# 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

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 instru-ments, 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

# Related Content

GPU Based Modified HYPR Technique: A Promising Method for Low Dose Imaging
Shrinivas D. Desaiand Linganagouda Kulkarni (2015). *International Journal of Rough Sets and Data Analysis (pp. 42-57).*
www.irma-international.org/article/gpu-based-modified-hypr-technique/133532

Further Development of an Application Framework for Computational Chemistry (AFCC) Applied to New Drug Discovery
J. Tindle, M. Gray, R.L. Warrender, K. Gintyand P.K.D. Dawson (2014). *Contemporary Advancements in Information Technology Development in Dynamic Environments (pp. 92-110).*
www.irma-international.org/chapter/further-development-of-an-application-framework-for-computational-chemistry-afcc-applied-to-new-drug-discovery/111607

Big Data, Knowledge, and Business Intelligence
G. Scott Ericksonand Helen N. Rothberg (2018). *Encyclopedia of Information Science and Technology, Fourth Edition (pp. 943-950).*
www.irma-international.org/chapter/big-data-knowledge-and-business-intelligence/183806

The Bees Algorithm as a Biologically Inspired Optimisation Method
D.T. Phamand M. Castellani (2015). *Encyclopedia of Information Science and Technology, Third Edition (pp. 285-294).*
www.irma-international.org/chapter/the-bees-algorithm-as-a-biologically-inspired-optimisation-method/112336

Meeting Gender Gaps in Information and Communication Technology (ICT): How Can Creativity Make a Difference?
Chunfang Zhou (2019). *Gender Gaps and the Social Inclusion Movement in ICT (pp. 212-229).*
www.irma-international.org/chapter/meeting-gender-gaps-in-information-and-communication-technology-ict/218446