supercomputers (MTAGS)*. 2012.

[27] Goecks, Jeremy, Anton Nekrutenko, and James Taylor. "Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences." *Genome biology* 11, no. 8 (2010): 1.

[28] Wilkins-Diehr, Nancy, Sergiu Sanielevici, Jay Alameda, John Cazes, Lonnie Crosby, Marlon Pierce, and Ralph Roskies. "An Overview of the XSEDE Extended Collaborative Support Program." In *International Conference on Supercomputing*, pp. 3-13. Springer International Publishing, 2015.