Matthew J. Bietz
Research Statement

My research focuses on *the design and use of information technologies to support distributed collaborative knowledge work*, with a particular focus on the cognitive aspects of large-scale data sharing and use. We are in an era of Big Data: 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. Yet we continue to struggle with how to create knowledge out of this "data deluge." My work shows that as the size of datasets increase, they also tend to become more sociotechnically complex. More people will be involved in the creation and use of the data, the data will be expected to serve a larger number of purposes, and datasets will be expected to remain relevant over longer periods of time. Understanding this complexity drives my research.

My work lives at the intersection of three core themes:

1. **Design for Data**: How can we design better systems for data creation, organization, sharing, and use? What are the cognitive and social consequences of technologies for data management, analysis, and visualization?
2. **Distributed collaborative work**: How do people collaborate over distance, what makes that work successful (or not), and how can we develop strategies and information technologies that make collaboration easier and more effective?
3. **Large-scale information infrastructures**: How do the "cyberinfrastructures" that provide the foundations for Big Data science and engineering come into being, and how do they support new forms of knowledge production? How must the design process change to address different scales of technological artifacts?

As an example of how these themes come together, I am in the second year of a 3-year grant from the National Science Foundation's *Virtual Organizations as Sociotechnical Systems* program entitled "Virtual Standards Development Organizations: Enhancing Interoperability in Data-Intensive Science." This project focuses on grass-roots virtual organizations that are working to design data standards for the genomic sciences. These organizations bring together scientists, technologists, and other stakeholders from academic, corporate and government sectors around the world to create standards that allow interoperability across genomic sequencers, large-scale databases, and high-performance data analysis systems.

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, and have received funding from the National Science Foundation, the Intel Science and Technology Center for Social Computing, and the Robert Wood Johnson Foundation. 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 Approach to Design

As someone who is interested in infrastructures and large-scale systems, I see a need for a set of design practices that embrace complexity across multiple scales. My understanding of design is heavily influenced by Furnas' elaboration of the Mosaic of Complex Adaptive Systems (MoRAS) [12] and by the conception of infrastructure proposed by Star and Ruhleder [19]. Both envision a complex design space of multiply interconnected dynamic systems—"including people's heads, organizations, communities, markets, and cultures" [12, p.205]—that respond and adapt to each other. There has been a tendency to bound design within relatively narrowly-defined and insulated domains: IT, organizations, architecture, policy, medicine, etc. My work demonstrates repeatedly that these are not independent design spaces, and that changes in one domain have contexts and consequences in the others. The scope of today's societal challenges requires an interdisciplinary approach to design that embraces the dynamic and complex nature of interactions among artifacts, systems, networks, and infrastructures.

My interest in infrastructure should not suggest that I limit my design interest to large-scale systems. Rather, I advocate for an approach that brings an understanding of broad contexts into the design process regardless of scale. Recently, building on my interest in self-quantification, I have begun working with students to design new types of wearable sensors for monitoring one's self and environment. This provides an opportunity to ask questions about how individuals develop a sense of self and use sensors and data as tools to understand their bodies and lives. This work is aligned with critical design [10] and critical making [17]: we are using design as a way to explore cognitive, social, cultural, and political issues around information technology.

My research on the design of cyberinfrastructure suggests a need for improved design processes and practice. One aspect of this is to extend our notion of design beyond the initial phase of early innovation and specification. Designed artifacts and systems often have long lives, and they tend to develop and change over time [5; 8; 14]. Ribes and Finholt [18] describe the infrastructural tension between development and maintenance, but my work shows that design and innovation occur in all phases of the infrastructural lifecycle: "the day-to-day work of maintaining… cyberinfrastructure is not fundamentally different than the work of developing new capabilities" [3, p.908]. Design for long-lived systems and sustainability will not be successful if we restrict the design focus to the beginning of a project.

Another finding from my research is that design is a value-laden process—the values of designers, users, and other stakeholders affect both the design process and its outputs [4; 13]. I have been involved with the Values in Design movement, including being faculty for the 2012 VID graduate student workshop. Values in Design provides a framework for exploring the value propositions of designed artifacts and

systems (and the cognitive and social consequences of those values), understanding values and potential value conflicts among stakeholders, and developing a set of practices for value-driven design.

I see my work fitting into each of Frayling's categories of relationships between research and design [11; 20]. I conduct *research into design* when I study the work and design processes through which infrastructures are created. I engage in *research for design* in the development of new design methods and practices like Values in Design. In both my engagement with critical design and my collaborations with developers of infrastructure, I explore *research through design*, using the act of making as a way to probe and transform the world. Design will continue to be an important part of my work as a scholar.

Research Projects

My research has spanned a number of projects, including (most recent first):

- **Virtual Standards Development Organizations**: I am the PI on this NSF-funded project to study the collaborative development of data standards for the genomic sciences.
- **From Cyberspace to Outer Space:** I am collaborating with Janet Vertesi (Princeton), Paul Dourish (UC Irvine) and Melissa Mazmanian (UC Irvine) to study the distributed organization and data sharing of large-scale planetary exploration teams, including the *Cassini* spacecraft mission to Saturn.
- **Health Data Exploration:** A variety of health-relevant parameters can now be easily captured via wearable devices, smartphones and other tools that may yield insights about personal and population health. With funding from the Robert Wood Johnson Foundation, we are exploring how to make these data streams available in meaningful and responsible ways for public health and social science research.
- **Genomics and the Quantified Self:** This project, supported by the Intel Science and Technology Center for Social Computing, is beginning to characterize the ways that health and activity data from personal genetic testing and personal sensing devices allow for and support new understandings of the self and well-being.
- **Collaboration Success Wizard:** I am working with Gary Olson and Judith Olson (UC Irvine) on a project to validate and expand the Theory of Remote Scientific Collaboration [16]. 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.
- **Trust in Distributed Software Engineering:** In a project led by David Redmiles (UC Irvine), I studied the role of trust in globally distributed software development. This work showed how both culture and geographic location play important roles in the way trust forms in software teams.
- **Cyberinfrastructure for Metagenomics:** As a post-doctoral scholar, I worked with Charlotte Lee (University of Washington) 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.
- **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.
- **International AIDS Research Collaboratory**: As a doctoral student, I worked with an internationally distributed HIV/AIDS research project to study and improve communication and data sharing.

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.

Design for Data

Increasingly knowledge work depends on massive collections of data. These collections are often produced through the contributions of many 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 for designing valid and effective systems to support 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 [9]. 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 [4]. Metagenomics 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 broad scientific community, sub-communities organize around particular research questions, and each has a particular stake in what the database should provide [15]. Through the design of these database, these sub-communities engage in active negotiation of community boundaries and scientific meaning.

Building on my work in the metagenomics area, I received a 3-year NSF grant to examine the development of community data standards in genomics. This project focuses on virtual standards development organizations in the biological sciences, including the Genomic Standards Consortium and the BioSharing project. These organizations bring together biologists, bioinformaticists, information scientists, and technologists to develop standards covering data interchange formats, minimum

metadata standards, standard operating procedures for DNA sequencing, and ontologies for describing genomic data. This study will provide an opportunity to investigate the promise and limits of standards in large-scale, big data science.

Two recent projects explore the ways that data-centric technologies and artifacts shape knowledge production. In a study of NASA's Cassini Mission, I am investigating the processes of data calibration across multiple scientific teams to understand how the mix of physical instruments, computational algorithms, social processes, and individual expertise are all enrolled in the creation of knowledge about Saturn and its moons. With a similar motivation, but in a very different domain, I am studying the ways that self-quantification technologies like pedometers, sleep monitors, and personal genetic testing are being used by individuals to develop knowledge about and a new relationship with their bodies and selves. I expect to publish on both of these projects within the next year.

Supporting Distributed Collaborative Knowledge Work

Improving distributed collaboration is a theme that runs through all of my research, and it has been a specific focus in several projects. 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.

As a doctoral student, I studied HIV/AIDS researchers who were conducting collaborative research at laboratories in the United States, the United Kingdom, and southern Africa. The researchers were performing comparative studies at multiple clinics and laboratories, and 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. We worked with the scientists to implement technologies and practices for distributed lab meetings [7]. 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 certain miscommunications that arose out of computer-mediated discussions between African students and their mentors in the US. 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.

One of the benefits of the kind of work I do is the opportunity to have a practical impact in the world. With Gary Olson and Judy Olson, I am working on the "Collaboration Success Wizard," an online survey designed as both a research instrument and a diagnostic tool for collaborative projects in science, engineering, and other domains [2]. Data from the Wizard will be used to validate and refine the theories of scientific collaboration, 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 their collaboration.

Large-Scale Information Infrastructures

I also work to understand the processes that underlie the creation and use of cyberinfrastructure. 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.

As a post-doctoral researcher I worked 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 investigated 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 [6]. 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 provides a framework for approaching the design of large information systems as a multi-faceted and multi-layered process of bringing organizational, individual, and technological entities into productive alignments.

A key concern in the development of cyberinfrastructure is understanding how to sustain these infrastructures over the long term. Science tends to operate on relatively short funding cycles, but cyberinfrastructures are expected to operate on decades-long time scales, and it is crucial to preserve the knowledge contained in them even longer. I have published several articles that explore how developers work to ensure the long-term sustainability of cyberinfrastructures and scientific knowledge [3; 5; 8; 14].

My Scholarly Commitment

My research grows from a fascination with the ways that people develop information practices and create meaning through everyday collective activity. I depend on the rich data produced by field

methods. I rely on qualitative and quantitative observation, interviews, and critical analysis of artifacts to build theory about the design of data and information technologies. But it is important to me to match the methodological approach to the question at hand, and I also conduct experiments and surveys that complement and validate the fieldwork.

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. I am working with the Values in Design Laboratory at UC Irvine to approach the design of artifacts, systems, and infrastructures as a way to make positive societal impacts. 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, post-doctoral scholar, and research scientist. 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, Technology and Society (STS) communities. I have written successful grant proposals in order to develop my own research agenda.

As I move ahead in my career, I hope to advance my research agenda, collaborate with colleagues, and train the next generation of scholars. I look forward to realizing my commitment to research, teaching, and service as 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., Abrams, S., Cooper, D., Stevens, K. R., Puga, F., Patel, D., et al. (2012). Improving the odds through the Collaboration Success Wizard. *Translational Behavioral Medicine, 2*(4), 480-486.
3. Bietz, M. J., Ferro, T., & Lee, C. P. (2012). Sustaining the development of cyberinfrastructure: An organization adapting to change. In *CSCW '12: Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work* (pp. 901-910). New York: ACM.
4. 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.
5. Bietz, M. J., & Lee, C. P. (2012). Adapting cyberinfrastructure to new science: Tensions and strategies. In *Proceedings of the 2012 iConference* (pp. 183-190). New York: ACM Press.
6. Bietz, M. J., Lee, C. P., & Baumer, E. P. S. (2010). Synergizing in cyberinfrastructure development. *Computer Supported Cooperative Work, 19*(3-4), 245-281.

7. Bietz, M. J., Naidoo, M., & Olson, G. M. (2008). International AIDS Research Collaboratories: The HIV Pathogenesis Programme. In G. M. Olson, A. Zimmerman & N. Bos (Eds.), *Scientific Collaboration on the Internet* (pp. 351-364). Cambridge, MA: MIT Press.

8. Bietz, M. J., Paine, D., & Lee, C. P. (2013). The work of developing cyberinfrastructure middleware projects. In *Proceedings of the 2013 Conference on Computer Supported Cooperative Work* (pp. 1527-1538). New York: ACM.

9. Birnholtz, J., & Bietz, M. J. (2003). Data at work: Supporting sharing in science and engineering. In *Proc. GROUP 2003* (pp. 339-348): ACM Press.

10. Dunne, A., & Raby, F. (2001). *Design Noir: The Secret Life of Electronic Objects*. Basel, Switzerland: Birkhäuser.

11. Frayling, C. (1993). Research in art and design. *Royal College of Art Research Papers, 1*(1), 1-5.

12. Furnas, G. W. (2000). Future design mindful of the MoRAS. *Human-Computer Interaction, 15*(2), 205-261.

13. Koehne, B., Bietz, M. J., & Redmiles, D. (2013). Identity Design in Virtual Worlds. In Y. Dittrich, M. Burnett, A. Mørch & D. Redmiles (Eds.), *End-User Development: Proceedings of the 4th International Symposium, IS-EUD 2013* (pp. 56-71). Berlin: Springer.

14. Lee, C. P., Bietz, M. J., Derthick, K., & Paine, D. (2012). A Sociotechnical Exploration of Infrastructural Middleware Development. In *CSCW '12: Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work* (pp. 1347-1350). New York: ACM.

15. 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.

16. 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.

17. Ratto, M. (2011). Critical Making: Conceptual and Material Studies in Technology and Social Life. *The Information Society, 27*(4), 252-260.

18. Ribes, D., & Finholt, T. A. (2009). The Long Now of Technology Infrastructure: Articulating Tensions in Development. *Journal of the Association for Information Systems, 10*, 375-398.

19. Star, S. L., & Ruhleder, K. (1996). Steps toward an ecology of infrastructure: Design and access for large information spaces. *Information Systems Research, 7*(1), 111-134.

20. Zimmerman, J., Forlizzi, J., & Evenson, S. (2007). *Research through design as a method for interaction design research in HCI*. Paper presented at the Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.