Disciplinary Histories, Funders' Priorities, and Competing Research Approaches: Organizational Communication Forces that Influence the Emerging Design of Cyberinfrastructure in the Early 21st Century US

Kerk F. Kee, Ph.D.
Assistant Professor Department of Communication Studies
Chapman University, Orange, California 92866
Doti Hall, Room 205
949-391-9988 (Google Voice)
714-532-6036 (Office)
714-628-7237 (Fax)
kerk.kee@gmail.com
www.ekerk.com

Abstract

Over the last decade or so, the emerging phenomenon of cyberinfrastructure (CI) and e-science has received much attention, enthusiasm, and funding support in the scientific community in the US and the world. While the vision is compelling, CI is still in its early development stages. This study investigates the organizational communication forces (i.e., disciplinary histories, funders' priorities, and competing research approaches) that drive early CI designs, and the ways in which these forces manifest between institutions and agents involved in CI projects. Theoretically, the paper concludes with a call to pursue studying infrastructures in organizational and applied communication research. Practically, the paper provides specific recommendations for how stakeholder groups (i.e., professional associations in various disciplines, universities that host supercomputing centers, etc.) can actively shape the trajectories of future CI designs to promote a healthy CI ecosystem.

Key Words: Cyberinfrastructure, e-Science, and Design.

Kee, K. F. (2013, November). *Disciplinary histories, funders' priorities, and competing research methodologies: Organizational communication influences behind the emerging designs of cyberinfrastructure*. Paper presented at the National Communication Association annual conference, Washington, DC.

Disciplinary Histories, Funders' Priorities, and Competing Research Approaches: Organizational Communication Forces that Influence the Emerging Design of Cyberinfrastructure in the Early 21st Century US

Over the last decade or so, there has been much enthusiasm in the United States, the European Union, and elsewhere in the world for the emerging phenomenon of cyberinfrastructure (CI) and the e-science it supports (Barjak, Eccles, Meyer, Robinson, & Schroeder, 2013; Edwards, Jackson, Bowker, & Williams, 2009; Jankowski, 2007). CI is a powerful infrastructure consisting of a dispersed network of communication technologies, computational and visualization tools, big data, remote instruments, interdisciplinary experts, supercomputers, etc. that enables virtual environments and virtual organizations to support large-scale, data-intensive, and distributed e-science (Atkins et al., 2003; Fry & Schroeder, 2009; Lee, Dourish, & Mark, 2006; Ribes & Lee, 2010). This approach to science has been hailed as "one of the most successful modern methods for experimental scientific discovery" (Getov, 2008, p. 30). However, the emerging CI model can also impact the evolution of future science and its trajectory given the emergence of dominant CI designs. The purpose of this study is to explore and identify the organizational communication forces at the macro-level that drive the emergence of the dominant CI designs in the early 21st century US.

Although the term 'cyberinfrastructure' was coined and made prominent in the highly influential Blue Ribbon Advisory Panel Report authored by Atkins and colleagues in 2003, it was not until 2006 when the National Science Foundation (NSF) officially established the Office of Cyberinfrastructure to provide a coordinated effort to support the grassroots CI phenomenon in the US (Seidel, Muñoz, Meacham, & Whitson, 2009). According to the 2013 NSF budget request document to Congress (NSF, 2012), OCI budget for research from 2005 to 2013 shows an upward trajectory, except for a drop in 2010 following a huge spike in 2009. The 2013 budget request from Congress for OCI is $218,270,000, a proposed increase of 3.1% from 2012.

Due to its collaborative nature, CI projects have also been referred to as 'collaboratories' (Finholt, 2003; Olson, Zimmerman, & Bos, 2008). In a review of more than 200 projects in the US, Bos and colleagues (2007) identified and described seven collaboratory types: distributed research centers, shared instruments, community data systems, open community contribution systems, virtual communities of practice, virtual learning communities, and community infrastructure projects. While these seven types of collaboratories have unique characteristics, the present study takes a macro perspective and treats these as parts of a whole, the larger phenomenon of cyberinfrastructure.

Infrastructure scholars have written about the issue of CI design. For example, Edwards and colleagues (2009) observe the criteria of stability and sustainability are key CI design factors for domain scientists who use CI for their science. On the other hand, significant innovation and research contribution through risky systems are what drive computer and computational scientists in their motivation for designing CI software and hardware. Given its distributed nature, Karasti and colleagues (2002) argue that the temporal and spatial dimensions present important challenges to CI design. In an NSF report of a workshop on lessons for new scientific cyberinfrastructure, Jackson and colleagues (2007) asserted, "Speaking of cyberinfrastructure as a machine to be built or a technical system to be designed tends to downplay the importance of social, institutional, organizational, legal, cultural, and other non-technical problems developers always face" (p. 7). Therefore, the present study focuses on the organizational communication forces at the macro-level that drive the technical aspects of CI design.

No infrastructure can emerge without an initial design, so it is inevitable that a design is directly and indirectly chosen in early projects that form the foundation of a national cyberinfrastructure. However, Steve Jackson and colleagues (2007) maintain that a key challenge of CI design is the realization of its path dependent nature, “what you do now affects what you can build in the future” (p. 36). Leigh Star and Geoff Bowker's (2006) argue to theorize infrastructure as an “installed base” (p. 231), and to notice how a new technology “wrestles with the inertia of the installed base and inherits strengths and limitations from the base” (p. 231). The implication of an ‘installed base’ is aligned with the argument advanced by communication scholars Mark Aakhus and Sally Jackson (2005) in that the design of a technology creates or amplifies preferences for users’ actions through its affordances, and design also cuts off possibilities through its inherent constraints.

This trajectory poses important concerns to the scientific community at large and the universities involved in the CI phenomenon with their campus-based supercomputing centers, because the establishment of an installed base and the emergence of a dominant design threatens the basis for diversity of research problems in basic science. As Aakhus and Laureij (2012) argue that design choices “reveal how important aspect of organizational and professional life are organized around the possibilities for orchestrating interactivity to facilitate some forms of communication while inhibiting other forms of communication” (p. 42). Similarly, this argument applies to the case of CI. Sally Jackson (2010) raises the need to cultivate a healthy ecosystem of cyberinfrastructure. Therefore, the present study takes an interest in exploring the communication forces that can create a dominant CI design that threatens the cultivation of a healthy CI ecosystem.

The literature reviewed suggests that the issue of CI design is both a necessary dimension for CI emergence as well as a limiting dimension for a diverse CI ecosystem in the future. In other words, early CI tools cannot be developed without some design choices. However, these choices also lead CI design down a path that will promote future designs that fit the base, potentially leading to the emergence of a dominant design. Collectively, the literature points to the need to more systematically investigate the macro-level organizational communication forces that drive early CI design.

Drawing from Aakhus and Jackson's (2005) work on technology, interaction, and design, I applied a design stance to study early CI design and the organizational communication forces at the macro-level that drive its trajectory. Aakhus and Jackson argue that a given design of a technology implies how the activities supported by the technology should work. Furthermore, design is value-laden. Therefore, users, developers, scholars of CI need to realize what is considered a normal or ideal design implies particular values embedded in such an assumption. An important argument they put forth is that design is a designer's hypothesis about the way in which could things work, not the way things do or should be. Treating design as hypothetical not only allows us to notice the assumptions behind the hypothesis, but the consequences of a particular design in adoption and implementation (Aakhus, 2007). Finally, taking a design stance also means to recognize the practical knowledge demonstrated by designs that are successful and no successful during implementation.

By taking a design stance in this study, the notion of CI design needs to be further defined. First, Aakhus (2007) argues, "[communication] designs are distinguishable by considering what an institution, practice, procedure, or technology presupposes about communication" (p. 114). This argument can be applied to studying CI design when CI is treated as a computer-mediated communication phenomenon based on a dispersed network of tools, datasets, instruments, experts, supercomputers, etc. Therefore, CI designs can be investigated by studying what the institutions, practices, procedures, and technologies involved in CI projects presuppose about CI.

Second, Aakhus and Jackson's (2005) argue that "a design enterprise involves creating techniques, procedures, and devices that make forms of communicative activity possible that were once impossible or that realize an improved form of communicative practice" (p. 416). Therefore, this study investigates CI design by looking for influences and choices that drive the creation of techniques, procedures, and devices that make previously impossible e-science activity possible and/or improved them.

Third, Aakhus and Laureij (2012) define an organization as 'a collaborative course of action" (p. 43). Therefore, this study investigates CI designs by looking for design actions, activities, and choices that represent forms of collaborations between institutions (such as disciplines, funders, etc.) and agents (scientists, technologists, etc).

Therefore, the three argues about designs inform this study, which was guided by two interrelated research questions: What organizational communication forces (surrounding institutions, practices, procedures, and technologies) at the macro-level drive early CI design? How do these organizational communication forces manifest (between institutions and agents) in terms of design choices in early CI projects?

Methods

A total of 70 interviews with 66 participants from across 17 U.S. states and three other countries were conducted over a period of 32 months, from November 2007 to June 2010. The interviews were spread across four years with 10 participants in 2007, 42 in 2008, 16 in 2009, and two in 2010. Because most of the interviews were conducted in 2008, the analysis primarily reflects CI during this period. Recruitment of interview participants mainly relied on snowball sampling technique (Biernacki & Waldorf, 1981).

The 66 interview participants came from Texas (12), Illinois (11), California (10), Michigan (5), Indiana (4), Massachusetts (3), Arizona (2), Colorado (2), Louisiana (2), Washington (2), DC (1), Maryland (1), New York (1), Virginia (1), Ohio (1), Pennsylvania (1), Delaware (1), as well as Australia (2), Germany (1), and the UK (1). The geographic affiliations refer to the primary locations of the participants at the time of the interviews. Participants include 52 males and 14 females. Participants' primary professional roles were diverse, including pioneering domain scientists who used CI to conduct science (15), computational technologists who built CI (12), a range of administrative directors and program managers at supercomputer centers and national research laboratories across the country (21), NSF program officers who helped allocate funding to CI projects (4), social scientists and policy analysts who studied and participated in CI projects (12), and experts from commercial industry (2).

The shortest interview was 15 minutes and the longest was 2 hours and 16 minutes, averaging approximately one hour each and were conducted in person with 19 of the participants and over the phone with the remaining 51 participants. All the interviews were audio recorded except for two, due to technical difficulty and following one participant's request. However, notes were taken immediately after these two interviews. All interviews were transcribed verbatim, yielding more than 600 pages of single-spaced text for analysis.

A qualitative approach (Tracy, 2010) guided by grounded theory (Corbin & Strauss, 1990) and Owen's (1984) criteria of recurrence, repetition, and forcefulness were employed in the present study. Data analysis was conducted in NVivo 10 qualitative software with multiple iterations. First, selective coding was performed by going through the transcripts, flagging instances related to the focus of this paper, which is CI design and related issues, such as design choices, design model, project decisions, etc. Then data analysis employed open coding and axial coding to allow key themes to emerge, based on Owne's criteria. As a result, three key themes of disciplinary cultures, funders' priorities, and competing research approaches emerged. In order to ensure the maximum capture of excerpts related to these three key themes, text queries with key words related to the three key themes was performed in NVivo 10. The resulted text was reanalyzed using axial coding and constant comparisons. A general framework with two subthemes under each of the three main themes is reported and elaborated with selected excerpt in the next section.

Findings

Disciplines Communicate Historical Orientations Through Routines and Norms

The first source of organizational communication force on CI design is disciplinary histories, especially the histories related to the Internet and desktop computers. More specifically, the design of CI tools developed out of early projects tend to reflect different disciplinary data practices (e.g., assumptions, techniques, formats, online sharing, etc.) and software architectures (desktop versus supercomputing scales). Using big data in aggregation is a key CI characteristic. Therefore, data practices of pioneering scientists present an important source of design influence.

Raw scientific data is a rich space to explore design, as it reflects the disciplinary perspectives of the scientists who collected the data. An informant emphasized that it is important to pay attention to "data techniques, things about the data, methods that are different between the disciplines…" (Social Scientist, Arizona, 19 February 2009). Therefore, this section explores the data practices of scientists in the context of early CI projects.

Data Practices. Scientists from different disciplines have different perspectives and assumptions that guide the way they approach science. These differences can be seen in problem definition, data collection, and data interpretations. An informant explained, "A hydrologist versus and [an] environmental engineering [researcher], they may be looking at the same phenomena but have completely different ways of characterizing it and having different ways of describing the data they've collected" (Social Scientist, Michigan, 2 February 2008). Another informant noted, "Oftentimes they have different formats, they have different assumptions that went behind how they interpreted their data" (Computational Chemist, Massachusetts, 22 February 2008). Therefore, data practices of pioneering scientists based on their disciplinary orientations drove early CI design at the data level. As CI development continues based on data accumulated over time on scientific problems defined by pioneering scientists, an informant alerted, "You need to make sure that you… understand the discipline specific aspects, while at the same time, understand the communities that want to use the data across disciplines" (Technologist, Colorado, 21 April 2008). In other words, if reusability of datasets beyond the original projects is desirable, CI design needs to take into consideration how scientists in other disciplines define, collect, and interpret the data on the same scientific problem. However, in practice, this may be an impossible task because most CI projects are highly specialized investigations based by pioneering scientists. From a CI design standpoint at the data level, future scientists may need to follow the disciplinary perspectives of pioneering CI scientists, and future datasets may need to conform to the data formats chosen by pioneering scientists so future datasets are interoperable with existing datasets.

Disciplines differ in terms of their existing database systems, which give some scientists compatible advantages with the CI model. CI employs supercomputing resources to analyze, visualize, and simulate big data from aggregation, so disciplines with database systems compatible with computational tools and the way they are designed to operate would have an advantage to thrive in the CI ecosystem. In fact, some fields that have engaged in CI projects longer than others have built in their disciplinary cultures and data practices the compatible formats. An informant shared his observation of the scientists involved in a successful CI project called GEON (GEOsciences Network),

> Within GEON, the range of experience was so varied, everything from paleobotanists who worked with very small-scaled data, often on Excel spreadsheets is how they kept their data, or maybe on Microsoft Access. So various sort of low-level, what I consider today, non-scalable, non-interoperable database systems. So we've got that range on the one side and then you've got quite a few different kind of geologists… The geophysicists in particular, or the seismologist, have long experiences, decades of experience, of working with the distributed computing, supercomputing with database models (Social Scientist, Michigan, 11 March 2008).

Certain disciplines, such as geophysics and seismology, already have database models that are compatible with the CI model involving distributed computing techniques and supercomputing resources. The design compatibility of database models will likely continue giving geophysicists and seismologists advantages, and build their disciplinary cultures into the CI database model, and vice versa.

The aggregation of datasets into big data requires the sharing of small datasets. However, the practice of data sharing is also disciplinarily rooted. Data sharing practices reflect different disciplinary histories. A recurring observation is that biology, as represented by the branch of bioinformatics, has a different historical blueprint that sets their data sharing practices a part. For example, an informant reflected:

> [E]verything in biology is pretty open… 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…Whereas biology, bioinformatics, was born 10 years ago. So it grew up as the Internet was growing up. So biology almost started doing cyberinfrastructure [in terms of data sharing] without thinking…" (Informatics Researcher, Physicist, and Computational Technologists, Indiana, 14 February

> 2008).

Disciplinary cultures are not born out of vacuum, but a complex ecology of key events, leaders, institutions, etc. The birth of the Internet is one of these key events. Due to the parallel developments of bioinformatics and the Internet, bioinformatics researchers started sharing data online almost naturally. When a mindset is natural, behaviors and actions become automatic.

Computing Architectures. Disciplinary cultures as design influence are also reflected in at the level of computing architectures (desktop versus supercomputing scales). As the previous informant suggested, the concurrent historical developments of bioinformatics and the Internet, bioinformatics researchers are comfortable with the open data sharing vision of CI. In the meantime, their disciplinary perspective tied to historical use of desktop computing limits some bioinformatics researchers in terms of their capacity of large-scale data computations and simulations. An informant observed,

> Physicists tend to be more computational sophisticated than biologists at this point. The biologists have come to computing lately with the deluge of digital information coming out of sequencers – probably the last 10 years primarily. Bioinformatics has exploded. Lots of good ideas, but they kind of think at the desktop scale. So they've incorporated a lot of desktop computation into their sort of normal workflows but aren't running leading scale, supercomputing systems yet… They don't have the common framework of FORTRAN and differential equations which are the basics for physics and engineering style simulations. It's PERL scripts and desktop codes and more data analysis than simulation… The way that they do science is not the way that the… physicists, electrical engineers, mechanical engineers, who've been in the [supercomputing] game for a while and were trained in a different era, do science. (Center Administrator, Texas, 28 February 2010).

The limitation of programming at the desktop scale is that when the dataset gets too big, the processing of the data slows down. In response to slow processing with desktop scale programming, some scientists think they run out of memory or they cannot solve a problem based on big data. Instead of turning to supercomputers by operating at the supercomputing scale, some change their problem so it can be solved at the desktop scale. These outcomes reflect the perspectives and assumptions stemmed from disciplines cultures. As an informant put, "A lot of these ideas [related to CI] are running contrary to the way people were brought up [in their home disciplines]" (Physicist and Computational Technologist, Louisiana, 18 February 2008).

The two pairs of contrast between chemistry and biology as well as biology and physics show the dialectical complexity of CI design, because on one hand, bioinformatics researchers are comfortable with open data sharing because of their disciplinary history parallel to the Internet. On the other hand, their disciplinary history also limits their experience with a supercomputing scale programming architecture that is more aligned with CI. This dialectics shows that CI is a highly complex phenomenon, one that may not fully manifest in a given discipline and/or project. Therefore, the macro CI design is constructed based on a string of loosely connected prototypes and exemplars in a direct and indirect fashion.

Moreover, CI elements such as programming codes, datasets, software applications, etc. are all interrelated as a cluster in practice. One has no value if other elements are missing. Therefore, to practice CI is to use these elements in combination. CI is moving towards an ecosystem based on elements developed out of different projects in a bottom-up fashion. If a CI element turns out successful, it will provide values and be adopted by others in the scientific community. An informant shared about CI software and related databases for adoption,

> … [T]hat software has to be so good that people on different continents and different places with different perspectives, are all going to want to use this software and use this database… [If] the software is kind of idiosyncratic, then it's hard to know that it will be adopted widely. If it gets adopted, then presumably people will approve it and add to it (Computational Chemist, Massachusetts, 22 February 2008).

A CI element, such as a piece of software, may be adopted beyond the original project and be used by scientists across the US and the world. With its worldwide adoption, the disciplinary perspective and assumptions embedded in the design of the software will also be adopted by other scientists. When new users add to it, they further solidify the foundational design along with its disciplinary perspective and assumptions that guide scientific practices. An informant profoundly stated, "… it is not just a matter of building the infrastructure, the technological infrastructure". She continued, "but it is also a matter of changing the way people collect data, for example, and changing the hardware that people are using or agreeing on some common sets of things to use. So it is also about changing practices" (Social Scientist, California, 5 March 2008).

Funders Communicate Priorities Through Funding Programs

Funding serves as a deterministic influence on CI design with explicit priorities promoted by funding programs for scientists, technologists, and supercomputing center administrators at research universities. This deterministic theme is forcefully articulated in these quotes: "[W]e live in a world where we either get funding and we do publications or we don't get where we're going. You don't get tenure or I don't get to stay on in my soft money position" (Administrator, California, 1 February 2008) and "So funding is often times dictates a lot of what goes on in science and in research and institutions" (Administrator, Texas, 24 April 2008).

To receive funding from an agency means to accept the social and political priorities implied in the funding solicitation, because every solicitation promotes a specific research agenda that supports certain priorities over others. An informant aptly put, "Funding is very goal oriented" (Informatics Researcher, Physicist, Technologist, Indiana, 14 February 2008). Therefore, the design of a CI tool is not value-neutral. The priorities communicated via funding are built into the design of CI tools developed by funded scientists and/or technologists.

Funding for Scientists and their Scientific Projects. Most CI projects are funded on a three to five years cycle. However, given its emerging nature, every piece of CI is an endeavor that requires multiple cycles of funding in order to reach stability and maturity. An informant explained, “Even after five years, it may just take that amount of time to get the basic infrastructure up and running.” (Social Scientist, California, 5 March 2008). In order to sustain the CI tools being developed in the original projects, the pioneering scientists as the principal investigators (PIs) have to think longer term. The long-tern sustainability concern can lead to two possible scenarios: failed projects without continuing funding and forcing tools from the original project into a new project. A project can fail at the end of a funding cycle when it is dependent on a single agency. An informant shared:

> The problems are really magnified by the fact that faculty receives grants from NIH [National Institutes of Health] and other places, which did not encourage collaboration and joint usage of technology, but rather, waived them off and said – If you go out and get a Sun workstation on your desk, put a couple of Condors together, and then when the funding runs out, [you are] left with this big bill to run these machines (Policy Analyst, DC, 8 February 2008).

Funding agencies often have their individual agenda and priorities. Majority of their programs are not motivated to fund something that will promote a different set of priorities pursued by another agency. Furthermore, priorities of funding programs within agencies change and evolve in response to the political environment they are embedded in. In the US, the National Science Foundation, the National Institutes of Health, Department of Energy, Department of Defense, and other federal funding agencies receive their own funding from Congress. They need to promote different sets of priorities in order to set themselves apart to receive funding from Congress. If their priorities hugely overlap with another agency, they risk not being able to justify their budgets, thus their very existence.

In order to avoid letting a project fail, PIs attempt to push their initial CI tool into another project, even if it is not the perfect fit. An emerging CI tool requires continuing funding to be further developed by the pioneering scientists before it becomes adopted and sustained by a community of users. An informant noted,

> The project [funding] will end and there will be nothing there in place to carry successful cyberinfrastructure forward. That has to be carried forward either by other projects, follow on projects, or

> if lucky, it's developed a community that often – the project's not funded to develop the community. They're only developed – if the project is funded – to develop a piece of cyberinfrastructure… [T]he group will get very attached to a particular technology that's developed and push it into places where it was never intended…because they need funding to continue their work (Technologist, Colorado, 21 April 2008).

An emerging CI tool becomes more complex when the priorities built into the first layer of design is covered up by the set of priorities from continuing funding, but both layers co-exist in the CI tools. The design becomes more complicated over time as more layers are added, if the PIs are able to secure continuing cycles of funding.

Funding for Technologists, Administrators, and their Supercomputing Centers. Although the discussion thus far appears to highlight how factors associated with scientists are critical design influences because CI tools are driven by scientific problems, it is important to note that many of these tools are designed by technologists at supercomputing centers, managed by administrators. Furthermore, in the case of scientists developing their own tools, they need to run their analyses using computing resources provided by these centers. Therefore, funding does not only influence domain scientists, but also technologists at and administrators of supercomputing centers at various research universities. Despite being housed within research universities, these centers heavily depend on external funding. An informant clarified, "We get more federal funding than we get from our campus… [We] maintain a very strong campus approach even with limited campus dollars, and a very strong national approach because that's where a lot of the funding is." (Administrator, Texas, 2 January 2008). In order to increase the chances of bringing in continuing federal funding, supercomputing centers have to be strategic and selective of the CI projects they pursue. One of the main considerations is the potential of a project that will bring in additional funding subsequently. An informant revealed:

> People won't admit to this out loud… [but] one of the questions that always came up at [a major supercomputing center] and it came up at [another supercomputing resource provider] was – Will this project bring in additional funding? Will this be a project that will look good to funding agencies? And it's true. It's about survival. If you're not bringing in grants, if you're not bringing in money, if the researchers aren't sharing grant money and seed money with you, you're going to flounder. You need

> those resources to survive. It's just that simple. Campuses may fund the supercomputing center for the startup, but they really want to see you get off the university's buck (Technologist, Illinois, 18 November 2008).

Under the condition that scientists, technologists, and supercomputing centers administrators critically depend on external funding, the design of sustaining CI tools is a complex repository of priorities embedded in the funding received by domain scientists and supercomputing centers in their individual and joint projects. This makes a study of CI as a technology to support scientific work fundamentally different from most other studies of workplace technologies in applied and organizational communication. An informant stated, "[S]upporting cyberinfrastructure is not the same as supporting an information system in a traditional organization" (Social Scientist, California, 5 March 2008).

Research Approaches Communicate Methodologies Through Competitions

A disciplines is defined by the thriving theories, methodologies, and approaches pursued by the scientists in the field. Similar to many other fields of work, academia is highly competitive and the leaders in the CI movement have to be savvy at maneuvering the competitive landscape of diverse theories, methodologies, and approaches.

Seeking Funding as a Form of Competition. To build on the second theme of funding previously discussed, the act of seeking funding is a form of competition. Furthermore, the funding proposal review process is a form of competition as well, as reviewers may favor or criticize a proposal based on the reviewer's own research approaches. The competition among PIs and the desire to get one's approach built into the foundation of CI to support future e-science can be detected as early as during the proposal review stage. An informant shared,

> [I]n writing for NSF proposals to support this work, we had some reviews back where people would say – How can the scientists who are using this promise to make their software available for everyone else because then they have control and then they could sabotage the rest of the community… Particularly if

> that group that is developing the software is also doing their own science with it (Physicist and Computational Technologist, Louisiana, 18 February 2008)

The peer-reviewed process at many major funding agencies is not blind and confidential. In other words, reviewers know the identities of the PIs of the proposals they review. In the event that there are competitions between the theories, methods, or approaches pursued by the PIs and reviewers, the tensions can play out in the review process. However, many respectful reviewers simply raise the legitimate concerns and point out the ethical implications of funding a project that can give certain groups of scientists control over the work other scientists in the field.

Competition During CI Development. In addition to the competition that manifests during the proposal stage, such competition can also play out during the CI development stage. First, competitive scientists can be forceful and straight-forward about pushing their approaches to be built into CI design from the start. An informant shared,

> [W]e'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 embed my methodology in the way the software works and exclude my competitor over there. We had to try explain to them that cyberinfrastructure could be used to let them have sort of a – have a fair fight on the science side that we get the data together, run both of their methodologies on it, let them compare in detail the results of their methodologies coming out the other side… I think that sort of pushed the competition into the realm of can you get money to get cyberinfrastructure around your methodology, which we sort of backed out of by trying to make the cyberinfrastructure a neutral relative to their two methodologies (Administrator, Illinois, 23 January 2008).

While creating a neutral ground for multiple approaches to equally compete is a good solution, not every CI project has the resources that allow such a solution. As discussed in the second theme, funding is critical in driving CI projects and their designs, the limitation of funding mixed in the inherent competition sometimes lead to difficult design choices. In such a situation, prominence and credibility of the project participants in competition as well as the funders can impact the trajectory of the design and development of a CI tool. An informant recounted the following situation:

> [T]here are people who are building tools who have very different views of what's the successful way of building a tool to manage and to address the various technical needs that are coming up. In particular, there

> were 2 systems that were being developed in parallel that were essentially workflow management systems. One of them was web-based and one of them was a downloadable – what they call client-based – so it runs on your desktop instead of running through your browser… But because of the nature of the personalities involved, and… the nature of who had more credibility, who was better friends with who – what ended up happening was at one point, I think it was a combination of NSF essentially intervening and saying – You can't afford to be spending your money developing two potential pathways for developing the software and you need to end one of them. They chose instead the web-based one" (Social Scientist, Michigan, 17 March 2007).

When resources are limited, early CI projects faced the constraints of not being able to let multiple choices develop concurrently. When one is chosen over another, long-term affordances and constraints begin to accumulate even at small scale.

Qualitative analysis of interview data based using grounded theory procedures and Owen's criteria yielded three main themes and two subthemes for each as discussed above. Table 1 below provides a summary of these themes.

Table 1. Organizational communication forces at the macro-level that drive early cyberinfrastructure design and their manifestations in influencing design choices in cyberinfrastructure projects.

1.	Disciplines Communicate Historical Orientations Through Routines, Procedures, and Norms
	a. Data Practices
	b. Computing Architectures
2.	Funders Communicate Priorities Through Funding Programs
	a. Funding for Scientists and their Scientific Projects
	b. Funding for Technologists, Administrators, and their Supercomputing Centers
3.	Research Approaches Communicate Methodologies Through Competitions
	a. Seeking Funding and Proposal Review as a Form of Competition
	b. Competition During Cyberinfrastructure Development

Conclusion

This study began with the research questions, "What organizational communication forces (surrounding institutions, practices, procedures, and technologies) at the macro-level drive early CI design? How do these organizational communication forces manifest (between institutions and agents) in terms of design choices in early CI projects?" A systematic analysis of qualitative interviews with domain scientists, computational technologists, center administrators, and social scientists/policy experts involved in CI project revealed three main sources of influence: disciplinary histories, funders' priorities, and competitive research approaches. First, disciplinary histories in the forms of data practices and computing architectures influence the design of projects that give scientists advantage on setting data standards and compatibility with early CI model. Second, funders' priorities drive the choices made by scientists, technologists, and center administrators in the pursuit of various projects in order to sustain the emerging tools being developed and the survival of supercomputing centers. Finally, the competition among rivalry methodologies and research approaches play out in the proposal review and tool development stages, as competitors attempt to push the trajectory of CI design given the limitation of funding resources. This study suggests several implications, both at the theoretical and practical levels.

Theoretical Implications

First, findings of this study suggest the increasing importance of studying infrastructure instead of discrete technologies in applied communication research. While it is sometimes necessary to focus the scope of a study by solely looking at a discrete technology in a specific communication context, findings from this study suggest that multiple technologies within an infrastructure or platform interrelate in a complex way, and the interrelatedness gives deeper meaning for each. For example, in the case of CI examined in the present study, the dataset is meaningful only when processed by a software application, powered by a network of supercomputers.

Second,

Practical Implications

Given that early CI design can be the installed base of the emerging CI for future science to come, Baker and colleagues (2002) argue to design infrastructures for heterogeneity. Aakhus and Jackson (2005) argued, “A strong theoretical response include not only noticing these affordances and constraints but also actively participating in shaping their directions” (p. 412). Here I suggest two specific ways in which stakeholder groups can actively participate in shaping future CI designs for heterogeneity in order to promote a healthy CI ecosystem.

First, in order to avoid or simply reduce the rise of dominant disciplines in the CI ecosystem, stakeholder groups could actively reach out to disciplines that are historically not aligned with the CI model at this point. Although the enthusiasm for CI is most prominently felt in the disciplines historically more aligned with the current CI model, the investment in less dominant or currently disconnected disciplines will likely promote a healthy CI ecosystem in the long run.

Second, early and current CI projects are mostly carried out by either domain scientists (and their graduate students and post-docs) who know enough about computational sciences so they are about create their own CI tools or by joint projects between domain scientists and computational technologists in funded projects. In a study of the science court, Aakhus (1999) explains that the selection of judges should not include researchers working in the dispute area, with organizational affiliations to the area, or people with a clear bias toward one side of the dispute. By applying this argument to the case of CI, funders could explore creating new funding mechanisms to initiate and support technologists whose goal is to build CI tools that could serve as many disciplines and methodologies as possible.

Finally, as CI is an emerging approach to research, the research communities at large and universities across the US and the world could be more involved in guiding its development. CI as an enterprise is too costly and too time consuming for re-designing. Therefore, thoughtful considerations should be given at the early stages. Major associations within different fields and administrators of universities can ask three critical questions to help them reflect on their own relationships with the emerging CI movement: “What is our field’s history with the Internet, computing architecture, and the way we collect and share digital data with the scientific community?”, “Who is funding our CI projects, campus supercomputing centers, and which agenda are the projects advancing?”, and “For which theory, method, and research approach in our field is a piece of CI built?” Among many other possible questions, these three that stemmed from the findings of this study will help leaders of different disciplines across the sciences, social sciences, and humanities as well as universities across the world negotiate their relationships with emerging CI.

The macro structures layer includes social networks, teams, organizations, institutions, communities, fields, disciplines, and other macro entities and networks that tie people together because of common characteristics, goals, purposes, and relationships. These macro structures usually maintain existing practices and cultures among the people involved, beyond the confine of time and space. This is similar to the notion of a ‘virtual organization’ in cyberinfrastructure literature (Bird, Jones, & Kee, 2009; Foster, Kesselman, & Tuecke, 2001), ‘invisible college’ in traditional diffusion literature (Estabrooks et al., 2008; Gmür, 2003; Lievrouw, 1989; Rogers, 2003), and Structuration Theory’s notions of rules and resources (DeSanctis & Poole, 1994; Giddens, 1984). The macro structures provide the environmental context in which cyberinfrastructure adoption is analyzed and understood.

Refererences

Aakhus, M. (1999). Science court: A case study in designing discourse to manage policy controversy. *12*, 20-37.

Aakhus, M. (2007). Communication as design. *Communication Monographs, 74*(1), 112-117. doi: 10.1080/03637750701196383

Aakhus, M., & Jackson, S. (2005). Technology, interaction, and design. In K. Fitch & R. Sanders (Eds.), *Handbook of language and social interaction* (pp. 411-437). Mahwah, NJ: Lawrence Erlbaum.

Aakhus, M., & Laureij, L. V. (2012). Activity, materiality, and creative struggle in the communicative constitution of organizing: Two cases of communication design practice. *Language & Dialogue, 2*(1), 41-59.

Atkins, D. E., Droegemeier, K. K., Feldman, S. I., Garcia-Molina, H., Klein, M. L., Messerschmitt, D. G., . . . Wright, M. (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/.

Baker, K. S., Bowker, G. C., & Karasti, H. (2002). *Designing an infrastructure for heterogeneity in ecosystem data, collaborators and organizations.* Paper presented at the Proceedings of the 2002 annual national conference on Digital government research.

Barjak, F., Eccles, K., Meyer, E. T., Robinson, S., & Schroeder, R. (2013). The emerging governance of e-infrastructure. *Journal of Computer-Mediated Communication, 18*(2), 1-24. doi: 10.1111/jcc4.12000

Biernacki, P., & Waldorf, D. (1981). Snowball sampling: Problems and techniques of chain referral sampling. *Sociological Methods and Research, 10*, 141-163.

Bird, I., Jones, B., & Kee, K. F. (2009). The organization and management of grid infrastructures. *Computer, 42*(1), 36-46.

Bos, N., Zimmerman, A., Olson, J., Yew, J., Yerkie, J., Dahl, E., & Olson, G. (2007). From shared databases to communities of practice: A taxonomy of collaboratories. *Journal of Computer-Mediated Communication, 12*(2), 652-672. doi: 10.1111/j.1083-6101.2007.00343.x

Corbin, J., & Strauss, A. (1990). Grounded theory research: Procedures, canons, and evaluative criteria. *Qualitative Sociology, 13*(1), 3-21.

DeSanctis, G., & Poole, M. S. (1994). Capturing the complexity in advanced technology use: Adaptive structuration theory. *Organization Science, 5*(2), 121-147.

Edwards, P. N., Jackson, S. J., Bowker, G. C., & Williams, R. (2009). Introduction: An agenda for infrastructure studies. *Journal of the Association for Information Systems, 10*(5), 364-374.

Estabrooks, C., Derksen, L., Winther, C., Lavis, J., Scott, S., Wallin, L., & Profetto-McGrath, J. (2008). The intellectual structure and substance of the knowledge utilization field: A longitudinal author co-citation analysis, 1945 to 2004. *Implementation Science, 3*(1), 49.

Finholt, T. A. (2003). Collaboratories as a new form of scientific organization. *Economics of Innovation and New Technology, 23*(1).

Foster, I., Kesselman, C., & Tuecke, S. (2001). The anatomy of the grid: Enabling scalable virtual organizations. *International Journal of High Performance Computing Applications, 15*(3), 200-222. doi: 10.1177/109434200101500302

Fry, J., & Schroeder, R. (2009). Towards a sociology of e-research: Shaping practice and advancing knowledge. In N. W. Jankowski (Ed.), *e-Research: Transformation in scholarly practice* (pp. 35-53). New York: Routledge.

Getov, V. (2008). e-Science: The added value for modern discovery. *Computer, 41*(11), 30-31.

Giddens, A. (1984). *The constitution of society: Outline of the theory of structuration*. Berkeley, CA: University of California Press.

Gmür, M. (2003). Co-citation analysis and the search for invisible colleges: A methodological evaluation. *Scientometrics, 57*, 27-57.

Jackson, S. (2010). Assessing cyberinfrastructure impact. *The Coalition for Networked Information Meeting.* *http://www.youtube.com/watch?v=_lLW4CrS7tc*.

Jackson, S. J., Edwards, P. N., Bowker, G. C., & Knobel, C. P. (2007). Understanding infrastructure: History, heuristics, and cyberinfrastructure Policy. *First Monday, 12*(6), Retrieve on June 12, 2008 from http://www.uic.edu/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/1904/1786.

Jankowski, N. W. (2007). Exploring e-science: An introduction. *Journal of Computer-Mediated Communication, 12*(2), 549-562.

Karasti, H., Baker, K. S., & Bowker, G. C. (2002). Ecological storytelling and collaborative scientific activities. *ACM SIGGROUP Bulletin, 23*(2), 29-30.

Lee, C. P., Dourish, P., & Mark, G. (2006). The human infrastructure of cyberinfrastructure. In P. Hinds & D. Martin (Eds.), *CSCW '06: Proceedings of the Conference on Computer Supported Cooperative Work, Banff, Alberta, Canada* (pp. 483-492). New York: ACM Press.

Lievrouw, L. A. (1989). The invisible college reconsidered: Bibliometrics and the development of scientific communication theory. *Communication Research, 16*, 615-628.

NSF. (2012). FY 2013 budget request to Congress. *http://www.nsf.gov/about/budget/fy2013/pdf/EntireDocument_fy2013.pdf*.

Olson, G. M., Zimmerman, A., & Bos, N. (Eds.). (2008). *Scientific collaboration on the Internet*. Cambridge: MIT Press.

Owen, W. F. (1984). Interpretive themes in relational communication. *Quarterly Journal of Speech, 70*(3), 274-287.

Ribes, D., & Lee, C. P. (2010). Sociotechnical studies of cyberinfrastructure and e-Research: Current Themes and Future Trajectories. *Computer Supported Cooperative Work (CSCW), 19*(3), 231-244.

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

Seidel, E., Muñoz, J., Meacham, S., & Whitson, C. A. (2009). A vision for cyberinfrastructure. *Computer, 42*(1), 40.

Star, S. L., & Bowker, G. C. (2006). How to infrastructure. In L. A. Lievrouw & S. M. Livingstone (Eds.), *Handbook of new media: Social shaping and social consequences of ICTs* (pp. 230-245). Thousand Oaks, CA: Sage Publications.

Tracy, S. J. (2010). Qualitative quality: Eight “big-tent” criteria for excellent qualitative research. *Qualitative Inquiry, 16*(10), 837-851. doi: 10.1177/1077800410383121