



Supporting the “Scholarship” in E-Scholarship

The opportunities of cyberinfrastructure have been both heralded and hyped. The express purpose of cyberinfrastructure is to enable *e-scholarship*: new forms of scholarship that are more information-intensive, data-intensive, distributed, collaborative, and multidisciplinary.¹ But what does that mean for scholars, for academic libraries, and for campus information technology? What cyberinfrastructure strategies should colleges and universities implement to support the “scholarship” in e-scholarship?

Information-Intensive Scholarship

Retrieving whole books, articles, and other documents is no longer sufficient for scholarly research. Faculty and students want to mine documents or other textual works—whether for molecules, materials, or mavens, depending on their field of study. Rarely do people read documents linearly, even on paper: they read abstracts, conclusions, and bibliographies; they look for tables, figures, and diagrams. What is new in the digital environment? Information can be extracted in smaller units, mashed up, and recombined—preferably with attribution to the original sources. Faculty and students alike need assistance in learning how to think with these tools and services if they are to ask truly new questions with them.

Data-Intensive Scholarship

The “data deluge” started in the sciences, with highly instrumented research in physics, astronomy, chemistry, and molecular biology, and has spread to most other fields. Embedded sensor networks are now essential technologies for the environmental sciences. Internet communication provides rich traces of human activity for social scientists to study. Educators are logging interactions with simulated experiments, collaborative tools, and embedded assessments. Scholars in the humanities, not to be outdone, are mining texts and modeling cities.

Although the data deluge presents the most immediate challenge for information technology strategy, academic planning, and research infrastructure, it is also the area of e-scholarship most subject

to hype. *Wired* recently pronounced that science no longer needs theory, models, metadata, ontologies, or “the scientific method”: mining the data deluge replaces all of them.² That glib claim obscures the epistemological problems of ascertaining “facts” and classifying them in appropriate cognitive and cultural structures. It also obscures the challenges of distributing trust in digital environments. Making sense of long strings of bits is difficult enough in the few fields that have agreed on data structures, metadata, and ontologies. In most fields, individual investigators still manage their data in spreadsheets or databases of local origin, with names of data elements and variables assigned anew for each research project and instrument. Essential information about instrument calibrations, field conditions, and data transformations may be stored elsewhere, if documented at all.

Although many funding agencies now require data management plans and the deposit of data, enforcement of data-release rules is often inconsistent, and much remains to be negotiated about definitions and about the scope of the data to be managed, preserved, and shared. Are the data raw, processed, or published? Physical specimens? Codebooks? Interview instruments? Field notebooks? Blogs and wikis describing the research process? In what condition must data be released, and in what time frame (six months, two years, after publication)? How long must data be kept (until the end of the project, five years, ten years, forever)? What requirements for de-identification of

human subjects apply? In the few fields for which data repositories exist (e.g., genomics, seismology, and social surveys), a plan for depositing data at the end of a project may be sufficient. In most fields, data are likely to remain the responsibility of the investigator or the campus, raising an array of opportunities and challenges.

Data collections are valuable not only for research but also for instruction and for “citizen science” and analogues in the social sciences and humanities, offering new forms of outreach for colleges and universities.³ Data may become the “new special collections” for libraries, a possibility that has implications for collection development.⁴ Data curation is more complex than document curation because of the variation in form and in the amount of domain expertise required. Strategies for data curation will require the involvement of individual faculty, the campus research enterprise, the library, and both instructional and information technology services.

Distributed Scholarship

Regardless of whether scholars are working alone or in research partnerships, they want access to their information and data resources 24x7, from anywhere. In addition to the services offered by the library and campus data centers, they want access to their own resources and to those of their research partners, students, and others. Shared sites to post files and track progress are becoming essential research management tools. Moving large files between machines, sometimes with peer-to-peer technologies, raises security concerns in information technology circles. Security must be addressed, but not at the price of shutting down the legitimate sharing of content among research partners and between faculty and their students.

Online chat (with or without accompanying video) is useful not only for providing quick answers to questions but also for participating in meetings and even for giving guest lectures. Researchers blog daily reports from remote field deployments, whether in California, Costa Rica, or Peru. Grant proposal budgets include line items for conference calls and Skype accounts. As individual investigators and project participants select their preferred technologies, duplication of effort and investment often

proliferates. The campus information technology department should lead the faculty, rather than follow, by offering attractive services for distributed communication.

Collaborative Scholarship

Collaboration usually requires distributed access to resources, even if the partners are on the same campus. Access is but a baseline, however. In addition, collaboration often requires agreements on sharing data before, during, and after the project. These agreements, in turn, involve considerations of ownership, intellectual property, confidentiality of human subjects records, and sometimes contracts regarding third-party data restricted to certain personnel and time frames. Other aspects to be negotiated are supervision of students and staff and allocation of authorship credit. Extra care is required when multiple institutions or countries are involved, since rules and practices for data ownership, curation, and release may vary widely.

Collaborative projects have increased dramatically in size and frequency over the last several decades, but that does not mean they have become easier to manage. A large study recently determined that physical distance between partners complicates projects even more than does differences in disciplines.⁵

Multidisciplinary Scholarship

The challenges of working across disciplines are not new to cyberinfrastructure, but technology can either ease or exacerbate these issues. As problems become more complex and as collaboration across distance becomes easier (at least in principle), the temptation is to add partners. However, as the differences between participating disciplines increase, so does the overhead to negotiate agreements on topics such as how data are collected, how they are managed, and how credit is assigned. The nuanced finding of one field may be old news to another, leading to disagreements and resentment.

While better tools and technologies for collaboration certainly will advance e-scholarship, the harder challenges are the human ones. Sufficient experience with virtual organizations now exists to offer guidance in establishing effective collaborations that involve multiple investigators,

disciplines, and/or campuses; thus, the most painful pitfalls should be avoidable.⁶ Campus research officers need to be careful matchmakers and counselors when promoting collaborative and multidisciplinary cyberinfrastructure research projects.

Whither E-Scholarship?

E-scholarship, as a form of scholarship enabled by cyberinfrastructure, should be viewed as evolution more than revolution. The pace of that evolution varies widely within and between disciplines, campuses, and countries. Distributed and multidisciplinary collaborations are both facilitated and complicated by cyberinfrastructure. Similarly, the changing forms of information and the spreading data deluge offer not only a wealth of new research opportunities but also a daunting array of new challenges. Colleges and universities can minimize the challenges and maximize the opportunities by implementing campus cyberinfrastructure strategies that focus less on the technology per se and more on advances in scholarship and learning—that is, strategies supporting the “scholarship” in e-scholarship.^e

Notes

This article benefited greatly from comments by Geoffrey Bowker, Sayeed Choudhury, Amy Friedlander, Andrew Lau, Marcia Linn, Clifford Lynch, and Libbie Stephenson.

- Christine L. Borgman, *Scholarship in the Digital Age: Information, Infrastructure, and the Internet* (Cambridge: MIT Press, 2007).
- Chris Anderson, “The End of Theory: The Data Deluge Makes the Scientific Method Obsolete,” *Wired*, June 23, 2008.
- NSF Task Force on Cyberlearning, Christine L. Borgman et al., *Fostering Learning in the Networked World: The Cyberlearning Opportunity and Challenge*, June 24, 2008, <<http://www.nsf.gov/pubs/2008/nsf08204/nsf08204.pdf>> An excerpt from the Executive Summary for the report is reprinted on pp. 90–91 in this issue of *EDUCAUSE Review*.
- Sayeed Choudhury and Timothy L. Stinson, “The Virtual Observatory and the Roman de la Rose: Unexpected Relationships and the Collaborative Imperative,” *Academic Commons*, December 16, 2007, <<http://www.academiccommons.org/commons/essay/VO-and-roman-de-la-rose-collaborative-imperative>>
- Jonathon N. Cummings and Sara Kiesler, “Collaborative Research across Disciplinary and Organizational Boundaries,” *Social Studies of Science*, vol. 35, no. 5 (2005), pp. 703–722.
- National Science Foundation, *Beyond Being There: A Blueprint for Advancing the Design, Development, and Evaluation of Virtual Organizations*, Final Report from Workshops on Building Effective Virtual Organizations, May 2008, <<http://www.ci.uchicago.edu/events/VirtOrg2008/>>.

Christine L. Borgman is Professor and Presidential Chair in Information Studies at UCLA. She recently served as Chair of the National Science Foundation Task Force on Cyberlearning.