

EDUCAUSE Center for Applied Research

Research Bulletin

Volume 2009, Issue 2

January 27, 2009

# More People, Not Just More Stuff: A New Vision for Research Cyberinfrastructure

Paul B. Gandel, Syracuse University

Megan Oakleaf, Syracuse University

Jeffrey M. Stanton, Syracuse University

R. David Lankes, Syracuse University

Derrick Cogburn, Syracuse University

Elizabeth D. Liddy, Syracuse University



With the passing of George Carlin in June 2008, the world lost not only a great comedian but an insightful observer of the human condition. In his classic comic routine about “stuff,” Carlin cleverly illustrated how we all become attached to and constrained by the stuff we acquire and cling to—how our own riches can impose unintended consequences.

For researchers in higher education, easy Internet access to high-end computing, digital storage, and new electronic tools for analysis and collaboration have had some unanticipated consequences. At the threshold of almost limitless possibilities, the research and IT communities struggle to figure out how to take full advantage of all this new stuff—this emerging cyberinfrastructure abundance. To paraphrase Carlin, if we didn’t have all this “stuff,” we wouldn’t be in the situation we’re in.

Thanks to Moore’s Law and related principles of electronic abundance, we can expect computing power and bandwidth to continue to increase for the foreseeable future. We also can expect that storage will be ever more plentiful and inexpensive. Knowing this, it is incumbent on the research and IT communities, when thinking about cyberinfrastructure, to imagine a world in which the contents of your life work can fit in your pocket, the results of experiments are measured in terabytes, and the entirety of science and associated knowledge is a mouse-click away. As higher education information technologists race to develop more stuff for researchers, we must, together with them, consider the impacts.

Powerful cyber-research tools are too complex to be run solely by a scientist or investigator. Imagine a geneticist who wants to test a new protein folding technique. Large-scale supercomputing or a distributed grid computing process would speed the research, but the scientist might not have the first clue about how to access these resources or how to prepare to run a series of protein-folding simulations.

Scientists and researchers spend their careers mastering the skills, knowledge, and tools that form the core of their respective disciplines.<sup>1</sup> There is little time to simultaneously become experts in information management, networking, virtual or distributed collaboration, search and retrieval, archiving, user-interface development, or any of the other skills typically found within the information professions.<sup>2</sup> Furthermore, advances and convergences in cyberinfrastructure that have occurred over recent decades have themselves fueled a proliferation of scientific and engineering information—more findings, data sets, papers, conferences, journals, books, and so on. Even the brightest and most motivated of scientists struggle to keep up with the rapid pace of knowledge creation in their respective fields.<sup>3</sup> Likewise, as relative newcomers, students and junior scholars face an even larger challenge to master the range of developments in their fields. Finally, information infrastructure itself is in the process of an accelerating evolution. Gains in computing power, storage, transmission bandwidth, and other fundamental building blocks of cyberinfrastructure create frequent discontinuities in the economics of information technologies, while open-source software tools sprawl daily into innovative new application territories.<sup>4</sup>

In the new world of massive-scale computing, professional information technology facilitators are needed to work with researchers to identify the most effective analytical tools, data sets, and other resources to best achieve the research objectives. This research bulletin highlights a special academic program that has been developed by the School of Information Studies (iSchool) at Syracuse University for the express purpose of creating a new information professional: the cyberinfrastructure facilitator (CI-facilitator).

## Highlights of a New Vision for Research Cyberinfrastructure

The job title of CI-facilitator does not yet exist, nor is there a place—yet—where a student can learn to become a CI-facilitator. However, there is clearly a need for such facilitators in the research process to make sure systems are in place, data are available, resources are identified, and the entire research process moves forward in a timely and effective manner within the new and technologically complex environment in which many of our top research teams now operate. Where will these people come from, and how can they best be prepared for what is often a set of difficult and ill-defined tasks?

The iSchool is creating a program to educate and train this new breed of information professional. The program will focus on using current information infrastructure in innovative ways and teaching the cognitive skills needed to master new information infrastructure as it emerges. It will teach students to excel in the “three I’s”—information, infrastructure, and improvisation. Such a program must also provide students with the research skills they require to discover the needs of information users and how to adapt available technology to satisfy those needs. The program must also teach students how to:

- Forge the links between people, information, and tools in order to expose, mine, and process the mountains of scientific and engineering data that are continuously generated.
- Create information systems to integrate the communication, collaboration, and data management facilities that different communities need to form productive professional relationships.
- Understand the importance of subject-matter expertise.
- Work with scientists to represent, organize, and locate information resources.
- Understand technology adoption and use and the importance of user-centered design approaches.
- Facilitate collaboration for distributed groups via collaborative tools.

As the first step in creating a new program for educating and training CI-facilitators, the iSchool is developing a pilot demonstration project with the help of a grant from the NSF.<sup>5</sup>

## Project Goal and Structure

The goal of the demonstration project is to develop a program of education, training, internships, and core principles for advancing the adoption of cyberinfrastructure throughout the science, technology, engineering, and mathematics (STEM) disciplines. (While other disciplines would also benefit, this program is initially aimed at the STEM disciplines to keep the initial scope manageable for the pilot program.) The program will develop curriculum and internship opportunities to give students the capabilities to innovate with current information infrastructure and the metacognitive skills to master new information infrastructure as it emerges. Students will be trained to become vital members of the research enterprise, members who will work closely with STEM researchers to identify extant tools, data sets, and other resources that can be integrated in pursuit of shared research objectives, involving colleagues regardless of their geographic locations.

The project will be carried out in three phases:

- *Design:* Existing research on cyberinfrastructure skills will be combined with needs analyses conducted with STEM project partners. The result of this research synthesis will be used to design an educational program based on existing courses, modified courses, new courses, and related educational experiences (such as internships, technical training, stand-alone learning modules, and the like).
- *Prototype:* The newly devised curriculum will be tested with a cohort of master's-degree students who will be part of a two-year graduate-level certification program at Syracuse University.
- *Revision and evaluation:* Evaluation data from the cohort will be used formatively to revise the curriculum and summatively to assess the project as a whole. The result of the curriculum development, certification process, and evaluations will be used to create expertise for advancing the adoption of cyberinfrastructure.

## Defining Knowledge, Skills, and Tools

O\*NET (<http://online.onetcenter.org/>) is the U.S. Department of Labor's successor to the now outmoded Dictionary of Occupational Titles. O\*NET provides an empirically derived database of job descriptions based on formal job analysis, along with a collection of information resources for each job, including skills, abilities, activities, tools, and technology.<sup>6</sup> Not surprisingly, there is no job with cyberinfrastructure in the title or description, let alone with the specific title of CI-facilitator.<sup>7</sup> Nonetheless, the CI-facilitator roles that we envision do have their roots in existing work roles. In particular, we examined the skills and tools of seven different jobs in order to begin our needs assessment of the CI-facilitator training regime: computer systems analysts, database administrators, computer support specialists, training and development specialists, natural sciences managers, archivists, and audiovisual collections specialists. The

results of this analysis, summarized in Table 1 below, showed a striking degree of commonality in knowledge, skills, and tools across these seven jobs.

**Table 1. Commonality of Knowledge, Skills, and Tools/Technology Shared Across Seven Jobs**

Agreement Level	Knowledge	Skills	Tools and Technology*
100%	English language	Reading comprehension	Desktop computers
100%	Computers and electronics	Active learning	Notebook computers
100%	Customer and personal service	Active listening	Database management system software
86%	Science/math	Critical thinking	Object or component oriented development software
71%	Administration and management	Instructing/teaching	Mainframe computers
57%	Education and training	Written communication	Metadata management software

*\*Tools and technology data not available across all O\*Net jobs; percentages are approximate for this column.*

As Table 1 suggests, basic literacy, numeracy, and technology capabilities are fundamental to these jobs. Moreover, it is interesting to see the importance of a group of interrelated skills and knowledge that pertain to working with the “clients” or “users” that these job roles serve. In particular, the skills of active listening, customer service, instructing/teaching, and active learning appear prominently in most of these jobs. Given the CI-facilitator role that we have defined—“a vital member of the research enterprise who works closely with researchers to identify extant tools, data sets, and other resources that can be integrated into the process of pursuing a research objective”—this finding should not be surprising. Assisting scientists and engineers with their cyberinfrastructure needs requires a well-honed ability to elicit user requirements and translate those requirements into effective systems and services. At a purely intuitive level, it is easy to see how an individual with the knowledge, skills, and mastery of the tools described in Table 1 would be a welcome member of a science or engineering research team.

One other aspect of note in Table 1 is that the descriptions of knowledge, skills, and tools are necessarily expressed at a very broad level. This characteristic is intrinsic to O\*NET, which must cover a wide range of jobs with as small a collection of categories as possible. To obtain a more nuanced view of the knowledge, skills, and tools required for the CI-facilitator role, we asked subject-matter experts (PhD faculty and students from a variety of STEM disciplines) to brainstorm in these three categories. Table 2 displays the results of that effort.

**Table 2. Knowledge, Skills, and Tools/Technology Suggested by Subject-Matter Experts**

<b>Importance (1–10 scale)</b>	<b>Knowledge</b>	<b>Skills</b>	<b>Tools and Technology</b>
10	Domain knowledge in one or more areas of science and mathematics	Research skills (data elicitation, data analysis, scientific writing)	Database design tools
9	Architecture and operation of data networks	Communication skills (oral and written)	Content management system/website development tools
8	Information policy (access controls, intellectual property rights, licensing, privacy)	Service skills (working with people, determining user needs)	Server administration
7	Human-computer interaction	Database design skills	Large-scale computing (mainframes, supercomputing, grids)
6	Scripting, query, and programming languages	Cultural sensitivity, working with people with disabilities	Distributed collaboration tools

Although the subject-matter experts focused less on fundamental skills of literacy and numeracy than the O\*NET analysis showed, we believe these essential skills were assumed by our subject-matter experts. More importantly, these lists of knowledge, skills, and tools provide a much finer level of detail, particularly with respect to prevailing information technologies. Also worthy of comment, the subject-matter experts were adamant about the importance of math and science skills for the CI-facilitator. Individuals with in-depth knowledge of information technology alone may not be as valuable because they lack knowledge of the specialized nomenclature and needs in the specific domains of science and engineering.

Together, Tables 1 and 2 provide a very clear view of the educational challenges that we face in training a new generation of cyberinfrastructure professionals. The wide range of capability areas suggests that within the time constraints of a typical four-year bachelor's degree program, students are unlikely to obtain both domain-specific capabilities in science and engineering, as well as the plethora of needed technology and information capabilities. In this light, we have designed a master's-level curriculum, presented below, that takes a "multiple pathways" approach. We assume that the student entering the proposed master's program will have graduated with a bachelor's degree that included a major or dual major in a STEM area. Combined with our plans for an internship experience, this approach will ensure that the student involved in CI-facilitator preparation enters with substantial domain knowledge in one or more STEM areas. In the material below, we outline the existing and planned program elements that will prepare master's-level students for the CI-facilitator role.

### Program Elements

The foundation of this master's-level certification are the core knowledge areas identified in Table 2. The initial curriculum will be refined through the phases of the project. The

framework for the whole sequence is a 24-credit-hour graduate certification program that bridges technology, management, and information organization in traditional master's programs offered at the iSchool. The iSchool presently offers master's degrees in three programs: information management, library and information science, and telecommunications and network management. The selection of courses for the CI-facilitator program will cut across those three degree program curricula, with two courses drawn from each. We also plan to offer three new courses specifically aimed at this new program: Introduction to Scientific Information Management, Management of Large-Scale Datasets (with an alternative course choice, System Modeling and Simulation), and Distributed Collaboration and Emerging Technologies. The details of focus and content will be determined after we have conducted a needs assessment data collection with our STEM partners. These courses will serve as overviews that introduce CI-facilitator students to the range of activities and requirements inherent in the scientific lab environment. Thanks to the recent and current investments of NSF's cyberinfrastructure program, additional classroom-type resources exist for CI-facilitator education, resources that we plan to leverage for our program. One example is the University of Oklahoma Supercomputing Seminars, which began in fall 2007 and use a web-based teleconferencing format.

The internship experience will constitute the heart of our program. In our existing master's programs, experiential learning, both in internships and other settings, has proven to be a mark of distinction both for our students and for those who hire them. For the CI-facilitator program, we plan to place students in settings similar to the laboratory environments in which these individuals will eventually work. The evaluation plan discussed below includes specific outcomes and criteria that pertain to the internship experience. These outcomes will provide data about the quality and success of the internship experience from the perspectives of the students and the internship sponsors.

### Program Evaluation Strategy

The third phase of the program is evaluation. Program evaluation will be based on 11 general program outcomes that focus on stakeholders including students, faculty, internship sponsors, and employers, ordered here chronologically:

1. Students will enroll in CI-facilitator courses and complete the degree.
2. Students will report satisfaction with CI-facilitator courses and career preparation.
3. Students will demonstrate learning of cyberinfrastructure knowledge, skills, and abilities.
4. Faculty will provide quality CI-facilitator instruction.
5. Students will find placement in CI-facilitator internships.
6. Internship sponsors will report satisfaction with CI-facilitator students.
7. Student intern performance will reflect cyberinfrastructure knowledge, skills, and abilities.

8. Students will find placement in cyberinfrastructure jobs.
9. Team members will disseminate results of the project.
10. CI-facilitator graduates will demonstrate cyberinfrastructure knowledge, skills, and abilities.
11. Employers will report satisfaction with CI-facilitator graduates.

Because we are committed to gathering, using, and reporting both formative and summative assessment data, we have designed an iterative and continuous evaluation plan into the program. Most outcome areas will have three-step assessments that are staggered in time: fast (formative), short term (formative and summative), and long term (summative). The short-term and long-term summative evaluations will take place at the conclusion of each semester and at the end of the prototype program period as a whole. We will take a multi-source approach to this process, focusing primarily on student outcomes, along with the needs and expectations of the scientists and lab managers who will interact with the students during the internship period, as well as employers who provide positions for the students after graduation.

Outcome areas will be evaluated using multiple data sources as follows:

1. Student adoption of the program will be measured using course and program enrollment as well as degree completion.
2. Student satisfaction with the program will be measured using focus groups, interviews, exit surveys, and alumni interviews.
3. Student learning will be assessed using course tests and projects, a rubric, and case studies.
4. Instructor quality will be evaluated through course evaluations, instructor self-evaluations, instructor focus groups, and an analysis of curricular content in the CI-facilitator courses.
5. Internship placements will be evaluated using placement data.
6. Internship sponsor satisfaction data will be gleaned from informal feedback, surveys, and focus groups.
7. Intern performance will be evaluated by both students and sponsors using a rubric.
8. Student job placement will be tracked using career acquisition data.
9. Project dissemination will be evaluated by recording the dissemination activities accomplished such as a planned CI-facilitator website, conference presentations, and scholarly publications.
10. Postgraduate job performance and employer satisfaction will be done as a follow-up project but is beyond the scope of the initial project.

Each outcome also has a desired impact criterion whose specific metric and level will be set during the needs assessment with partner constituents in science labs and with professors active in the instructional part of the prototype program. In order to effectively measure student learning, the iSchool is also developing a set of rubrics to evaluate various aspects of this pilot program and to give students a set of criteria to measure their progress. The rubrics will cover:

- STEM subject-matter expertise
- Cyberinfrastructure subject-matter expertise
- Tool competency/infrastructure
- Research expertise
- Metacognition
- Creativity/innovation/improvisation
- User focus

The CI Learning Rubric will be used throughout the project to measure classroom learning and internship performance and to facilitate faculty evaluation of student assignments, internship supervisor assessment of student performance, and student self-evaluation. Throughout the project, the performance levels will be articulated and distinguished from each other so that the rubric captures growth and development of cyberinfrastructure knowledge, skills, and ability.

## What It Means to Higher Education

The long-term impacts of this project lie in the eventual indispensable contributions of the graduates of CI-facilitator training programs, both at Syracuse and at other institutions. Assuming the success of this demonstration project, a single instance of a CI-facilitator educational program will be run at a larger scale on the Syracuse campus. The findings and structure of this program will be disseminated for use by other institutions.

Our hope is not simply to create new educational programs but, more importantly, to fill a critical gap in our cyberinfrastructure and to create a new career path for our students. Increasing research productivity and taking full advantage of the technology tools at our disposal now and in the future are critical for all of higher education. Training and deploying CI-facilitators to key research teams can help increase research productivity and success. Moreover, creating this new position of CI-facilitator offers students an opportunity and professional career path to be directly involved in the research enterprise in a meaningful way. It can offer an alternative to the traditional PhD for students to play a meaningful role in the research enterprise.

Important questions about resources must be asked. Where will the resources come from to pay for these facilitators? We believe they will come from two sources. First is

the increased productivity of the research teams themselves as a result of the contribution of these facilitators. In other words, to some degree these facilitators may pay for themselves in the form of more successful grants and research funding. Second, there is an opportunity to shift resources. Rather than continue to have librarians wait passively for researchers to come to them, these resources could be shifted by retraining library personnel with science backgrounds to be CI-facilitators. The same may be true with computer center personnel. Both cases will require not only shifting resources and retraining personnel but, equally importantly, rethinking our organizational structures. CI-facilitators will need to be part of research teams and not part of centralized organizations, which are often far removed from the daily research activities of various research enterprises across campuses.

Developing a new profession of CI-facilitators will be challenging. Clearly this is an educational program that will evolve and change as we learn from experience. Organizations will have to learn how to best deploy, manage, and take full advantage of individuals who graduate from this program. Nevertheless, we believe the need for such professionals is great and the rewards for both the individuals entering into this career path and organizations taking full advantage of the skills of these people will be even greater.

## Key Questions to Ask

- How can the training of CI-facilitators best be integrated into existing educational programs?
- What are the best ways for institutions and research teams to deploy and take full advantage of CI-facilitators?
- How will we best assess the success of CI-facilitators?
- What professional support organizations will be needed, if any, to support the development of a new CI-facilitator profession?

## Where to Learn More

- Atkins, Daniel E. (chair), Kelvin K. Droegemeir, Stuart I. Feldman, Hector Garcia-Molina, Michael L. Klein, David G. Messerschmitt, Paul Messina, Jeremiah P. Ostriker, and Margaret H. Wright. *Revolutionizing Science and Engineering through Cyberinfrastructure*. Washington, D.C.: National Science Foundation, 2003. <http://www.nsf.gov/od/oci/reports/atkins.pdf>.
- Berman, Francine. "Making Research and Education Cyberinfrastructure Real." *EDUCAUSE Review* 43, no. 4 (July/August 2008), <http://connect.educause.edu/Library/EDUCAUSE+Review/MakingResearchandEducatio/46967>.
- Blustain, Harvey, with Sandra Braman, Richard N. Katz, and Gail Salaway. *IT Engagement in Research: A Baseline Study* (Research Study, Vol. 8) Boulder,

CO: EDUCAUSE Center for Applied Research, 2006, available from <http://www.educause.edu/ecar>.

- Freeman, Peter A. "Is 'Designing' Cyberinfrastructure—or, Even, Defining It—Possible?" *First Monday* 12, no. 6 (June 2007), [http://firstmonday.org/issues/issue12\\_6/freeman/index.html](http://firstmonday.org/issues/issue12_6/freeman/index.html).
- Sheehan, Mark C. *Higher Education IT and Cyberinfrastructure: Integrating Technologies for Scholarship* (Research Study, Volume 3). Boulder, CO: EDUCAUSE Center for Applied Research, 2008, available from <http://www.educause.edu/ecar>.

## Endnotes

1. Julie Thompson Klein, *Interdisciplinarity: History, Theory and Practice* (Detroit: Wayne State University Press, 1990); and Elaine Seymour, "Tracking the Processes of Change in U.S. Undergraduate Education in Science, Mathematics, Engineering, and Technology," *Science Education*, 86, no. 1 (2002): 79–105.
2. Marye Anne Fox and Norman Hackerman, eds., *Evaluating and Improving Undergraduate Teaching in Science, Technology, Engineering, and Mathematics* (Washington, D.C.: National Academies Press, 2003), <http://books.nap.edu/openbook.php?isbn=0309072778>.
3. Michael Gibbons, Camille Limoges, Helga Nowotny, Simon Schwartzman, Peter Scott, and Martin Trow, *The New Production of Knowledge: The Dynamics of Science and Research in Contemporary Societies* (Thousand Oaks, CA: Sage Publications, 1994); Diana M. Hicks and J. Sylvan Katz, "Where Is Science Going?" *Science, Technology & Human Values* 21, no. 4 (1996): 379–406; and Andrew Odlyzko, "The Rapid Evolution of Scholarly Communication," *Learned Publishing* 15, no. 1 (January 2002): 7–19, <http://www.si.umich.edu/PEAK-2000/odlyzko.pdf>.
4. Ilkka Tuomi, "The Lives and Death of Moore's Law," *First Monday* 7, no. 11 (November 2002): 4, [http://www.firstmonday.org/issues/issue7\\_11/tuomi/](http://www.firstmonday.org/issues/issue7_11/tuomi/); and Hal R. Varian, Joseph Farrell, and Carl Shapiro, *The Economics of Information Technology: An Introduction* (Cambridge: Cambridge University Press, 2004).
5. The project is supported through National Science Foundation (NSF) Grant # OCI 0753372, CI-Facilitators: Information Architects Across the STEM Disciplines. The NSF does not necessarily endorse the conclusions in this bulletin.
6. Rod McCloy, Gordon Waugh, Gina Medsker, Janet Wall, David Rivkin, and Phil Lewis, *Development of the O\*NET Computerized Work Importance Profiler* (Raleigh, N.C.: National Center for O\*NET Development); and Teresa A. Wagner, and Robert J. Harvey, "Job-Component Validation Using CMQ and O\*NET: Assessing the Additivity Assumption" (paper presented at the Annual Conference of the Society for Industrial and Organizational Psychology, Chicago, 2004), [http://harvey.psyc.vt.edu/Documents/2004siop\\_wagner\\_harvey.pdf](http://harvey.psyc.vt.edu/Documents/2004siop_wagner_harvey.pdf).
7. Francine D. Berman and Henry E. Brady, *Final Report: NSF SBE-CISE Workshop on Cyberinfrastructure and the Social Sciences* (National Science Foundation, 2005), <http://vis.sdsc.edu/sbe/reports/SBE-CISE-FINAL.pdf>; and Daniel Clery, "INFRASTRUCTURE: Can Grid Computing Help Us Work Together?" *Science* 313, no. 5786 (July 2006), 433–434, <http://www.sciencemag.org/cgi/content/full/313/5786/433>.

## About the Authors

*At Syracuse University, Paul B. Gandel (pgandel@syr.edu) is Professor, School of Information Studies; Megan Oakleaf (mogleaf@syr.edu) is Assistant Professor at the School of Information Studies; Jeffrey M. Stanton (jmstanto@syr.edu) is Associate Dean for Research and Doctoral Programs and Associate Professor at the School of Information Studies; R. David Lankes (rdlankes@syr.edu) is Associate Professor and Director of the Syracuse Information Institute; Derrick Cogburn (dcogburn@syr.edu) is associate professor in the School of Information Studies, research professor in the Maxwell School of Citizenship and Public Affairs, senior research associate in the Burton Blatt Institute for Innovation in Disability Studies, and Director of the Center for Research on Collaboratories and Technology Enhanced Learning Communities; and Elizabeth D. Liddy (liddy@syr.edu) is Trustee Professor and Dean of the School of Information Studies.*

### Copyright

Copyright 2009 EDUCAUSE and Paul B. Gandel, Megan Oakleaf, Jeffrey M. Stanton, R. David Lankes, Derrick Cogburn, and Elizabeth D. Liddy. All rights reserved. This ECAR research bulletin is proprietary and intended for use only by subscribers. Reproduction, or distribution of ECAR research bulletins to those not formally affiliated with the subscribing organization, is strictly prohibited unless prior permission is granted by EDUCAUSE and the authors.

### Citation for This Work

Gandel, Paul B., Megan Oakleaf, Jeffrey M. Stanton, R. David Lankes, Derrick Cogburn, and Elizabeth D. Liddy. "More People, Not Just More Stuff: Developing a New Vision for Research Cyberinfrastructure" (Research Bulletin, Issue 2). Boulder, CO: EDUCAUSE Center for Applied Research, 2009, available from <http://www.educause.edu/>.