MATTHEW J. BIETZ
RESEARCH STATEMENT

My work is driven by a desire to *explain and improve distributed collaborative knowledge work*. My research shows how the everyday practices of individuals and small groups can be brought together to produce larger collaborative effects. Many of today's societal and scientific challenges, especially in areas like human health and the environment, require the coordinated efforts of many people. I want to make these collaborations more effective.

I use sociotechnical theories to illuminate the complex interplay of social and technological phenomena. My methods include ethnographic observations of collaborative work, experimental studies of mediated communication, and interviews and surveys to understand larger collaboration patterns. Using multiple methods, I provide rich descriptions, develop robust theory, and create opportunities to improve collaborative practice.

I have a strong record of publication, and my work has appeared in journals and peer-reviewed conferences. I have experience preparing grant proposals, including a funded proposal to the NSF Virtual Organizations as Sociotechnical Systems program that supported my post-doctoral work. I have a strong commitment to my field, and I have organized conferences and workshops and served as a guest editor for a special issue of *CSCW*.

My research on distributed collaboration has spanned a number of projects, including:

- **International AIDS Research Collaboratory**: I worked with an internationally distributed HIV/AIDS research project to study and improve communication and data sharing.
- **Mediated Feedback Communication:** My dissertation involved experimental studies of how the affordances of communication media affect the way that critical feedback is delivered and interpreted.
- **Cyberinfrastructure for Metagenomics:** As a post-doctoral scholar, I worked with Dr. Charlotte Lee to investigate the development and use of a cyberinfrastructure for metagenomics research. Through qualitative interviews and observations in microbiology laboratories and software development teams, we explored the interactions of scientific practice and technologically-based infrastructures.
- **Collaborative Cyberinfrastructure Development:** During my post-doc, Dr. Lee and I co-wrote a proposal that was funded by the NSF. This comparative study of cyberinfrastructure development at two national supercomputing centers examined how human and technological resources are collaboratively shared and re-used across different scientific domains.
- **Trust in Distributed Software Engineering:** In a project led by Dr. David Redmiles, I am studying the role of trust in globally distributed software development. Early results suggest that both culture and geographic location play important roles in how trust forms in software teams.
- **Collaboration Success Wizard:** I am working with Dr. Gary Olson and Dr. Judith Olson on a project to validate and expand the Theory of Remote Scientific Collaboration [6]. Using a "translational science" paradigm, we have developed an online survey instrument that not only collects data for research purposes, but also provides feedback to participants about successful collaborative

practices. Our goal is not only to test and refine the theory, but also to create a useful and widely applicable instrument that can be used to inform and enhance collaborative work.

Three main themes drive my work and will be discussed in detail below. First, *I want to understand and improve distributed collaborative knowledge work*. How do people collaborate over distance, what makes that work successful (or not), and how can we develop strategies and technologies that make collaboration easier and more effective?

Second, *I focus on the development and use of infrastructures for collaboration*, especially in science and engineering. These "cyberinfrastructures" not only provide for the distribution of resources, but they also support the aggregation of local practices to produce greater combined effects. My work studies how these infrastructures come into being, and how they shape and are shaped by those who use them.

Third, *I have an abiding interest in the roles that data play in collaborative work*. Huge increases in data volume and shifts toward more data-intensive work are affecting not just academia but also corporate, government, military and intelligence sectors. In the face of the "data deluge," it is more important than ever to understand the collaborative creation, management, and use of large-scale data sets.

In the next sections I summarize findings from my recent work and outline my research expectations for the near future. While not a complete catalog of my research, these examples demonstrate how I approach the themes outlined above in the context of specific research projects.

Supporting Distributed Collaborative Knowledge Work

The scale and scope of collaborative activity has increased with new communication and information technologies and trends toward globalization of work. The distribution of collaborators across large geographic and cultural distances and the use of mediated communication technologies as the primary mode of interaction create challenges for successful collaboration. My research examines how distribution and technology affect interpersonal interaction within collaborative projects, and suggests strategies and technologies that can make collaboration more effective.

As a doctoral student, I studied HIV/AIDS researchers who were conducting research at laboratories in the United States, the United Kingdom, and southern Africa. The researchers were conducting comparative studies at multiple clinics and laboratories. It was crucial that the procedures in all sites be carried out in the same way so that the findings would be comparable. From interviews and observations, we discovered that communication was limited, not only by the high cost of international calls, but also by significant culture, status, and technology differences among the sites. Building on the traditional lab meeting as a space where scientific issues could be raised and addressed in a structured but friendly space, we worked with the scientists to implement technologies and practices for distributed lab meetings. These real-time interactions became an important forum for lab discussions, and the participants adopted the technology to support ad-hoc and mentor-student interactions as well. The framework we put in place is still being used to train clinicians and allow African doctors to discuss treatment plans with experts around the world.

Even though the lab meetings were considered successful, I observed the frustrations that arose out of computer-mediated discussions between African students and their mentors in the US. The senior scientists were providing feedback to their students, but giving critiques of someone's work is fraught with relational complexity. Negative critiques carry a risk of emotional and relational harm, and making feedback constructive rather than destructive requires careful crafting of the feedback message and

emotional cues. The students were having difficulty interpreting the feedback they were given, and that the mentors were having difficulty adjusting their feedback to the emotional state of the individual students.

My dissertation was an experimental investigation of this phenomenon. I conducted experiments in which participants used various communication media to give each other feedback about documents they had written. The experiments showed that communication media have important effects on the way feedback is communicated and interpreted [1]. For example, critics tend to give more feedback and explain their feedback more when giving feedback using videoconferencing instead of IM. When the recipients could not see or hear the critic, they were less likely to trust that the critic was giving honest feedback. Recipients are more likely to accept feedback delivered over videoconferencing instead of IM. The data also suggest that men and women may adapt to the affordances of the communication technology differently.

My concern for the affective and relational aspects of collaboration carries over into a current project with Dr. David Redmiles to understand the role of trust in globally distributed software engineering. We have conducted interviews with more than forty individuals who work on globally distributed software teams in a large engineering organization. Early results suggest that trust is affected by both geographic distance and cultural difference. In particular, difficulty in interpreting cultural cues or contextual information may lead to trust breakdowns. We are currently conducting interviews with software engineers from other companies that are not based in the United States to further explore culture and distribution effects.

One of the benefits of the kind of work I do is the opportunity to have a practical impact in the world. In a project with Dr. Gary Olson and Dr. Judy Olson, I am working on the "Collaboration Success Wizard," an online survey designed as both a research instrument and a diagnostic tool. The Wizard is based on the Theory of Remote Scientific Collaboration (TORSC), a theory that explains collaboration success as a product of several factors including the nature of the work, common ground, collaboration readiness, project management, and technology readiness [6]. In the current project, we have adopted a "translational science" paradigm that works to translate research findings into practical applications. The Wizard is an online survey administered to participants in collaborative projects in science, engineering, and other domains. Data from the Wizard will be used to validate and refine the TORSC, but it also provides immediate feedback to the participants. At the end of the survey, participants are provided with an individual report that gives feedback about collaborative strengths and weaknesses indicated in their answers. Each project is also provided with a summary report that details their collaborative practices and suggests strategies to improve the collaboration.

The Work of Infrastructure Development

A second thread in my research on distributed collaboration has been to understand the processes that underlie cyberinfrastructure development projects. Cyberinfrastructure refers to the complex sets of human and technological arrangements (of high-performance computing, high-speed networking, and large data repositories) that support distributed knowledge work. Cyberinfrastructure development provides rich fodder for study—the goal of cyberinfrastructure is to support collaboration, and the development of cyberinfrastructure is highly collaborative in and of itself.

I have been working on two separate studies: one following the development of cyberinfrastructure for metagenomics research, and a second comparative study of how supercomputer centers build cyberinfrastructure for multiple domains. In both of these projects I investigate the work of

cyberinfrastructure developers, especially the challenges of designing large-scale distributed systems in the context of collaborative virtual organizations.

In a 2010 *CSCW* journal article, we coined the term *synergizing* to theorize the collaborative work of developing cyberinfrastructures [3]. Synergizing refers to the work necessary to build and maintain the relationships that produce greater combined effects than individuals, groups, or organizations could make on their own. Synergizing highlights the relational nature of infrastructure and illuminates how developers work to bring organizational, individual, and technological entities into productive alignment. We follow up on this work in an article just accepted to the CSCW 2012 conference that uses the synergizing concept to explore how developers work to ensure the long-term sustainability of cyberinfrastructures.

I am currently developing a proposal to examine how synergizing activities are supported by community standards. Standards are a particular class of infrastructures that support collaboration by promoting interoperability. This project will focus on community standards organizations in the biological sciences, including the BioSharing project. BioSharing is a kind of community-of-communities, focused on "the production of an integrated standards-based framework" that brings together several relevant standards-making organizations in genomics. For example, the Genomic Standards Consortium (GSC) consists of a collection of biologists, bioinformaticists, and information technologists to develop standards covering data interchange formats, minimum metadata standards, standard operating procedures for DNA sequencing, and ontologies for describing genomic data. The GSC is also working to disseminate and promote adoption of their standards by working closely with stakeholder communities, advocating for standards adoption by equipment manufacturers and large databases, and even founding a journal to support publication of standards. The GSC has invited me to work with them, and I am developing a proposal to better understand how community-developed scientific standards support large-scale collaboration.

Data-Centric Collaboration

Increasingly knowledge work depends on massive collections of data. These collections are often produced through the contributions of large numbers of individuals, and they are typically so large that collaboration is the only way to fully analyze and understand them. I believe that understanding these collaborative practices of data generation, sharing, and analysis is imperative to ensuring the validity and effectiveness of these new ways of knowing.

In a 2003 paper, Jeremy Birnholtz and I compared data sharing practices in HIV/AIDS research, earthquake engineering, and space physics [4]. We found that data play an important role in the social as well as scientific life of the laboratory. Data are implicated in the creation and maintenance of community boundaries, in assigning credit and reputation, and in enculturating new researchers. These social considerations affect scientists' decisions to share or not share their data. Designing systems for data sharing requires recognizing that data are socially as well as scientifically meaningful.

My investigations in the field of metagenomics reveal that large shared databases act as *boundary negotiating artifacts*: they provide a venue where community boundaries can be defined and modified as decisions get made about which data to include in the database and how it should be represented [2]. Metagenomics refers to the study of populations of microorganisms using genetic techniques. Advances in laboratory and computational technologies allow researchers to directly sequence the DNA of entire populations of organisms (for example, all of the bacteria in a sample of seawater). These researchers rely on large centralized shared databases of genetic data which serve both as a baseline to which newly

sequenced DNA can be compared and as a model through which to understand the microbial world. Even though these collections are shared among the community, sub-communities organize around particular research questions and approaches have a particular stake in what the database should provide [5]. Through the design of these database, these sub-communities engage in active negotiation of community boundaries and scientific meaning.

Data-centric collaboration will be a key theme as my research moves forward. These issues are the focus of a workshop I will be leading at the CSCW 2012 conference. They also form the basis of a grant proposal (under review) to develop community-owned repositories of phylogenetic knowledge. Phylogenetics is a field of science concerned with understanding the evolutionary history of organisms by studying their genetic material. As the field grows and progresses, researchers are eager to have a centralized, high-quality repository for phylogenetic data. Two current databases—the Tree of Life Web and TreeBASE—provide a place to store data, but lack support for community-level curation, collaboration and interaction around the datasets. The proposed project would develop the social aspects of the systems in an attempt to engage and nurture data-focused communities of practice. These features would allow users to annotate existing data, engage in discussions around the datasets, and take collective responsibility for areas of the database. It is hoped that these interventions will help to improve the use, quality, and relevance of these databases for phylogenetics researchers.

My Scholarly Commitment

My research grows from a fascination with the ways that people develop practices and create meaning through everyday collective activity. I depend on the rich data produced by qualitative methods. I rely on observation, interviews, and critical analysis of artifacts to build theory about the role of information technologies in distributed collaborations. But it is important to me to match the methodological approach to the question at hand, and I also conduct quantitative studies (including both experiments and surveys) that complement and validate the qualitative work.

I believe that knowing and doing go hand in hand. In my work I strive to not only provide rich theoretical understanding of collaborative knowledge work, but also to make a difference in the world. Working with HIV researchers was a way to understand international collaboration, but also a way to improve the lives of AIDS patients in Africa whose doctors would now have better access to cutting-edge expertise. With the Collaboration Success Wizard, we are working to not only validate theory, but also to provide useful diagnostics and advice to collaborative projects. This approach is more than merely “giving back”—I believe that practical experience and theoretical knowledge produce and enhance each other, and that my research is stronger when it combines both.

I have a demonstrated record of publication as a doctoral student and post-doctoral scholar. My research has appeared primarily in the Human-Computer Interaction domain, especially in the Computer Supported Cooperative Work (CSCW) conferences and journals. I am active in the growing community of scholars studying cyberinfrastructure and e-science, and I have organized workshops and co-edited a journal special issue on the topic. I am also involved with Social Studies of Science (SSS) and Science and Technology Studies (STS) communities. I have been involved in writing successful grant proposals, and have begun to submit proposals as a PI to further my own research agenda.

As I move ahead in my career, I hope to establish a research group to push my research agenda forward and train the next generation of scholars. I am excited to get into the classroom and share what I know about collaboration, communication, and HCI with students. I want to be a tenure-track faculty member.

References

1. Bietz, M. J. (2008). Effects of communication media on the interpretation of critical feedback. In *Proceedings of the ACM 2008 Conference on Computer Supported Cooperative Work* (pp. 467-476). New York: ACM Press.
2. Bietz, M. J., & Lee, C. P. (2009). *Collaboration in metagenomics: Sequence databases and the organization of scientific work.* Paper presented at the ECSCW 2009, London.
3. Bietz, M. J., Lee, C. P., & Baumer, E. P. S. (2010). Synergizing in cyberinfrastructure development. *Computer Supported Cooperative Work, 19*(3-4), 245-281.
4. Birnholtz, J., & Bietz, M. J. (2003). Data at work: Supporting sharing in science and engineering. In *Proceedings of the 2003 International ACM SIGGROUP Conference on Supporting Group Work* (pp. 339-348). New York, NY: ACM Press.
5. Lee, C. P., Bietz, M. J., & Thayer, A. (2010). Research-driven stakeholders in cyberinfrastructure use and development. In *Proceedings of the 2010 International Symposium on Collaborative Technologies and Systems (CTS 2010)* (pp. 163-172). Chicago, IL: IEEE.
6. Olson, J. S., Hofer, E. C., Bos, N., Zimmerman, A., Olson, G. M., Cooney, D., et al. (2008). A theory of remote scientific collaboration. In G. M. Olson, A. Zimmerman & N. Bos (Eds.), *Scientific Collaboration on the Internet* (pp. 73-97). Cambridge, MA: MIT Press.