Barriers to the Adoption of New Collaboration Technologies for Scientists

Charlotte P. Lee
Department of Technical Communication
University of Washington
14 Loew Hall, Seattle, WA 98195
cplee@u.washington.edu

Matthew Bietz
Department of Technical Communication
University of Washington
14 Loew Hall, Seattle, WA 98195
matt@matthewbietz.org

ABSTRACT

In this paper we discuss some barriers to the adoption of new collaboration technologies, based on empirical qualitative research of cyberinfrastructure development. We discuss two projects: 1) research on collaborative technologies that were implemented in a project early in its development and 2) preliminary findings of based on interviews of stakeholders in a nascent project in a new field of scientific endeavor: metagenomics.

Author Keywords

Scientific collaboration, cyberinfrastructure, ethnography.

ACM Classification Keywords

H5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

INTRODUCTION

Scientific research is increasingly conducted by large, multi-institution and interdisciplinary project teams, processing exponentially vaster and more complex data flows and with similarly larger and more complex sociotechnical structures. The increase in both the amount and kinds of scientific data being generated is overwhelming current software tools. At the same time, large scale collaborations must work across time and space, with collaborators often being geographically distributed presenting additional challenges to collaboration.

The overwhelming increase in data is inextricably intertwined with an intense scramble by individuals, teams, and organizations to develop new software tools and the new sociotechnical arrangements needed to develop and support new tools and the infrastructure needed to support software tools and new scientific practices surrounding the collection, processing, and analysis of the data itself.

Empirical research across projects yields some interesting examples of barriers to the adoption of new collaboration technologies for scientists.

STUDY 1: THE BRAIN IMAGE STUDY (BIS) PROJECT

The major goal of BIS is to develop tools to make multi-site functional MRI studies a common research practice. The challenges are complex and heterogeneous, combining technical, scientific, and organizational elements. Currently data cannot be pooled between machines of different manufacturers or even between different research sites using machines from the same manufacturer, thereby limiting researchers to studies based on locally collected samples of patients and control subjects. These samples tend to be small due to the difficulty of locating and enrolling appropriate research subjects, limited access to expensive machines, and the labor intensive nature of conducting clinical assessments and in-scanner cognitive tests.

Multi-site studies can ameliorate the problem of inadequate sampling in medical research. To take advantage of the power of multi-site studies, variability across sites must be addressed. Variability in imaging equipment, data acquisition and analysis, and patient assessment compromise the value of multi-site imaging datasets. Completion of the technical and clinical goals of BIS will enable researchers to tap the power of large-scale, multi-site neuroimaging studies.

Methodology and Data Collection

We used ethnographic research methods to collect data on BIS. We have undertaken participant observation for several months at weekly on-site, teleconferencing, and videoconferencing meetings of various working groups and all-Function Network meetings. Because the work of BIS is distributed over time and space, and because most BIS workers only work on BIS part-time, a critical means of data collection has been through one-on-one interviews. Twenty interviews have been completed with individuals from ten different institutions. Pseudonyms have been used for names of projects and people.

CHI 2009, April 4–9, 2009, Boston, MA, USA.

Findings

While the science undertaken by BIS during the course of this study was cutting edge, the collaborative technologies were not particularly advanced. Rather, collaborative technologies such as data conferencing, desktop sharing, teleconferencing, video teleconferencing and, eventually, wikis were the primary collaborative technologies. Studying the barriers to the adoption of these technologies, however, points towards potential barriers to the adoption of newer technologies.

Mismatch Between Management and User Requirements

Data analysis reveals that there was a mismatch between management conceptions of necessary collaborative technologies and the conceptions of the users. The management of BIS was highly committed to the expensive and flashy video teleconferencing (VTC) system. The VTC failed to meet many of the collaborative needs of BIS.

One group of participants were comprised of high-profile researchers who travelled frequently. A majority of them indicated in questionnaires and follow-up interviews that regular teleconferencing better met their needs in that it worked anywhere and was appreciated for being low-tech, more reliable, and not requiring the expense or attention of additional equipment or software.

Another example of a mismatch regarding the VTC system was that of the informatics subgroup. The informatics subgroup was comprised primarily of database engineers, computer scientists, and programmers. They found that rather than video teleconferencing, their primary need was dataconferencing and screen sharing. As the informatics participants were sophisticated technology users, they quickly found and used inexpensive or free dataconferencing and screen sharing software tools.

Orlikowski [4] noted some time ago that for successful adoption to occur, systems must take into account differing incentives. Our study of BIS confirms this finding, and also underlines the importance of understanding large scale scientific collaborations as large, complex organizations that contain numerous groups of collaborators with different backgrounds and information needs. Also technologists themselves can no longer be considered apart from the scientific endeavor, rather they are an integral part of the conduct and advancement of science.

Need for Local Contributions to Global Resources

The BIS functioned successfully for years without a comprehensive list of participants. An official list was kept on the official website, but was months, and in some instances, years out of date. Turnover in scientific collaborations can be quite rapid due to staff changes across universities, within a university, or between projects particularly at the research associate, post-doc, and lower-level administrative levels and keeping track of personnel changes across geographically distributed groups is very challenging. The decision was made to make each research site responsible for maintaining its own list. Enabling local control and responsibility where appropriate is an important feature for collaboration technologies.

When BIS began, it relied heavily on a central website for disseminating information and documentation relating to the project. The website was considered the official information resource for the public. As a consequence of its official status, any time an individual from any one of the several research projects wish to have information or a document added to the official website, they needed to email one individual to procure permission and then a second email with the information would then need to be conveyed to the website manager. The process often took a few days and was sufficiently cumbersome that most of the scientists avoided using the BIS website for collaborative work. Servers were in place for large, entire data sets, but scientists tended to rely instead on email and email archives for the dissemination and storage comparatively small amounts of data, information, or documentation.

Several months into the project, one of the scientists suggested the use of a wiki. As this was a couple years ago, while most of the scientists had heard of wikis, many of them had never used one. However, thanks to its ease of use it was quickly embraced become a clearinghouse for a wide variety of coordinative (e.g. dates and times for various meetings) and substantive information.

STUDY 2: METAGENOMICS

Metagenomics is a new science that enables the study of microorganisms by extracting DNA directly from communities of environmental microorganisms, thus sidestepping the need for culturing or isolation [3]. Currently, there is little information on the vast majority of microorganisms present in Earth's different environments due to the difficulty of culturing them in the laboratory. The resultant wealth of genes and molecular structures deciphered from uncultured microorganisms has tremendous potential in the development of novel biocatalysts for industrial and medical applications [5]. Metagenomics also offers a way to gauge changes in biodiversity and environmental health. Metagenomic techniques require significant computational and data resources, and several groups are working on developing cyberinfrastructure for the field.

Methodology and Data Collection

We are in the second year of a three year study of the growth of the field of metagenomics, especially the relationship between new science and cutting edge computation. We have conducted in-depth interviews members of three overlapping groups connected to the Community Cyberinfrastructure for Advanced Marine Microbial Ecology Research Analysis (CAMERA): cyberinfrastructure developers (n=10), bioinformaticists (n=4), and microbiologists (n=3). Over a three month period, one of the authors spent more than 50 hours in on-site observation of a cyberinfrastructure development team. Observations included weekly group meetings as well as several scheduled subproject meetings and numerous ad-hoc meetings. Observations also included shadowing several of the team members, sitting in on casual conversations among the team, and noting general patterns of interaction among team members. We have also conducted observations and attended lab meetings of a microbiology lab and attended an international workshop on developing data standards for genomics and metagenomics. Data collection is ongoing.

Preliminary Findings

In some senses, metagenomics is an inherently collaborative science. Once DNA sequences have been assembled, the next step in a metagenomic study is to determine if the organism has already been discovered and catalogued. Genomes are characterized by their degree of similarity to other known genomes, which can provide clues to the functional properties of the organism. Scientists can explore the evolutionary history of organisms by comparing the amount of difference in the genomes of two organisms of the same species. All of these analyses require comparing the sample genome to enormous databases of known genomes. The quality of the science depends on having high-quality contributions from all members of the community.

Community-contributed databases are not new in this field. GenBank, for example, is an open-access database of nucleotide sequences that has been in operation since the early 1980s. Many other sequence databases exist for specific countries (e.g. the DNA Data Bank of Japan), for specific kinds of sequences (e.g. metagenomic sequences in CAMERA), or with specialized tools (e.g. RAST annotations in The SEED). Since 1996, genome researchers have been required to make their genome sequences public [2], and most journals now require scientists to have deposited the sequence into a public database before they can publish their results.

At the same time, metagenomics research is done primarily in the context of microbiology laboratories, and collaborative analysis tends to be more “traditional.” Research projects are typically housed within a single laboratory, or sometimes in two or three laboratories in a collaborative project. While data is considered a public resource, analysis is primarily the domain of the individual or small group.

This suggests that fertile ground for scientific social software is more likely to be found (at least initially) around data sharing activities. It can be useful to compare these databases to other large community contribution systems like wikis. Such comparisons may lead to potential social solutions for some of the problems inherent in the current generation of sequence databases.

Our informants often mention quality problems within sequence databases. While publications go through an extensive peer review process, database submissions do not, and many of the submitted sequences are never attached to a particular publication. Some sequences are also submitted in a “draft” form but are, for whatever reason, never completed. New sequencing technologies make it possible to re-analyze older samples to provide higher quality sequences, but these are uploaded as new sequences to the database. The result is that sequence databases contain many entries that are incomplete, duplicate, or just plain wrong.

At the same time, these databases have no established mechanisms for correcting errors. They have adopted a model from scientific publication: once the sequence is published in the database, it is not changed. It is difficult if not impossible for a future scientist to update, correct, or even post a comment about a sequence. Community norms, while functioning to advance science, also serve to enact barriers to the adoption of potentially useful collaborative technologies.

CONCLUSION

Collaboration technologies become more important as scientific collaborations become increasingly large, distributed, complex, and data intensive. At the same time that collaborations become increasingly complex, so do the collaborative technologies and the barriers to their adoption. As we have noted in this brief discussion barriers to adoption take many forms. Many barriers to adoption result from the interplay of the social and the technical in the sociotechnical system.

A holistic understanding of large scale scientific collaborations as large, complex organizations that contain numerous groups of collaborators with different backgrounds and information needs will help ensure that barriers to adoption are not inadvertently built in. In simple terms, collaborative technologies must be appropriate to the needs and abilities of the precise subgroups that they support. As we have seen, the needs of each user groups in cyberinfrastructure (CI) may often be relatively invisible to others, including but not limited to management. Therefore great care must be taken to identify the various user groups and their respective needs.

Related to the point above, another significant barrier to adoption is the desire to have the latest and most sophisticated collaboration technologies. While the technologies used to collect and analyze the data often must be cutting edge, specifically collaborative technologies are often best served by technologies that are simple, require very little learning, and are already easily accessible (e.g. wikis, and telephones). Most scientists were reluctant to invest more than very small amount of time to learn to use new technologies unless the benefits were substantial and related directly to their research.

Needs for local vs. global control of tools, information, and data resources can change over time and given how the organizational structures of CIs change over time, allowing flexibility so that over time control over resources can be reallocated or shared as appropriate will assist with adoption, but also perhaps sustained use over time.

In the metagenomics domain, there is absolutely an opportunity to make these databases more "social." Allowing community members to comment on, discuss, and modify existing sequences could provide significant added benefit to the database. Scientists could flag or fix errors, create links to related sequences, or combine duplicate sequences into single records. Someone could add important metadata about a sequence from other sources (e.g. a journal publication).

Of course, there are barriers to implementing such practices. Community norms revolve around authorship, and allowing one scientist to edit another's sequence may not be acceptable. Typical review processes in the sciences rely on human editors to select reputable and knowledgeable reviewers. Scientists may insist on limiting the ability to edit or comment in the database to screened contributors.

Finally, it would be unwise to assume that if these "social" features are implemented scientists would rush to use them. It is important to understand what incentives are necessary to convince scientists to spend their scarce time on improving the quality of these community resources.

ACKNOWLEDGMENTS

We thank the participants of BIS and CAMERA. This work has been funded in part by NSF Grant #0712994.

REFERENCES

1. Lee, Charlotte P., Paul Dourish, Gloria Mark. 2006. "The Human Infrastructure of Cyberinfrastructure." Computer Supported Cooperative Work Conference (CSCW), Banff, Canada.
2. Marshall, E. Bermuda Rules: Community Spirit, With Teeth. *Science, 291*, 5507 (2001), 1192.
3. National Research Council (U.S.). Committee on Metagenomics: Challenges and Functional Applications. (2007). *New science of metagenomics: revealing the secrets of our microbial planet*. National Academies Press.
4. Orlikowski, Wanda J. Learning from Notes: Organizational Issues in Groupware Implementation. In: Mantel, Marilyn and Baecker, Ronald M. (eds.) *Proc. of the 1992 ACM conference on Computer-supported Cooperative Work*, 1992, 362-369.

5 Smarr, L. (2006, March 21, 2006). Creating a Community Cyberinfrastructure for Advanced Marine Microbial Ecology Research and Analysis (a.k.a. CAMERA). *Strategic News Service*.