

Information Challenges in Collaborative Science

Jillian C. Wallis

University of California, Los Angeles
215 GSE&IS Building, Los Angeles, CA 90095
jwallisi@ucla.edu

Ayse Buyuktur

University of Michigan, Ann Arbor
1075 Beal Ave., Ann Arbor, MI 48109
abuyuktu@umich.edu

ABSTRACT

Collaborative research is on the rise, and presents difficult information challenges that must be overcome to make this mode of research effective. For instance distributed research is regularly plagued by the “distance matters” problem just as multi-disciplinary research is often plagued by misaligned terminology and assumptions. These problems can potentially be alleviated using information channels and mediation in new ways, as well as balancing technology with policy and incentives. In this panel we engage with information challenges in collaborative sciences from a variety of perspectives, with the hope that a rich discussion will emerge regarding these challenges.

Keywords

Cyberinfrastructure, methodology, HCI, digital libraries, policy

PANEL TOPIC

Knowledge production and knowledge management in the sciences are being transformed through initiatives such as Cyberinfrastructure in the U.S. and eScience or eResearch in the U.K., Europe, and elsewhere. Science - and other forms of scholarship - are becoming more collaborative, distributed, information-intensive, and data-intensive. Research about the information practices of these endeavors is an emerging area in information science and technology. Large-scale collaborative science projects are very difficult topic to study, however, due to their distributed, emergent, and partial character. Any single perspective — whether information-seeking, information-behavior, data-handling, knowledge organization, digital libraries, or archival records (just to name a few) — risks examining only one part of the elephant. New kinds of research methods and new kinds of collaborations are needed to explore cyberinfrastructure and eResearch. Challenges include theoretical and methodological assumptions of

ethnographic and place-based wholeness.

Cyberinfrastructure supports communities that are spatially remote, linked through information infrastructure and networks of various forms. They attract and connect a heterogeneous crowd of researchers, builders, funders, and users, few of whom live in the project as a central professional or personal identity. They live on paper and in imagination before and sometimes after they exist in hardware, software, and practice. They're virtual in the double sense of the word: mediated through technology; but also in some substantial way (as measured by traditional ethnographic standards) never fully "there."

This panel explores methodological and analytic responses to the challenge of writing “thick descriptions of thin phenomena” for cyberinfrastructure and other distributed scientific enterprises. We explore tools, methods, and strategies we have used in our research, and ongoing problems with which we and others continue to struggle. Effective solutions to these challenges will be central to information science's ongoing engagement with the development, deployment, and management of advanced information infrastructure in the sciences, social sciences, and the humanities. They may also provide new clues and strategies for practicing science studies at scale.

The panel consists of researchers who are actively engaged in the study of cyberinfrastructure communities. Each has a different theoretical or methodological perspective. They represent a range of career stages, from doctoral students writing dissertations to senior scholars in the field. The moderator will present a brief overview of the issues to be addressed. Each panelist will then make a 7-10 minute opening statement about a specific methodological issue he or she is addressing in current research. These statements are expected to provoke a rich methodological and theoretical discussion with the ASIST audience about the challenges of research on large-scale collaborative scholarship from an information science orientation.

PANEL MODERATOR

Christine Borgman, Professor & Presidential Chair in Information Studies, University of California, Los Angeles

Christine L. Borgman is Professor and Presidential Chair in Information Studies at UCLA. She is the author of more than 180 publications in the fields of information studies, computer science, and communication. Both of her sole-

authored monographs, *Scholarship in the Digital Age: Information, Infrastructure, and the Internet* (MIT Press, 2007) and *From Gutenberg to the Global Information Infrastructure: Access to Information in a Networked World* (MIT Press, 2000), have won the Best Information Science Book of the Year award from the American Society for Information Science and Technology. She is a lead investigator for the Center for Embedded Networked Systems (CENS), a National Science Foundation Science and Technology Center, where she conducts data practices research. She chaired the Task Force on Cyberlearning for the NSF, resulting in the report, *Fostering Learning in the Networked World*.

PANELISTS

Steven Jackson, University of Michigan, Ann Arbor
David Ribes, Georgetown University

Recent studies of large-scale distributed practice in the sciences have taught us important things about space and place as props and barriers to distributed collective action, but they have had relatively little to say about time. I will argue for the importance of *temporal alignment* as a neglected but crucial element underpinning distributed collective practice in the sciences (as well as other spheres of distributed collective activity). Specifically, I argue that joint scientific work is organized around four separate and potentially dissonant temporal registers, or 'rhythms' – *phenomenal, organizational, biographical, and infrastructural* – and that efforts to align such rhythms constitute an important and under-recognized aspect of collaborative work.

Steven Jackson is an Assistant Professor at the University of Michigan School of Information, where he also directs the school's Information Policy program and is part of the university-wide Science Technology and Society and Science Technology and Public Policy programs. His research interests include the practice and governance of large-scale collaborative science, IT and international development, and public/citizen engagement with scientific practice and decision-making through IT artifacts and processes.

David Ribes is an Assistant Professor in Communication, Culture and Technology Program at Georgetown University. His research and teaching interests lie at the intersection of sociology, philosophy and history of science & technology. His work focuses on the emerging phenomena of Cyberinfrastructure (i.e., networked information technologies for the support of science) and how these are transforming the practice and organization of contemporary knowledge production. His primary methods are ethnographic and archival.

Jillian C. Wallis, University of California, Los Angeles
CENS is a multi-disciplinary, distributed research center focused ostensibly on researching on one topic, sensor networks, from many angles. Researchers collecting data

together have different concepts of what counts as data, and surrounding norms for data cleaning, analysis, and preservation. Data itself is difficult to fix in time from a document standpoint as a dynamic resource with very little "there" there, but this is further complicated by acting as a shared resource within the collaboration. Disciplinary differences cause unintended effects within the collaboration, for example one collaborator making data available without not realizing that it impinges on another's claims to publish from it. This also affects the long-term survival of data for reuse by others due to little understanding of who is responsible for datasets to be propagated forward, and answer any questions about the data for interpretation, creating challenges for digital libraries or repositories of data.

Jillian Wallis is currently a PhD student at UCLA. Her research focuses on data practices of science and technology researchers at the Center for Embedded Networked Sensing (CENS), a UCLA-based National Science Foundation Science and Technology Center. She received her Masters in Library and Information Science at UCLA in 2005.

Geoffrey C. Bowker, University of Pittsburgh

In most large scale scientific collaborations, there is very little clarity about the role of the information manager. Some come from a domain science and move into the role; others come from computer science; few come from the iSchool. A central issue has been the lack of a clear career path. There a number of ways of professionalizing the field, but each way has significant disadvantages to complement their positive features. Longstanding theories of professionalization (Andrew Abbott; Donald Schon) do not account for this new working environment.

Geoffrey C. Bowker is Professor and Senior Scholar in Cyberscholarship at the School of Information Sciences, University of Pittsburgh. He is the author of *Memory Practices in the Sciences* (2006), which won the ASIST prize for "Best Information Science Book 2006" as well as the 2007 Ludwig Fleck Prize of the Society for Social Studies of Science (4S). His research interests include cyberinfrastructure for the sciences, critical reading of databases, classification and its consequences, science and technology studies, and the history of information practices.

Archer Batcheller, University of Michigan, Ann Arbor

Across many sciences, software development teams are building software that aspires to revolutionize scientific work. But building "perfect" software is not enough, and a project that only builds software will fail. These software teams also need to "sell" their software, provide support, offer workshops to the community, and generate useful examples. Some teams recognize the importance of these ancillary information activities, while others neglect or belatedly implement them. Just as teams developing

software for scientists must be sure to incorporate support activities, we must be certain not to neglect these activities in social studies of software teams.

Archer L. Batcheller is a doctoral candidate in the School of Information at the University of Michigan. His research focuses on the growth of cyberinfrastructure to support scientific work, with a particular focus on software engineering and requirements processes among software development teams engaged in building systems for scientists. He received his bachelor's degree from Princeton University in Electrical Engineering in 2005.

PANEL ORGANIZATION

The session will begin with the moderator's introduction to the common topic of the session followed by participants' brief 10 minute presentations. We will then open the floor to discussion between presenters and with the audience. The presentations themselves to act as common starting points for a collective discussion, and topics will emerge organically from discussion. That said, below are some of the topics we expect will come to structure the conversation:

- Distanced collaboration
- Data sharing, metadata creation, maintaining repositories

- Interdisciplinary collaboration and crossing boundaries (e.g., information management, science, computer science)
- Methods or best-practices for collaboration
- Developing long-term partnerships
- Reward structures and career pathways
- Funding and institutional challenges

SIGS WHO MIGHT FIND THIS PANEL OF INTEREST

- Critical Issues (CRIT)
- Digital Libraries (DL)
- Human-Computer Interaction (HCI)
- Information Architecture (IA)
- Information Needs, Seeking and Use (USE)
- Information Policy (IFP)
- International Information Issues (III)
- Knowledge Management (KM)
- Management (MGT)
- Scientific & Technical Information Systems (STI)
- Technology, Information, & Society (TIS)