DISCIPLINARY HISTORIES, FUNDERS' PRIORITIES, AND COMPETING THEORIES: ORGANIZATIONAL FORCES BEHIND CYBERINFRASTRUCTURE DEVELOPMENT IN E-SCIENCE

Kerk F. Kee

The University of Texas at Austin

kerk.kee@gmail.com

4S Conference, Oct 31, 2009, Crystal City, VA.

Kee, K. F. (2009, October). Disciplinary histories, funders' priorities, and competing theories: Organizational forces behind the development of cyberinfrastructure in e-science. Paper presented at the *Society for Social Studies of Science* annual conference, Arlington, VA.

INTRODUCTION

Cyberinfrastructure (CI) refers to a collection of information, communication, computer technologies and human experts (Atkins et al, 2003). According to Stewart (2007), "Cyberinfrastructure consists of computing systems, data storage systems, advanced instruments and data repositories, visualization environments, and people, all linked together by software and high performance networks to improve research productivity and enable breakthroughs not otherwise possible" (¶. 3). In the influential *Atkins Report*, Atkins and colleagues (2003) further state that CI is "an effective and efficient platform for the empowerment of specific communities of researchers to innovate and eventually revolutionize what they do, how they do it, and who participates" (p. 5). In essence, cyberinfrastructure represents a collection of machines and humans, as well as the social interactions and cultural practices surrounding the meshing of the two. CI is inherently social, political, and organizational. The study of CI to support science can includes an examination of the organizational forces related to its design during CI development.

One approach to examine the organizational issues related to the design and development of CI is to employ the lens of organizational communication. An organizational communication approach posits that "communication theory can be used to explain the production of social structures, psychological states, member categories, knowledge, and so forth rather than being conceptualized as simply one phenomenon among these others in organizations" (Deetz, 2001, p. 5). The focus would be on the process of organizing through symbolic interaction that showcases that the communication is the organization rather than on 'communication' within an 'organization' (Hawes, 1974). Organizational communication perspectives are "especially concerned with the constitutive role of communication in shaping organizational reality and with examining how communication serves the interests of some organizational interest groups more than others" (Mumby & Stohl, 1996, p. 57). The communicative constitution of organizations (Putnam & Nicotera, 2009) is particularly applicable for a study of cyberinfrastructure design and development because like constitution theory, the analysis focuses on how cyberinfrastructure is built up over time by commitments across institutions and people.

In this paper, I attempt to identify three organizational issues related to the development of CI by taking an organizational communication approach. More specifically, I ask the research question: "What organizational forces communicate influence to the design of cyberinfrastructure during development?" By asking this question, I highlight three organizational forces that communicate influence to the design of CI. I present the preliminary findings in terms of questions to be asked as we design CI to support science and emerging groups. For the rest of the paper, I will explain the research methods and data collection process, the three key findings, and a brief conclusion.

METHODS AND DATA COLLECTION

Data collection for this project started in November 2007 at a supercomputer center in Texas and the project eventually expanded into 65 participants (with 68 separate interviews) by August 2009 across 17 US states and three countries in addition to the US. Interviews were conducted with over three years with 8 participants in 2007, 41 in 2008, and 16 in 2009. The shortest interview is about 15 minutes and the longest is 2 hours and 16 minutes. The interviews average approximately one hour each and were conducted in person with 16 of the participants and over the phone with the remaining 49 participants. This study sampled participants based on specialized knowledge on the topic, so snowball recruitment (Johnson 1990; Sætre, et al. 2007) was chosen as the appropriate strategy. In addition to in-depth interviews, observation was also a part of the data collection process. I attended and observed the 2008 Supercomputing Conference (SC08) in Austin, Texas, and several public events organized by the Texas Advanced Computing Center in 2007, 2008, and 2009. The third data source came from careful reviews of homepages of the National Science Foundation and key supercomputer centers across the country.

The 65 interview participants come Texas (11), Illinois (11), California (10), Michigan (5), Indiana (4), Virginia (3), Massachusetts (3), Arizona (2), Colorado (2), Louisiana (2), Washington (2), DC (1), Maryland (1), New (1), Ohio (1), Pennsylvania (1), Delaware (1), as well as Australia (2), Germany (1), and the UK (1). Interviewees' professional roles in CI projects are diverse, including domain scientists who use CI to conduct science (16), computational technologists who work to build CI (12), a diverse range of administrative directors and program managers at supercomputer centers and national research laboratories across the country (20), NSF program officers who help allocate funding to CI projects (3), and social scientists and policy analysts who are sensitive to the social, organizational, and political dynamics of CI projects (12). These roles are the primary roles of the participants at the time of the interviews. However, many wear multiple hats and have multiple backgrounds and disciplinary expertise. Some technologists are professors of computer science and they also engage in research. However, they are labeled as technologists in this study because they are involved in the technological aspects of CI projects

Partially influenced by the philosophy of grounded theory (Corbin and Strauss 1990), sense making and data analysis of the interviews took place throughout the entire data collection process. However, a more systematic thematic analysis was completed after all the interviews were transcribed. This analysis of this paper is based on a general qualitative procedure (McCracken 1988) that leads to a specific thematic analysis of interview transcripts (Owen 1984). McCracken's (1988) procedure has the researcher follow these steps: (a) sort out important data from unimportant data; (b) examine the slices of data for logical relationships and contradictions; (c) re-read the transcripts to confirm or disconfirm emerging relationships and to scan for the general properties of the data; (d) identify the general themes and sort the themes in a hierarchical fashion, while discarding those that prove useless in the organization; and (e) review the emergent themes for each of the transcripts and determines how they can synthesized into a still wider set of overarching themes.

FIRST QUESTION: DISCIPLINARY HISTORIES

The first question asks, "*What is a field's history with the Internet and computer-supported cooperative work?*" The Internet, computers, and a wide range of emerging information and communication technologies shape today's organizational life [4, 6, 7]. Every scientific field has a history with the Internet and computer technologies at work. This history will affect how scientists in a field approach CI, and how CI design will impact their work. The younger the field, the more receptive the scientists will be to doing computer-supported cooperative work on CI. As an interview participant shares,

> [L]et's take bioinformatics [as an example]… The use of the Internet to do the science is dominant. That's very different, say from chemistry… Because the field is much older… If you look at biology versus chemistry, there are dramatic differences in the field, in the nature – everything in biology is pretty open, as far as I know. It is all put on the Internet. And chemistry – nothing is open and almost nothing is put on the Internet… I think it mostly reflects history… Chemistry was born a very long time ago. The electronic support was born 30 years ago. It is inconsistent with the Internet cyberinfrastructure model. Whereas biology, bioinformatics, was born 10 years ago. So it grew up as the Internet was growing up. So biology almost started doing cyberinfrastructure without thinking… That field is richly cyber-enabled…. Science is evolutionary... If the previous step was on the Internet, the next step probably has to be on the Internet, by definition. So you're not able to not do cyberinfrastructure." (Professor of Informatics, Professor of Computer Science, and Professor of Physics, Indiana, 14 February 2008).

As we examine the issue of CI design, we need to distinguish the different disciplines of science, and take into consideration a field's history with the Internet and computer technologies. Scientific practices within a field are shaped by its history, and these practices change slowly by evolving over a long period of time. The design of CI should closely match existing practices, if CI is to be adopted and implemented to support a particular branch of science. Compatibility (Rogers, 2003) with existing practices is key.

Furthermore, an effective design for one field may not be equally useful for another. Different fields developed their unique ways of doing science, and these organizational practices rooted in past successes are difficult to change. CI design has to acknowledge the complexity and diversity in a wide range of disciplines and fields in science. If we need to build more CI to support different branches of science, funding is a key organizational issue to consider next.

SECOND QUESTION: FUNDER'S PRIORITIES

The second question to consider asks, "*Who is funding the CI project, and which agenda is the project advancing?"* Although science is often assumed to be a neutral endeavor simply to improve human conditions in the society, the organizational influence associated with funding agencies behind the scene is not neutral nor value free. Funding agencies only fund projects that promote and advance their missions and agendas. If an agency is to fund a particular project, the money is given only to conduct science relevant to the agenda of the funding agency, and not the agenda of another. Below is what a policy expert reveals,

> In some cases, in larger institutions, the problems are really magnified by the fact that faculty receive grants from NIH and other places, which did not encourage collaboration and joint usage of technology, but rather, waived them off…If [they] go out and get a Sun workstation on [their] desk, … put a couple of Condors together, and then when the funding runs out, they're left with this big bill to run these machines… They need help programming it if it's going to be used for anything other than the research they did initially." (Policy Expert, Washington DC, 8 February 2008)

Funding is perhaps the most powerful driving force behind large-scale science in the US. A CI project is very expensive, and without funding from agencies such as NSF and NIH, no CI can be built. These funding agencies allocate resources to CI projects on a limited term basis. Once the allocated funding is used up, and if the project cannot secure continuing support, the operation comes to an end, including the development of CI for the project. In addition, the agenda of the funding agency influences the design of CI to prioritize activities in scientific research. What does not serve the agenda does not get built into the design.

Furthermore, funding is not neutral. By receiving funding from a particular agency, acceptance of the agency's agenda is implied. Therefore, while discussing the design of CI to support science and emerging groups, it is important to keep in mind the political priorities communicated through funding to a particular project. If a CI project is to continue, the project has to continue advancing the agenda. The development of CI is not only closely aligned with the funding agency's agenda, the design of CI is also closely tied to the theoretical and methodological competitions within a field. This observation turns us to the third question.

THIRD QUESTION: COMPETING THEORIES

The third question asks, "*For which theory or method is the CI built?*" There are competing theories and methodologies within any disciplines and fields in science. A vibrant scientific community engages in a healthy debate about the different ideologies and approaches to doing science. However, when it comes to the design of CI to support science, we inevitably encounter the competition among these different groups of scientists who hold different philosophies of science. As the last informant in this position paper points out,

> We've been in disputes with people essentially having two different – not quite theories, but two methodologies to approach a problem. They would come to the cyberinfrastructure folks and say – We're glad to be on the project and of course you're going to include my methodology in the way the software works and exclude my competitor over there. (Supercomputer Center Administrator, Illinois, 23 January 2008)

The decisions made before and during the process in which CI is designed to support science and emerging groups involve persuasions, arguments, or even conflicts between groups. The theory or methodology selected to guide the design of CI determines which theoretical and methodological camp gains ground in advancing its approach to science. Scientists compete to influence CI design in favor of their own orientation, and persuade computational technologists to write codes and build applications that will support their method. This is a process to indirectly weed out competing theories and methodologies in the field. The design of CI becomes a contested terrain among competing groups of scientists.

CONCLUSION

In this paper, I employ an organizational communication approach to highlight three organizational issues that could affect the design of CI to support science. Through careful analysis of interviews from 65 participants, I presented three questions to consider while designing CI. The three questions include: "*What is a field's history with the Internet and computer-supported cooperative work?*"; "*Who is funding the CI project, and which agenda is the project advancing?";* and "*For which theory or method is the CI built?*" These questions reveal that the history of a field, funder's priority, and theoretical/ methodological commitment of scientists can influence decisions that go behind the design of CI to support science. A few implications can be drawn from these observations. I will discuss them in the last paragraph.

First, given the limited resources to build CI, the design of CI is best to be flexible in order to adapt to a wide range of scientific fields. When there are discipline specific requirements, parts of CI can be built to cater to these needs. Second, while it is important to bring older disciplines on board with CI, perhaps a good approach could be to focus on younger fields in order to create a critical mass in the overall scientific community. A critical mass naturally helps speed up diffusion and adoption. Third, CI projects may benefit from staying with one primary funding agency, or closely allying agencies, as trying to satisfy different agendas simultaneously or subsequently is difficult, especially when (re)building CI can be extremely costly. Fourth, CI design may best be neutral by creating a platform through which competing theories and methodologies can be tested on equal ground.

CONTRIBUTIONS TO STS

This paper contributes directly to STS scholarship in two ways. First, this paper extends Star and Bowker's (2006) argument of infrastructure as an "installed base" (p. 231) to consider a recursive relationship between organizational forces and infrastructure design. Star and Bowker contend a new technology "wrestles with the inertia of the installed base and inherits strengths and limitations from the base" (p. 231). This paper highlights the organizational forces that get built into the 'installed base' for future science.

Second, this paper shows that infrastructure design is political and complex, and an OC lens adds to STS scholarship. Bijker (1995) calls STS scholars to pursue "political questions" (p. 255). By employing an OC lens and presenting the findings in a form of questions, this paper reveals the political-cultural relevance of organizational forces in infrastructure design.

REFERENCES

Atkins, D. E., Droegemeier, K. K., Feldman, S. I., Garcia-Molina, H., Klein, M. L., & Messina, P. (2003). Revolutionizing science and engineering through cyberinfrastructure: Report of the National Science Foundation blue-ribbon advisory panel on cyberinfrastructure. Washington, DC: National Science Foundation. Retrieved December 19, 2006 from http://www.communitytechnology.org/nsf_ci_report/.

Corbin, Juliet and Anselm Strauss (1990): Grounded Theory Research: Procedures, Canons, and Evaluative Criteria. Qualitative Sociology, vol. 13, no. 1, pp. 3-21.

Deetz, S. (2001). Conceptual foundations. In F. M. Jablin, & L. L. Putnam (Eds.), The New Handbook of organizational communication: Advances in theory, research, and methods (pp. 3-46). Thousand Oaks, CA: Sage.

Hawes, L. (1974). Social collectives as communication: Perspectives on organizational behavior. *Quarterly Journal of Speech, 60,* 497-502.

Johnson, Jeffrey C. (1990): *Selecting Ethnographic Informants*. Newbury Park, CA: Sage.

McCracken, Grant (1988): *The Long Interview: Qualitative Research Methods*. Newbury Park, CA: Sage.

Mumby, Dennis K. and Cynthia Stohl (1996): Disciplining Organizational Communication Studies. *Management Communication Quarterly,* vol. 10, no. 1, pp. 50-72.

Putnam, Linda L. and Anne M. Nicotera (eds.) (2009): Building Theories of Organization: The Constitutive Role of Communication. New York: Routledge.

Rice, R. E., & Bair, J. H. (1984). New organizational media and productivity. In R. E. Rice & Associates (eds)., The new media: Communication, research and technology (pp. 198-215). Beverly Hills, CA: sage.

Rogers, E. M. (2003). Diffusion of innovations (5 ed.). New York: Free Press.

Sætre, Alf S., Jan O. Sørnes, Larry D. Browning and Keri K. Stephens (2007): Enacting Media Use in Organizations. Journal of Information, Information Technologies and Organization, vol. 2, pp. 133-158.

Scott, C. R. (1999). Communication technology and group communication. In L. R. Frey (ed.), D. S. Gouran, & M. S. Poole (assoc. eds.), The handbook of group communication theory and research (pp. 432-472). Thousand Oaks, CA: Sage.

Scott, C. R. (2003). New communication technologies and teams. In R. Y. Hirokawa, R.S. Cathcart, L. A. Samovar, & L. D. Henman (Eds), *Small group communication theory and practice: An anthology* (8th ed., pp. 134-147).

Stewart, C. (2007, March). Indiana University cyberinfrastructure newsletter. Retrieved November 1, 2008 from http://racinfo.indiana.edu/newsletter/archives/2007-03.shtml.