Encyclopedia of Information Science and Technology, Third Edition

Mehdi Khosrow-Pour
Information Resources Management Association, USA

Volume IX
Categories: Net-Sus

Information Science REFERENCE
An Imprint of IGI Global

Managing Director: Lindsay Johnston
Production Editor: Jennifer Yoder & Christina Henning
Development Editor: Austin DeMarco & Jan Travers
Acquisitions Editor: Kayla Wolfe
Typesetter: Mike Brehm, John Crodian, Lisandro Gonzalez, Deanna Zombro
Cover Design: Jason Mull

Published in the United States of America by
Information Science Reference (an imprint of IGI Global)
701 E. Chocolate Avenue
Hershey PA, USA 17033
Tel: 717-533-8845
Fax: 717-533-8661
E-mail: cust@igi-global.com
Web site: http://www.igi-global.com

Library of Congress Cataloging-in-Publication Data

Encyclopedia of information science and technology / Mehdi Khosrow-Pour, editor.
pages cm
Includes bibliographical references and index.
ISBN 978-1-4666-5888-2 (hardcover) -- ISBN 978-1-4666-5889-9 (ebook) -- ISBN 978-1-4666-5891-2 (print & perpetual access) 1. Information science--Encyclopedias. 2. Information technology--Encyclopedias. I. Khosrow-Pour, Mehdi, 1951-
Z1006.E566 2015
020.3--dc23

2014017131

British Cataloguing in Publication Data
A Cataloguing in Publication record for this book is available from the British Library.

All work contributed to this book is new, previously-unpublished material. The views expressed in this book are those of the authors, but not necessarily of the publisher.

For electronic access to this publication, please contact: eresources@igi-global.com.

Cyberinfrastructure, Science Gateways, Campus Bridging, and Cloud Computing

Craig A. Stewart
Indiana University, USA

Richard Knepper
Indiana University, USA

Matthew R. Link
Indiana University, USA

Marlon Pierce
Indiana University, USA

Eric Wernert
Indiana University, USA

Nancy Wilkins-Diehr
University of California, San Diego, USA

INTRODUCTION

Computers accelerate our ability to achieve scientific breakthroughs. As technology evolves and new research needs come to light, the role for cyberinfrastructure as "knowledge" infrastructure continues to expand. This article defines and discusses cyberinfrastructure and the related topics of science gateways and campus bridging; identifies future challenges in cyberinfrastructure; and discusses challenges and opportunities related to the evolution of cyberinfrastructure, "big data" (data-centric, data-enabled, and data-intensive research and data analytics), and cloud computing.

BACKGROUND

The evolution of cyberinfrastructure as a concept spans some three decades. The earliest references in 1976 (Sorkin, 2006) and in a Clarke and Hunker press briefing (1998) mention "cyber-infrastructure" in the context of cyber threats and cybersecurity.

Cyberinfrastructure in today's sense originated in the NSF-funded supercomputer centers program of the 1980s (National Science Foundation, 2006). The NSF centers delivered and supported supercomputers, which were generally accessed individually, often with users logging into a system that served as a front end to such supercomputers. Using multiple supercomputers in concert was at first practically impossible. This began to change in the late 1980s. Projects such as the CASA testbed, (Messina, 1991a, 1991b) linked multiple supercomputers together to support distributed scientific workflows. The NASA Information Power Grid (Johnston, Vaziri, & Tanner, 2001) provided a production grid of multiple supercomputers connected by a high-speed network.

These two projects advanced the grid concept in computer science and computational science. The computing architecture implied in the term made intuitive sense. An early definition of grid computing reads:

A computational grid is a hardware and software infrastructure that provides dependable, consistent, pervasive, and inexpensive access to high-end computational capabilities. (Foster & Kesselman, 1998)

Other grid types based on function include data grids and collaboration grids. Semantic grids and peer-to-peer systems are grids distinguished by the

DOI: 10.4018/978-1-4666-5888-2.ch645

characteristics of the protocols and interactions between components (Fox, 2006).

At the turn of the century two major projects developed major grid infrastructure in the USA. Three different projects developing grid technology to analyze data for physics data led to today's Open Science Grid (Open Science Grid, 2013). In 2001, the NSF funded the TeraGrid, computational, storage, and visualization resources in a grid that spanned the US.

The term "cyberinfrastructure" in its sense of *knowledge infrastructure* was introduced in 2001 by Dr. Ruzena Bajcsy in her charge to a National Science Foundation Advisory Panel led by Dr. Daniel Atkins. She wished to "create a program on cyberinfrastructure that would involve the broader computer science/ information technology community" (Bajcsy, 2013). According to Freeman (2007) this effort "led to the creation of a term for infrastructure that attempts to capture the integration of computing, communications, and information for the support of other activities (especially scientific in the case of NSF)." The NSF report created by the Atkins-led NSF Advisory Panel "Revolutionizing Science and Engineering Through Cyberinfrastructure," now known as "the Atkins report," clarified: "The newer term cyberinfrastructure refers to infrastructure based upon distributed computer, information and communication technology. If infrastructure is required for an industrial economy, then we could say that cyberinfrastructure is required for a knowledge economy" (Atkins et al., 2003a).

Indiana University staff developed a definition more specific in terms of identifying components and function.

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. (Stewart, 2007)

The EDUCAUSE Campus Cyberinfrastructure Working Group and the Coalition for Academic Scientific Computation developed a definition based which includes teaching and learning:

Cyberinfrastructure consists of computational systems, data and information management, advanced instruments, visualization environments, and people, all linked together by software and advanced networks to improve scholarly productivity and enable knowledge breakthroughs and discoveries not otherwise possible. (Dreher et al., 2009)

CYBERINFRASTRUCTURE TODAY

Cyberinfrastructure is distinguished from other IT terms and concepts by the following elements:

- Geographically distributed IT resources, expressed in the phrase "linked together by software and [...] networks"
- People
- Capabilities advanced enough to create "knowledge breakthroughs not otherwise possible."

Examples

One of today's largest single examples is the eXtreme Science and Engineering Discovery Environment (XSEDE), "...the most advanced, powerful, and robust collection of integrated advanced digital resources and services in the world." This NSF-supported project replaces and expands on the NSF TeraGrid project, and is used by more than 10,000 scientists, teachers, and students (XSEDE, 2013a).

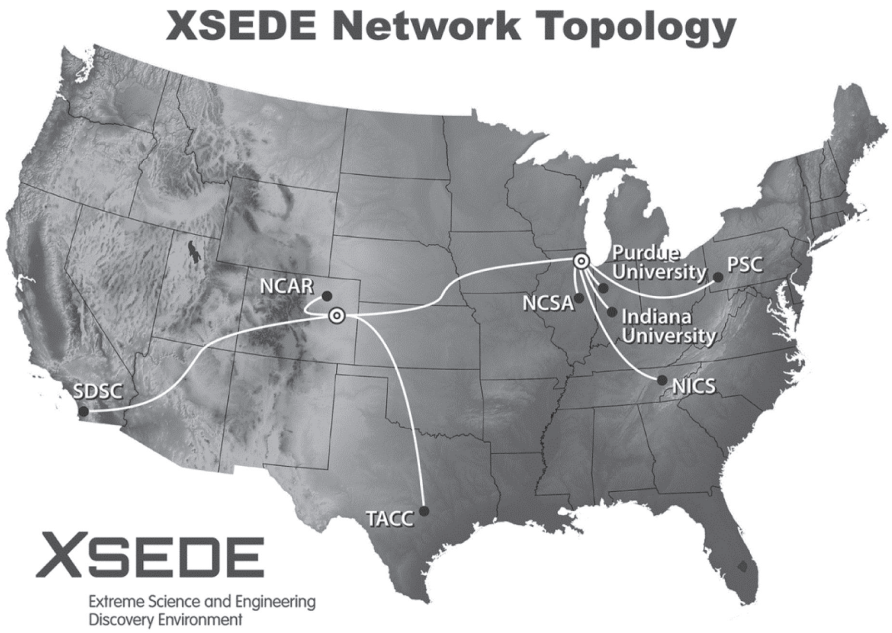
XSEDE exemplifies a large-scale infrastructure.

- It is a "single virtual system that scientists can use to interactively share computing resources, data, and expertise" (XSEDE, 2013b). It enables breakthroughs that would otherwise not be possible.
- It is physically distributed and tied together by networks (Figure 1).
- Expert support staff are critical.

Other examples of government-funded cyberinfrastructure projects include:

- The *Open Science Grid* data analysis cyberinfrastructure of thousands of smaller computers. It enabled analysis of Large Hadron Collider

Figure 1. The eXtreme Science and Engineering Discovery Environment (XSEDE, 2012). (© 2012, XSEDE. Used with permission.)



data, including evidence of the Higgs Boson that resulted in a Nobel Prize (Open Science Grid, 2013).

- *PRACE* (the Partnership for Advanced Computing in Europe, 2010) is building a pan-European cyberinfrastructure facility that includes the largest supercomputers accessible to most EU scientists.

Not all cyberinfrastructure systems are government funded. *BOINC* is a distributed computing grid of privately owned computers (BOINC, 2013). Though the nature of the network connecting the computers limits types of research calculations, as of 2011, BOINC's computational capability was among the largest in the US (Welch et al., 2011). Cyberinfrastructure has been widely adopted in the private sector, particularly in advanced engineering, medicine and pharmaceuticals, mining and oil exploration, finance, and manufacturing (Tabor Griffin Communications 1998).

Cyberinfrastructure may also serve to support a particular scientific domain or application. A specialized cyberinfrastructure was developed to support *Operation IceBridge* (OIB) (NASA, 2013), in which planes use sophisticated radar systems to study polar ice and map the bedrock base in Greenland and Antarctica. OIB (Figure 2) uses an in-plane computation and data storage cluster for real-time analysis of multiple radar data sources. Data are duplicated in flight and again when the plane lands. Analyzed data are communicated to North America via satellite. When the mission is complete raw data and data products are shipped back to the US.

OIB enables breakthroughs not otherwise possible. The in-plane cyberinfrastructure expert analyzes images, helps researchers identify polar features, and detects radar problems in real time.

Components of Cyberinfrastructure

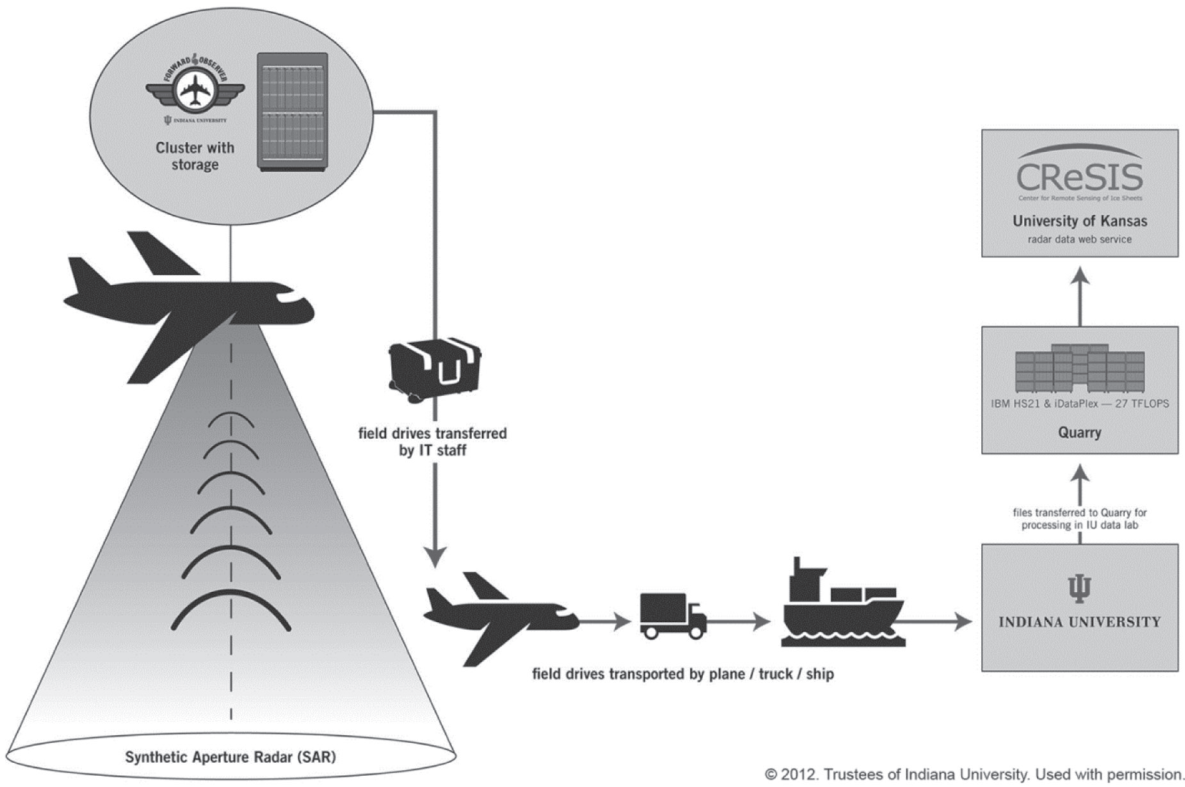
Supercomputers and data resources are described in other articles (cf. HPSS - High Performance Storage System (2012), Walgenbach, Simms, Miller, and Westneat (2010)). Other components are described below.

Middleware is defined by the NSF as follows.

Middleware refers to the software that is common to multiple distributed applications and is built atop the network transport layer and the operating system. Middleware manages interactions between remote resources and hides the underlying complexity so that

R

Figure 2. NASA Operation IceBridge Field Radar Data Processing Service. OIB cyberinfrastructure is a refinement of work begun by some of the authors with Geoffrey Fox (Guo, Singh, & Pierce, 2009; Hayden, Fox, & Gogineni, 2007)



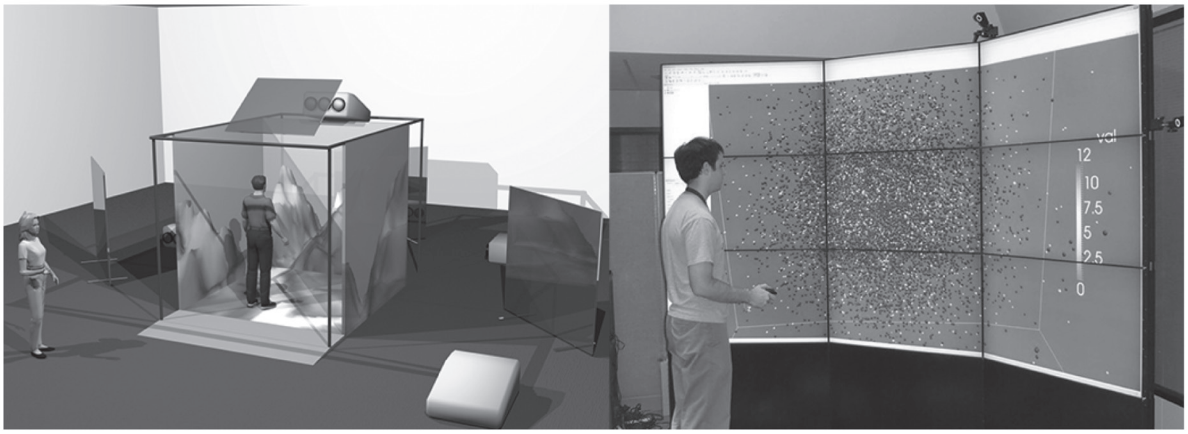
rapid development of new networked applications is enabled (National Science Foundation, 2004).

The Globus toolkit (Foster, 2005; Globus Online, 2013) is a middleware software suite whose functions include authentication and secure access to remote computing systems and data. Middleware includes scientific workflow systems that orchestrate the coordinated use of cyberinfrastructure and automate complex analyses. Examples include Apache Airavata (Marru et al., 2011), Kepler (Ludäscher et al., 2006), Taverna (Taverna, 2012), and Pegasus (Deelman et al., 2005).

Visualization systems — hardware (display systems, visualization computer systems, and interaction devices) and software (applications, libraries, middleware, and data format standards) — facilitate the visual understanding of data. A person must engage with the system.

Visualization was one of the earliest cyberinfrastructure components to promote distributed applications and high levels of interoperability, largely because of the network of homogeneous CAVE Automatic Virtual Environments (CAVEs) and smaller devices using similar software launched in the last half of the 1990s (NCSA, 2001). Users at multiple sites could synchronously interact with the same data sets and observe remote participants via virtual avatars while communicating over IP-based audio and video channels. CAVEs and similar devices introduced new capabilities for understanding complex 3D and 4D data from other cyberinfrastructure resources. Cost and scarcity limited their impact on day-to-day scientific investigation. In the 2000s came affordable PC-based graphics cards and digital light processing (DLP) projectors, and the subsequent powerful GPU cards and high-definition, stereoscopic flat panel displays. Consumer-level technologies spurred a range of innovative systems for stereoscopic (Geowall Consortium, 2006; Wernert et al., 2005) and ultra-resolution visualization (SourceForge, 2006), democratizing advanced visualization systems

Figure 3. At left is a CAVE, a room-scale visualization environment. At right is an ultra-high resolution tiled wall built in 2012 using commodity HDTV displays. (© 2012, Trustees of Indiana University. Used with permission.)



and techniques. Figure 3 shows a CAVE diagram and an ultra-high resolution tiled wall assembled from commodity HD televisions.

People are key, especially experts in what the NSF refers to as computational and data-enabled science and engineering (NSF Advisory Committee for Cyberinfrastructure Task Force on Grand Challenges, March 2011) in pushing the evolution of cyberinfrastructure to frontiers.

Science gateways are "a community-specific set of tools, applications, and data collections that are integrated together via a portal or a suite of applications" that can "support a variety of capabilities including workflows, visualization as well as resource discovery and job execution services" (Wilkins-Diehr, 2007). They help democratize access and enable the intuitive use of sophisticated systems. For example, in summer 2012 a 15-year-old student used NSF supercomputers through a science gateway to win his high school science fair (Graham, 2012).

Gateways serve unique communities, such as data discovery, management, and access (the Earth System Grid); some focus on computational requirements and workflows (GridChem); others (nanoHUB) have powerful collaborative aspects.

Gateways are often associated with larger-scoped research activities and may outlive the original project. The NSF-funded Cyberinfrastructure for Phylogenetic Research (CIPRES, 2003-2008) project enabled large-scale phylogenetic reconstructions. CIPRES is the most-used XSEDE science gateway, providing scalable, high performance versions of life sciences applications through a simple browser interface, so users can investigate problems without needing to understand the underlying cyberinfrastructure.

Scientific workflow systems used within a science gateway execute remote applications or capture details about a simulation so it can be reproduced. A composition tool allows scientists to create workflows for specific types of scientific computation. Workflows can then be deposited into an online registry service. Science gateways allow scientists to run the workflows and inspect results stored with an archive service.

Peer-to-peer (P2P) computing, in which a group of different computers connected by an overlay network act as peers (Schollmeier, 2001) has similarities with grid computing and could be a cyberinfrastructure component. The sense of "peers" distinguishes P2P computing from the heterogeneity in the definition of cyberinfrastructure.

FUTURE RESEARCH DIRECTIONS

Cloud Computing and Big Data

Cloud computing and big data present cyberinfrastructure challenges and opportunities. According to the National Institute of Standards and Technology (NIST),

Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool

R

of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction. (Mell & Grance, 2011)

Key in cloud computing are on-demand self-service, broad network access, resource pooling, rapid elasticity, and measured service. The "Infrastructure as a Service (IaaS)" model provides resources such as storage and networks, and the ability to deploy and run software, which can include operating systems and applications.

Cloud providers offer large numbers of virtual machines (VMs) on a charge-for-service basis, where prices depend on the processing capability and amount of memory and may deliver many identical VMs to hundreds, thousands, or tens of thousands of customers.

Cloud computing facilities often have internal networks with limited bandwidth. Rackspace offers internal network speeds of 40 to 600 Mbps (Megabits per second) (Rackspace Cloud Computing & Hosting, 2013); Amazon Web Services offers 250 Mbps (Amazon Web Services, 2013). In 2013, in a high performance computer, 1 Gbps (Gigabit per second) is a low-performance interconnect and a 10 Gbps internal network connection is common – 40 times faster than the internal network speed of the Amazon Web Services cloud. Cloud services have smaller memory per VM (2 to 8 GB) than a supercomputer component of a major cyberinfrastructure system (64 GB to a terabyte of RAM per node). But the accessibility and flexibility make it attractive for many loosely coupled applications.

Choosing cloud computing vs. more traditional cyberinfrastructure depends on priorities: the flexibility to purchase resources based on need vs. the ability to control one's own data. Commercial firms may choose the former; research organizations that depend on their own data stores may adopt cyberinfrastructure or private clouds (internal to a particular company).

Cloud computing and more traditional distributed cyberinfrastructure have complementary strengths and weaknesses. Neither a cloud nor a single supercomputer is cyberinfrastructure; either might be a component. Some claim cloud computing will obviate other forms of computing. Cloud computing and distributed cyberinfrastructure are more likely to both remain important in research and commerce.

There are business cases for selecting cloud computing or local servers (Brumec & VrčEk, 2013; Marston, Li, Bandyopadhyay, Zhang, & Ghalsasi, 2011). Choosing between cloud computing and cyberinfrastructure is more complicated. Cloud resources cannot always support the science applications that large-scale cyberinfrastructure can because low-latency internal networks or large amounts of memory are prerequisites. The cost of a supercomputer CPU hour may undercut one from a cloud provider. Such programs as FutureGrid (FutureGrid, 2013) and Grid 5000 Grid5000, 2013) enable experimentation in new cloud services.

Big data are characterized by size, structural complexity, high rate of production, and complexity in extracting meaning and in variety and nature of sources. Various cloud computing resources target types of big data problems, and include tools based on MapReduce (Apache Software Foundation, 2012; Indiana University, 2012). Big data and big compute (supercomputing) problems have a much in common.

Usability, Science Gateways, and Scientific Workflows

For years, researchers accessed advanced cyberinfrastructure systems through command-line interfaces. Today science gateways provide powerful graphical user interfaces, but tend to be specific to analysis task. Sometimes development occurs in a vacuum, without the long-term planning needed for success and longevity. With an incubation program, gateway developers could share experiences and code and focus on elements unique to their communities. They could also access expertise needed for long-term success in such areas as business plan development, marketing, software engineering, security, usability and licensing.

Scientific workflows help manage the many scientific processes that must work together to, for example, determine the structure of a chemical compound. Sometimes the scientist can do this manually; other times it is impossible.

Cyberinfrastructure needs improved collaboration tools. Today people stay current on friends through social media. Gateways can serve the same purpose in a science community and disseminate knowledge more quickly than reading scholarly articles. During the 2003 severe acute respiratory syndrome (SARS) breakout in Southeast Asia, dozens of scientists and medical researchers collaborated remotely to understand the source of the disease and its transmission. Collaboration tools sped up discovery and improved researcher safety. The nanoHUB gateway provides a collaborative cyberinfrastructure platform thousands of nanotechnology researchers use worldwide (nanoHUB.org, 2013). Continuing to develop science gateways and workflow tools is needed for cyberinfrastructure to have the same impact on science that social media has on communication.

Visualization

The democratization of visualization systems made them more accessible but also prompted organizations to develop custom software tools and techniques for their display systems. This resulted in poor hardware and software standardization and interoperability. Later, efforts by the NSF (including TeraGrid and XSEDE), the National Institutes of Health, and the Department of Energy led to new de-facto standards for more sustainable, scalable, and sharable visualization tools. New efforts are beginning to reduce the unproductive variability in advanced visualization displays (Electronic Visualization Laboratory, 2004; IQ-Station, 2013). Although visualization in advanced cyberinfrastructure still faces challenges and opportunities, two areas are (re-) emerging: remote visualization and the tighter integration of visualization into the scientific workflow.

Remote visualization goes back to remote X-Windows displays in the 1980s. With improved accessibility and greater need, visualization processes and tools are becoming more tightly integrated into the scientific workflow. Visualization is now critical in the exploration and analysis stages of data-intensive research. Web-enabled visualization technologies are launching new opportunities and needs for visualization tools in domain-specific science gateways. As visualization grows more important in scientific discovery and reasoning, the need grows for more intuitive visualization tools and infrastructure.

Citizen Science and eScience

Science gateways have had a profound impact on citizen science, or the public contribution to scientific discoveries. Dozens of projects use citizen science in weather data, archaeology, biology (whale communication, ocean floor species diversity), and medicine (cancer research). Individuals have donated millions of hours analyzing vast stores of data. Projects such as Galaxy Zoo make astronomical images broadly available to students, hobbyists, and professionals, and dozens of scholarly papers have been published as a result of these efforts (Zooniverse, 2012). eScience is "the large scale science that will increasingly be carried out through distributed global collaborations enabled by the Internet" (National e-Science Centre, 2010), and which requires access to massive data collections, computing resources, and high performance visualization.

Campus Bridging

This approach to providing basic research cyberinfrastructure aims to create consistency, flexibility, transparency, and virtual proximity. Users access distributed computational, storage, network, and visualization resources as if on the desktop. Internet2 NET+ services provide cost-effective, easily accessed cloud and video services. Genesis II provides a globally federated file system and grid queues that can submit jobs to multiple resources. The Open Science Grid's Campus Infrastructures Community BOSCO tool sets up job submission management from researchers' machines. As campus bridging strategies mature, researchers will be able to start, modify, and extend science workflows between their own systems and national research systems.

Evaluation, Power, Disaster Resilience, Security

Cyberinfrastructure challenges include documenting return on investment and the growing cost of electricity to operate its supercomputers. Its scale heightens the risk of malicious action, and the need for cybersecurity (Kshteri, 2013). On the other hand, physically disparate components provide disaster resilience and are important in industrial, sensitive, and university-based research applications. The density of power use at any one location is less than the system's total power consumption. These practical advantages and benefits suggest cyberinfrastructure will continue to exist in its own right and complement cloud computing.

CONCLUSION

Cyberinfrastructure has evolved from supercomputer centers in the US and European Union into an integrated and distributed suite of powerful and flexible resources that surpass the capabilities of massive supercomputers. Integrating supercomputers, data resources, visualization environments, and people, cyberinfrastructure extends the impact of information technology. In the private sector, it has led to new products, medical treatments, and improved business processes that combine to improve the quality of human life. The future offers unbounded opportunities for science and society as new tools for visualization, science gateways, campus bridging, citizen science, and cloud computing evolve and deliver new capabilities to the public and the scientific and technical communities worldwide.

ACKNOWLEDGMENT

This material is based upon work supported by the National Science Foundation under grants 0504075, 0451237, 0723054, 1062432, 0116050, 0521433, 0503697, and 1053575, and several IBM Shared University Research grants and support provided by Lilly Endowment, Inc. for the Indiana University Pervasive Technology Institute. Any opinions, findings and conclusions or recommendations expressed herein are those of the authors and do not necessarily reflect the views of the supporting agencies.

REFERENCES

Allen, B., Avery, P., Baker, K., Baker, R., Bauerdick, L., Branson, J., et al. (2001) *An international virtual-data grid laboratory for data intensive science.* Submitted to the 2001 NSF Information and Technology Research Program. Retrieved from http://www.phys.ufl.edu/~avery/ivdgl/itr2001/proposal_all.pdf

Amazon Web Services. (2013). Amazon EC2 FAQs. Retrieved May 2013, from https://aws.amazon.com/ec2/faqs/

Apache Software Foundation. (2012). Welcome to Apache Hadoop! Retrieved May 1, 2013, from http://hadoop.apache.org/

Atkins, D. E., Droegemeier, K. K., Feldman, S. I., Garcia-Molina, H., Klein, M. L., Messerschmitt, D. G., et al. (2003a) *Revolutionizing science and engineering through cyberinfrastructure: Report of the National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure*. Retrieved from http://www.nsf.gov/cise/sci/reports/atkins.pdf

Berman, F., Fox, G., & Hey, T. (Eds.). (2003). *Grid computing: Making the global infrastructure a reality*. Chicester, England: John Wiley & Sons. doi:10.1002/0470867167

BOINC. (2013). Home page. Retrieved April 30, 2013, from http://boinc.berkeley.edu/

Brumec, S., & VrčEk, N. (2013). Cost effectiveness of commercial computing clouds. *Information Systems, 38*(4), 495-508. doi: 10.1016/j.is.2012.11.002. Retrieved from http://dl.acm.org/citation.cfm?id=2442207

Clarke, R., & Hunker, J. (1998). Press briefing by Richard Clarke, National Coordinator for Security, Infrastructure Protection, and Counter-Terrorism; and Jeffrey Hunker, Director of the Critical Infrastructure Assurance Office. Retrieved April 30, 2013, from http://www.fas.org/irp/news/1998/05/980522-wh3.htm

Deelman, E., Singh, G., Su, M.-H., Blythe, J., Gil, Y., & Kesselman, C. et al. (2005). Pegasus: A framework for mapping complex scientific workflows onto distributed systems. *Science of Computer Programming*, *13*(3), 219–237. Retrieved from http://dl.acm.org/citation.cfm?id=1239653

Dreher, P., Agarwala, V., Ahalt, S. C., Almes, G., Fratkin, S., Hauser, T., et al. (2009). Developing a coherent cyberinfrastructure from local campuses to national facilities: Challenges and strategies. *EDUCAUSE*. Retrieved from http://www.educause.edu/Resources/DevelopingaCoherentCyberinfras/169441 or http://hdl.handle.net/2022/5122

Electronic visualization laboratory. (2004). SAGE: Scalable adaptive graphics environment. Retrieved April 30, 2013, from http://www.evl.uic.edu/core.php?mod=4&type=1&indi=281

Foster, I. (2005). Globus toolkit version 4: Software for service-oriented systems. In H. Jin, D. Reed, & W. Jiang (Eds.), *Network and parallel computing* (Vol. 3779, pp. 2–13). Berlin, Heidelberg: Springer. doi:10.1007/11577188_2

Foster, I., & Kesselman, C. (Eds.). (1998). *The grid: Blueprint for a new computing infrastructure*. Morgan Kaufmann.

Foster, I., & Kesselman, C. (Eds.). (2004). The grid 2, second edition: Blueprint for a new computing infrastructure. Morgan Kauffman.

Fox, G. (2006). Collaboration and community grids. *International symposium on collaborative technologies and systems*. Retrieved from http://dl.acm.org/citation.cfm?id=1153169.1153314

FutureGrid. (2013). Home page. Retrieved May 1, 2013, from http://futuregrid.org/

Geowall Consortium. (2006). GeoWall. Retrieved April 30, 2013, from http://www.geowall.org/

Graham, R. (2012). SDSC's CIPRES science gateway clarifies branches in evolution's 'tree of life'. *UC San Diego News Center*. Retrieved from http://ucsdnews.ucsd.edu/pressrelease/sdscs_cipres_science_gateway_clarifies_branches_in_evolutions_tree_of_life

Grid5000. (2013). Home page. Retrieved May 1, 2013, from https://www.grid5000.fr/

GriPhyN. (2003). Grid physics network in ATLAS. Retrieved April 30, 2013, from http://www.usatlas.bnl.gov/computing/grid/griphyn/

Guo, Z., Singh, R., & Pierce, M. (2009). Building the PolarGrid portal using web 2.0 and OpenSocial. In *Proceedings of the 5th Grid Computing Environments Workshop*. Portland, OR.

Hayden, L., Fox, G., & Gogineni, P. (2007). Cyberinfrastructure for remote sensing of ice sheets. *TeraGrid 2007*. Madison, WI. Retrieved from http://cerser.ecsu.edu/citeam/teragrid07.pdf

HPSS - High performance storage system. (2012). Home page. Retrieved April 30, 2013, from http://www.hpss-collaboration.org

Indiana University. (2012). Twister iterative MapReduce. Retrieved May 1, 2013, from http://www.iterativemapreduce.org/

Indiana University Pervasive Technology Institute. (2011). Campus bridging. Retrieved May 1, 2013, from http://internal.pti.iu.edu/campusbridging

IQ-Station. (2013). Home page. Retrieved April 30, 2013, from http://iq-station.com/

Johnston, W. E., Vaziri, A., & Tanner, L. A. (2001, 4-7 Mar 2001). The State of NASA's Information Power Grid. *Global Grid Forum*. Amsterdam, Netherlands. Retrieved from http://hdl.handle.net/2060/20010111389.

Korab, H., & Brown, M. D. (1995). Virtual Environments and Distributed Computing at SC'95: GII Testbed and HPC Challenge Applications on the I-WAY. Supercomputing '95. San Diego, CA.

Kshetri, N. (2013). Privacy and security issues in cloud computing: The role of institutions and institutional evolution. *Telecommunications Policy, 37*(4–5), 372-386. doi: dx.doi.org/10.1016/j.telpol.2012.04.011 Retrieved from http://www.sciencedirect.com/science/article/pii/S0308596112000717

Ludäscher, B., Altintas, I., Berkley, C., Higgins, D., Jaeger, E., & Jones, M. et al. (2006). Scientific workflow management and the Kepler system: Research articles. *Concurrency and Computation*, *18*(10), 1039–1065. Retrieved from http://dl.acm.org/citation.cfm?id=1148454 doi:10.1002/cpe.994

Marru, S., Gunathilake, L., Herath, C., Tangchaisin, P., Pierce, M., Mattmann, C., et al. (2011). Apache airavata: A framework for distributed applications and computational workflows. In *Proceedings of the 2011 ACM workshop on Gateway computing environments*. Seattle, WA.

Marston, S., Li, Z., Bandyopadhyay, S., Zhang, J., & Ghalsasi, A. (2011). Cloud computing - The business perspective. *Decision Support Systems*, *51*(1), 176–189. Retrieved from http://dl.acm.org/citation.cfm?id=1943810. doi:10.1016/j.dss.2010.12.006

Mell, P., & Grance, T. (2011). *The NIST definition of cloud computing*. Retrieved from http://csrc.nist.gov/publications/nistpubs/800-145/SP800-145.pdf

Messina, P. (1991a). Casa gigabit network testbed. *Supercomputer '91 Conference*.

Messina, P. (1991b). Parallel computing in the 1980s: One person's view. *Special issue: Practical parallel computing: Status and prospects*. nanoHUB.org. (2013). Home page. Retrieved April 30, 2013, from http://nanohub.org/

NASA. (2013). IceBridge - Antarctic 2012. Retrieved April 30, 2013, from http://www.nasa.gov/mission_pages/icebridge/index.html

National e-Science Centre. (2010). Defining e-science. Retrieved May 1, 2013, from http://www.nesc.ac.uk/nesc/define.html

National Science Foundation. (2004). NSF middleware initiative (NMI) Program Solicitation NSF 04-555. Retrieved 30 April 2013, from http://www.nsf.gov/pubs/2004/nsf04555/nsf04555.htm

National Science Foundation. (2006). Cyberinfrastructure: From supercomputing to the TeraGrid. Retrieved November 7, 2012, from http://www.nsf.gov/news/special_reports/cyber/fromsctotg.jsp

NCSA. (2001). CAVERN users aociety. Retrieved April 30, 2013, from http://cavernus.ncsa.illinois.edu/

NSF Advisory Committee for Cyberinfrastructure Task Force on Grand Challenges. (2011, March). *Final report*. Retrieved from http://www.nsf.gov/od/oci/taskforces/TaskForceReport_GrandChallenges.pdf

Online, G. (2013). Globus toolkit homepage. Retrieved April 30, 2013, from http://www.globus.org/toolkit/

Open Grid Forum. (2013). Home page. Retrieved April 30, 2013, from http://www.gridforum.org/

Open Science Grid. (2013). Home page. Retrieved April 30, 2013, from http://www.opensciencegrid.org/

OpenScientist. (2011). Finalizing a definition of "citizen science" and "citizen scientists." *OpenScientist.* Retrieved from http://www.openscientist.org/2011/09/finalizing-definition-of-citizen.html

Particle Physics Data Grid. (2006). Home page. Retrieved April 30, 2013, from http://web.archive.org/web/20120913025324/http://ppdg.net/

Partnership for Advanced Computing in Europe. (2010). PRACE in a few words. Retrieved April 30, 2013, from http://www.prace-ri.eu/PRACE-in-a-few-words

Rackspace Cloud Computing & Hosting. (2013). Cloud servers pricing. Retrieved May 1, 2013, from http://www.rackspace.com/cloud/servers/pricing_b/

Schollmeier, R. (2001). A definition of peer-to-peer networking for the classification of peer-to-peer architectures and applications. In *Proceedings of the first international conference on peer-to-peer computing*. Retrieved from http://dl.acm.org/citation.cfm?id=883282

Sorkin, S. (2006). *Goods in transit*. Matthew Bender and Company Inc.

SourceForge. (2006). Chromium introduction. Retrieved April 30, 2013, from http://chromium.sourceforge.net/doc/index.html

Stewart, C. A. (2007). Welcome. *IU Cyberinfrastructure News*. Retrieved from http://internal.pti.iu.edu/ci/iu-cyberinfrastructure-news-march-2007

Stewart, C. A., Simms, S., Plale, B., Link, M., Hancock, D., & Fox, G. (2010). What is cyberinfrastructure? In *Proceedings of SIGUCCS 2010*. Norfolk, VA. Retrieved fromhttp://dx.doi.org/10.1145/1878335.1878347

Storm, D. (2012). Meet the fastest, most powerful science machine in the world: Titan supercomputer. *Security Is sexy.* Retrieved from http://blogs.computerworld.com/high-performance-computing/21260/meet-fastest-most-powerful-science-machine-world-titan-supercomputer

Tabor Griffin Communications. (1998). *High-performance computing contributions to society*. Tabor Griffin Communications.

Taverna. (2012). Home page. Retrieved April 30, 2013, from http://www.taverna.org.uk/

Walgenbach, J., Simms, S. C., Miller, J. P., & Westneat, K. (2010). Enabling Lustre WAN for production use on the TeraGrid: A lightweight UID mapping scheme. In *Proceedings of TeraGrid'10*. Pittsburgh, PA: Retrieved from. doi:10.1145/1838574.1838593

Welch, V., Sheppard, R., Lingwall, M. J., & Stewart, C. A. (2011) *Current structure and past history of US cyberinfrastructure (data set and figures)*. Retrieved from http://hdl.handle.net/2022/13136

Wernert, E., Boyles, M., Huffman, J. N., Rogers, J., Huffman, J. C., & Stewart, C. A. (2005). The john-e-box: Fostering innovation, inclusion, and collaboration through accessible advanced visualization. In *Proceedings of the 2005 Conference on diversity in computing*. Albuquerque, New Mexico, USA. Retrieved from http://portal.acm.org/citation.cfm?doid=1095269

Wilkins-Diehr, N. (2007). Special issue: Science gateways -- Common community interfaces to grid resources: Editorials. *Concurrency and Computation*, *19*(6), 743–749. Retrieved from http://dl.acm.org/citation.cfm?id=1232528. doi:10.1002/cpe.1098

XSEDE. (2012). XSEDEnet map. Retrieved April 30, 2013 from https://www.xsede.org/documents/10157/169907/xsedenet.pdf

XSEDE. (2013a). Home page. Retrieved from https://www.xsede.org/

XSEDE. (2013b). Overview. Retrieved April 30, 2013, from https://www.xsede.org/overview

Zooniverse. (2013). Galaxy zoo. Retrieved April 30, 2013, from http://www.galaxyzoo.org/

ADDITIONAL READING

National Institute of Standards and Technology. (2011). The NIST definition of cloud computing (SP 800-145). Retrieved Nov 20, 2012, from http://csrc.nist.gov/publications/PubsSPs.html#800-145

National Institute of Standards and Technology. (2011). The NIST definition of cloud computing (SP 800-145). Retrieved Nov 20, 2012, from http://csrc.nist.gov/publications/PubsSPs.html#800-145

National Science Foundation Cyberinfrastructure Council. (2007). Cyberinfrastructure vision for 21st-century discovery. Hey, Tony, Tansley, Stewart, & Tolle, Kirstin (Eds.). (2009). The fourth paradigm: Data-intensive scientific discovery. Redmond, WA: Microsoft Research.

National Science Foundation Cyberinfrastructure Council. (2007). Cyberinfrastructure vision for 21st-century discovery. Hey, Tony, Tansley, Stewart, & Tolle, Kirstin (Eds.). (2009). The fourth paradigm: Data-intensive scientific discovery. Redmond, WA: Microsoft Research.

Taylor, I. J., Deelman, E., Gannon, D. B., & Shields, M. (Eds.). (2012). *Workflows for e-science: Scientific workflows for grids*. New York, NY: Springer-Verlag.

Taylor, I. J., Deelman, E., Gannon, D. B., & Shields, M. (Eds.). (2012). *Workflows for e-science: Scientific workflows for grids*. New York, NY: Springer-Verlag.

KEY TERMS AND DEFINITIONS

Campus Bridging: The seamlessly integrated use of cyberinfrastructure operated with other local or remote cyberinfrastructure as if they were proximate to the user.

Cloud Computing: Aims to deliver on-demand, affordable access to a distributed, shared pool of computing and storage resources, applications, and services usually via the Internet to a large number of users.

Citizen Science: The work of individuals or teams of amateur, non-professional, or volunteer scientists who conduct research, gather and analyze data, perform pattern recognition, and develop technology, often in support of professional scientists.

Computational Grid: Hardware and software infrastructure that provides access to geographically distributed computational resources. Data grids focus on data analysis. Both can be components of cyberinfrastructure.

Cyberinfrastructure: Cyberinfrastructure consists of computational systems, data and information management, advanced instruments, visualization environments, and people, all linked together by software and advanced networks to improve scholarly productivity and enable knowledge breakthroughs and discoveries not otherwise possible.

E-Science: Computationally intensive science carried out through distributed global collaborations enabled by the Internet, involving access to large data collections, very large scale computing resources and high performance visualization.

Science Gateways: Community-developed tools, applications, and data integrated via a portal or a suite of applications, usually in a graphical user interface, and customized to the needs of specific communities.

Scientific Workflows: Sets of tasks done in a specific order during computational experiments. Tasks are usually scientific applications that may run on more than one resource in the cyberinfrastructure. Workflows accommodate conditional decisions, loops, and interactivity with human monitors at various stages.