Managing CI Centers:

An Agenda for Organizational Scholarship and Cyberinfrastructure Innovation*

Nicholas Berente
University of Georgia
berente@uga.edu

Jennifer Claggett
University of Georgia
claggett@uga.edu

James Howison
University of Texas at Austin
jhowison@ischool.utexas.edu

Cory Knobel
University of California, Irvine
cknobel@uci.edu

Joe Rubleske
University of Georgia
rubleske@uga.edu

* A report from the NSF-sponsored "Managing CI Centers Workshop" in Athens, Georgia, on October 6-8, 2011. The workshop is part of an ongoing series of projects aimed at understanding and supporting the successful design and management of computationally-intensive collaborative research centers. We gratefully acknowledge the support of the U.S. National Science Foundation's Office of Cyberinfrastructure (OCI grant #1059153).[1]

Workshop Panelists: Stan Ahalt (RENCI); Jay Boisseau (TACC); Debbie Crawford (Drexel University); Bill Feiereisen (Intel); John Leslie King (University of Michigan); Laurie Kirsch (University of Pittsburgh); Patricia Kovatch (Mt. Sinai); David Moses (PSC); Nicole Radziwill (James Mason University); Barry Schneider (NSF); Sandra Slaughter (Georgia Tech); Phil Westmoreland (ICSE); Susan Winter (NSF).

---

[1] **Citation**: Berente, N., Claggett, J., Howison, J. Knobel, C. and Rubleske, J., (2012) "Managing CI Centers: An Agenda for Organizational Scholarship and Cyberinfrastructure Innovation." Available at SSRN: http://ssrn.com/abstract=2128872.

Managing CI Centers:

An Agenda for Organizational Scholarship and Cyberinfrastructure Innovation

Introduction

Advances in computational technology have transformed science. An increasing portion of scientific research is geographically distributed, cross-disciplinary, and reliant on a sociotechnical computational infrastructure that has come to be known as *cyberinfrastructure* (Atkins *et al*., 2003). These large-scale scientific collaborations involve a distinct and innovative form of organizing -- what we refer to as *cyberinfrastructure centers* (or "CI centers"). CI centers are those organizations charged with developing the digital infrastructure for the next generation of scientific activity.

CI centers are faced with diverse forms of complexity, and often the management of such centers involves dealing with multiple, unique challenges. CI centers are simultaneously charged with radical technological innovation and flexible service provisioning, while navigating deeply entrenched institutional (including university) arrangements. They are especially unique in that they manage long-term technology innovation trajectories in the context of short-term project funding (Edwards *et al*., 2007; Ribes and Finholt 2009; Karasti *et al*., 2010). Further, CI centers are increasingly virtual organizations, in that they are distributed geographically and are comprised of multiple, temporary or partially "drifting" arrangements (Lee *et al*., 2006; DeSanctis and Monge 1999). As such, these CI centers are in some ways similar to other organizations, but at the same time represent distinct forms of organizing. In many ways, CI centers may represent organizational forms of the future (Hedberg *et al.*, 2000). Existing literature on cyberinfrastructure does not, however, attend to the specific ways in which CI centers are unique, or how they might help us better understand other innovative organizational forms, or how they can be successfully managed.

In this report we make two key claims. First, we argue that CI centers are a unique and compelling form of post-industrial organization that can provide valuable insight to organizational scholars -- particularly in the increasingly important area of infrastructural innovation (cf., Tilson *et al*., 2010; Nicolini *et al*., 2012). Second, we propose that engaged organizational scholarship can provide relevant, actionable insight to CI center managers, funding agencies, and others who are involved with computationally-intensive science practice and policy -- a possibility that has been raised by Cummings and Kiesler (2011) and Lee *et al*. (2006).

In the following pages we support these claims with examples from our own research into CI center management (see Berente 2010; Herbsleb & Howison 2009). We presented these findings to a panel of CI center executives, senior organizational scholars, and cyberinfrastructure policymakers (see Appendix) in an NSF-sponsored "Managing CI Centers" workshop, and asked them to critically scrutinize and comment on these findings.

(1) CI Centers as Cyberinfrastructure "Stewards": To date, the bulk of studies into organizing around cyberinfrastructure focus on project teams (e.g., Ribes and Lee, 2010; Spencer *et al*., 2011). Our research extends this work by situating these project teams within their organizational contexts, and focusing on an essential role played by CI centers, namely, that of *cyberinfrastructure steward*. This form of stewardship entails combining a service ethos with a responsibility for sustaining elements of a digital infrastructure over the long term. This is accomplished through the coordination of knowledge resources (e.g., people, hardware, software, documents) across projects. As stewards of cyberinfrastructure, CI centers are positioned centrally within contemporary scientific and technological networks. Accordingly, we also propose that CI centers, through their position at a cyberinfrastructure network's "structural fold" (Vedres and Stark, 2010), effectively enable cyberinfrastructure innovation over time.

(2) CI Center Leaders as Entrepreneurs: Leadership in CI centers is usually associated with project management (Cummings and Keisler 2007; 2011; Karasti et al., 2010). Our research has found that, in practice, CI center leadership more accurately resembles entrepreneurship. CI center leaders come from a variety of backgrounds with diverse dispositions. We find that the attention of CI center leaders is a critical element of cyberinfrastructure evolution.

(3) Resource scarcity and Cyberinfrastructure Innovation: The single most important concern of CI center leaders involves access to the appropriate resources - both funding and human resources - to enable them to innovate. Although, in their view, more resources enable greater innovation - our findings indicate that this is not the whole picture. We find that the interplay of resource abundance and resource scarcity result in a sort of "generative tension" that fosters innovation.

(4) Scientific Software Ecosystems: Software is a critical element of cyberinfrastructure. Our research has examined the production of software in science and compared it with both commercial software production and production in the open-source software communities. Software development in CI centers, whether for internal or external use, largely takes place in relatively small, cohesive groups known as "software projects". Increasingly, though, CI centers are looking to participate in "software communities" either by attracting outside contributors or, less frequently, by contributing to existing software communities. Yet our work also shows a third role, as yet relatively unexplored in science: the software ecosystem steward, working at the broader community level to establish incentives, collaboration infrastructure, and governance for coordinated development.

(5) Assessing CI Center Impact: The impact of CI centers is often invisible but it is broad and far-reaching - resulting in everything from technology innovation (like key elements of the Internet) to groundbreaking science in multiple disciplines. However, because of the invisibility of infrastructure, assessing their contributions is one of the critical challenges for CI centers and policymakers. While difficult, fair assessment is imperative when navigating various institutional arrangements, acquiring resources, and comparing potential cyberinfrastructure investments that will shape future scientific patterns of action.

In addition to critically assessing the research findings themselves, the objective of the workshop was to identify directions for organizational scholarship to explore that would benefit CI center managers, researchers, and policymakers. In the remainder of this report, we briefly reprise each of these five findings and incorporate comments from the panelists. Then we report on an agenda for organizational scholarship and CI center management resulting from the discussions of the workshop.

CI Centers as Cyberinfrastructure "Stewards"

Much of the existing literature on organizational issues around cyberinfrastructure tends to emphasize project teams and the management of these teams (Lee *et al*., 2006; Spencer *et al*., 2006; Spencer *et al*., 2011). This focus makes sense because the related work is organized around and funded through discrete projects undertaken and delivered by project teams. However, cyberinfrastructure development transcends any single project and involves particularly long time horizons that exceed the scope of short-term plans and funds by years and even decades (e.g., Ribes and Finholt 2009; Karasti *et al*., 2010). A central concern, then, is the question of how CI can be managed in a way that promotes short-term project success *and* long-term sustenance.

We propose that CI centers are important for reconciling these competing demands through their role as *cyberinfrastructure stewards*. Drawing from notions of stewardship advanced by Davis *et al.* (1997) and Dicke (2002), cyberinfrastructure stewardship entails (1) cultivating and maintaining a service ethos and (2) sustaining elements of the cyberinfrastructure beyond the term of any given project. Next we discuss each of these in turn.

Cultivating and Maintaining a Service Ethos

The cultivation and maintenance of a service ethos (or philosophy) by CI centers entails engagement in two ongoing activities. First, it is contingent on the adherence to a *service-dominant logic* (Vargo and Lusch, 2004). An organization that adheres to a service-dominant logic (SDL) acknowledges that its mission is to serve customers; accordingly, it creates value through customer satisfaction. A central idea of an SDL is that the organization enlists customers as active participants in the design of a service blueprint and in the production (or delivery) of an instance of the service. In other words, an SDL-adhering organization sees customers as co-designers and co-producers of its services.

Cultivating and maintaining a service ethos also is contingent on the continuous enabling and performing of computational science. Indeed, this endeavor -- *the enabling and performing of computational science -- is the basic (or fundamental) service provided by CI centers*. Providing this service involves three broad tasks: maintaining existing technological resources; coordinating the application of technological and human resources; and developing new technological and human resources that will satisfy (anticipated) future needs.

Technological resources refer here to digital technologies situated conceptually and architecturally within one of five layers of a *cyberinfrastructure stack*. As Figure 1 illustrates, these five layers include (from bottom to top) hardware, systems software, a programming environment (or development platform), software applications and codes, and domain-specific portals and "gateways." Each of these layers, in turn, comprises a variety of resources. The hardware layer, for example, includes resources related to processing, storage, and networking. Embedded within each layer is a set of standards that enables interoperability between resources within and across layers. Historically, CI centers have assumed full responsibility for the bottom three layers of the cyberinfrastructure stack while ceding some control to scientists (users) at the top two layers of the cyberinfrastructure stack.

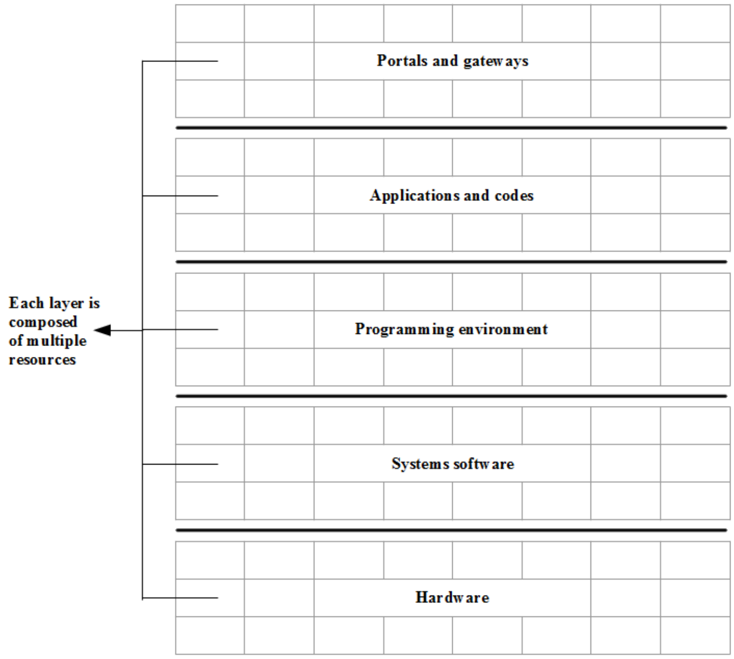

Figure 1: Cyberinfrastructure as a stack of layered digital technologies

The second task associated with enabling and performing computational science -- coordinating the application of technological and human (knowledge) resources -- is a particularly important task. A large CI center may be actively engaged in dozens of computational science projects (e.g., the co-development of parallelized code which simulates the movement of the Earth's mantle), each of which demands a unique set of technological resources (e.g., certain types of processors, a certain software platform, etc.) and human resources (e.g, knowledge of how to parallelize sequential code). As a result, many CI center managers spend considerable time and effort mobilizing a collection of resources assembled for the express purpose of executing a single project, and typical CI centers simultaneously coordinate multiple such projects.

Sustaining Cyberinfrastructure in Perpetuity

Cyberinfrastructure stewardship also entails sustaining elements of the cyberinfrastructure over an extended period of time, or what Ribes and Finholt (2009) termed "the long now." This endeavor involves ongoing, interrelated efforts to anticipate cyberinfrastructure trends and future needs and to secure funding for CI centers and projects. To accomplish these aims, CI center knowledge workers gain experience in mastering digital technologies and in collaborating with domain scientists in order to enable innovative forms of computational science. A CI center's effectiveness in mastering the digital technologies and effectively collaborating clearly hinges on its ability to secure funding. The ability to secure funding, in turn, hinges on the strong ideas borne of collaboration and administration. Cyberinfrastructure stewardship is thus highly sensitive to self-reinforcing patterns of positive (or negative) feedback that are often described in terms of virtuous (or vicious) cycles (Garud and Kumaraswamy, 2005). From a network point of view, CI centers that manage these cycles effectively occupy a "structural fold" (Vedres and Stark, 2010) in a cyberinfrastructure network, effectively enabling cyberinfrastructure innovation over time.

Cyberinfrastructure Centers at the "Structural Fold" of Cyberinfrastructure Networks

A central tenet of social network analysis is that some actors (including organizations) try to identify and occupy brokerage positions within sparse networks that are rich in "structural holes"

(Burt, 1995). In these positions, the actor bridges two unconnected actors in order to access novel information (e.g., a new idea, process, or practice) that can be leveraged to create new knowledge (Obstfeld, 2005; Burt, 1995). However, because this actor typically does not have a history of close interactions with its new connections, its ability to act on this new knowledge tends to be limited.

It is not surprising, then, that a handful of innovation network studies have proposed ways in which actors can bridge one or more structural holes (in order to *get* new ideas) while also belonging to an alliance (in order to *act* on new ideas) (see, e.g., Obstfeld, 2005). This aim can be achieved by the actor that occupies a *structural fold* (Vedres and Stark, 2010). "Generative recombination" that creates knowledge and results in cyberinfrastructure innovation requires access to diverse technical and human resources, as well as a deep familiarity with those resources (Vedres & Stark 2010). In short, and as illustrated graphically in Figure 2, the occupant of a structural fold must be a "multiple insider" with membership in more than one cohesive group.

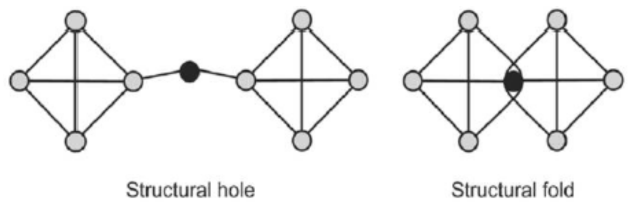

Figure 2: Vedres and Stark's (2010) structural fold juxtaposed with Burt's (1995) structural hole

The idea of a structural fold is valuable to the study of cyberinfrastructure because it suggests that the ability of a cyberinfrastructure network to create knowledge and innovate may be enhanced when one or more actors occupy a structural fold position. Our research indicates that many CI centers satisfy the conditions for occupying this position: they interact closely with many of their customers (e.g., customer as service co-producer), have an intimate familiarity with cyberinfrastructure resources, and they maintain ties with customers and other stakeholders representing a wide variety of knowledge domains. Thus, our research identifies the important role of CI centers in the cyberinfrastructure context. Next we will discuss the role of the leaders within these centers.

CI Center Leaders as Entrepreneurs

CI centers generally have minimal organizational structures, and even the largest centers have only a handful of people in organizational leadership positions. This leadership (with the occasional input from advisory boards) comprises the governance of such centers. The leadership, however, is critical to virtually every part of a CI center. Center employees point to leaders when explaining successes and failures of the CI centers and when describing the selection of projects and the direction of the centers. Many center leaders founded their center, are principal investigators for major sources of funding to the centers, and are heavily involved with a broad variety of center activities. At the same time these leaders are highly educated and talented domain specialists who have multiple academic and practice-oriented pursuits. In many ways, center leaders are like entrepreneurs who are actively involved in virtually every aspect of their organization and whose attention shapes the patterns of activity for the organization (Ocasio 1997). Because they are so important to centers, and their attention is pulled in so many directions, we investigated where and why these leaders focus their attention and the implications of this attention on cyberinfrastructure innovation. A stream of organizational scholarship focuses explicitly on leadership attention, which can provide a lens for making sense of cyberinfrastructure center leadership.

Attention is a scarce resource (Deutsch & Deutsch 1963) and the things that people chose to pay attention to varies by the individual (Schneider & Shiffrin 1977). Management literature offers an "attention-based" view of the firm (Ocasio 1997), since things that do not capture a decision maker's attention are not acted upon. Key leaders in cyberinfrastructure centers are similar to other contemporary organizational contexts in that they are bombarded by information and competing demands for their attention (Hansen & Hass 2001), often in the form of competing projects. By examining the literature on attention, paying particular attention to Herbert Simon's (1957) original thoughts on attention, we suggest there are two important insights that begin to explain the patterns surrounding leadership attention of successful CI centers: (1) the influence and situational factors of the various domains in a CI center ecosystem, (2) the different modes of attention necessary to generate and complete the cycles of projects that feed a CI center.

CI centers nearly always serve two domains simultaneously. They are developing and redeveloping cutting edge cyberinfrastructure (computer science, engineering, and applied mathematics) to better serve cutting edge domain science (physics, chemistry, biology, humanities, etc.). In other words, the cyberinfrastructure being developed is often the tool to answer a question in a domain science. While serving both of these groups, they exist in a unique organizational structure that receives resources on a project-by-project basis that contains its own management challenges. Leaders may closely associate with the computer science roots or the domain science the center serves. In this way, we see similarities to Simon's (1957) insights about the effects of goal-alignment on attention as a way to mobilize attention patterns. Leaders are likely to pay attention to problems they feel are directly related to their personal goals and values (which Simon refers to as "inducements").

This resonated with the panelists, but they added that some leaders may be mobilized not by one domain group or the other as much as the idea of the intersection between the two. The term "renaissance scientist" was used on Day 1 to describe a sophisticated cross-disciplinary player, but it is often used to describe a person who is expert in both his cyberinfrastructure and domain area – a la da Vinci (see Figure 3). We suggest that while such capability is a worthy ideal, a more attainable and useful model is having adequate complementary expertise.

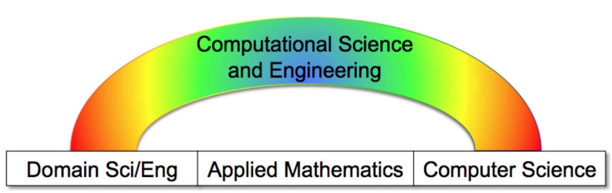

Figure 3: The domain of the renaissance scientist
(image from Phil Westmoreland)

The second insight that we glean from the attention literature and see evidence of in CI center leadership is the difference between two modes of attention: (1) initiating - when attention is initially directed at the task and (2) sustaining – the maintained attention on the task for a long period of time.

The initiating mode of attention is punctuated by a burst of creative ideas to solve a large problem, and shares many traits with entrepreneurial literatures which describes the entrepreneur as charismatic, passionate, and lacking the ability to delegate. The sustaining mode of attention is a long-term commitment to putting the "plan" into place and involves implementing the solutions decided upon during the initiating mode. It shares similarities with professional management phases which focus on implementing procedures and controlled growth. It seems that centers must iterate between initiating and sustaining modes of attention in order to constantly have innovative ideas that secure sources of funding, yet maintain the sustained attention of following-through with the solutions and see them to fruition (illustrated in Figure 4).

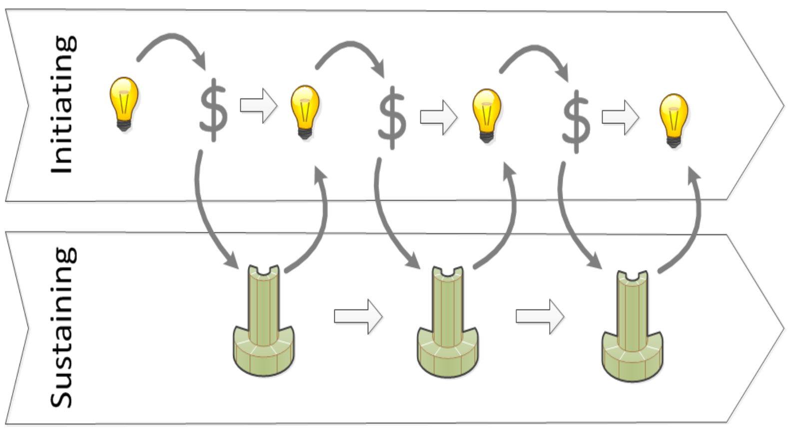

Figure 4: The interplay of two CI center modes of attention over time

Innovations take the form of technological advances as well as organizational enhancements that change the way cyberinfrastructure is delivered by the center. Therefore, this attention based view seems promising in order to understand what types of innovations come from each mode of attention and how centers balance between the two modes. Further, we found that successful centers with high levels of performance iterate and balance between initiating and sustaining modes of attention.

Several center participants suggested this tension between initiating and sustaining modes of attention resonated with their experiences inside CI centers. Further, we believe a fruitful area of research will be to combine the two insights from our attention-based view of CI centers and explore how inducements aligned with innovation areas and attention modes interact. We offer a possible framework from this line of thinking in Table 1.

One suggestion from the panelists was to consider how the goals of the center interplay with the expected output from individuals. For example, if a center is focused on furthering domain science, and is therefore expected to enable domain science publications, it is likely that those focused on the domain science will fall into initiating modes of attention, where those focused on computer science will tend towards sustaining modes. Our research suggests a continuous cyclical (clockwise) dynamic around the framework shown in Table 1.

	Organizational Innovation	Technology Innovation
Initiating Behaviors	**"Mobilized around the Problem"** Large-scale "systems builders"	**"Mobilized Around the Solution"** Bring resources together – project initiator
Sustaining Behaviors	**"Enabling Solutions"** Creates solutions and plans from resources at hand	**"Delivering Solutions"** Improving and supporting technology

Table 1: Modes of attention among CI center leaders and forms of innovation

Whether we characterize scientists as entrepreneurs or as renaissance scientists, it is clear that CI centers are strongly dependent on the leadership of very few key individuals who are involved with virtually all aspects of the center. One of the classic tenets of entrepreneurship research is that individuals can more effectively guide entrepreneurial organizations in the early years through creativity and dedication, but this same energy - often resulting in micromanagement - can adversely affect the organization as it grows (known as "the Founder's Trap," see Adizes 2004). Throughout the growth of an organization's evolution, gradually increasing levels of professional management have been found to have a positive influence as an organization grows (Adizes 2004). However, as the panel indicated, there is a limited role for professional management given the institutional setting of most CI centers, which are typically affiliated with one or more universities. Going forward it will be important to investigate how and when the introduction of professional management can benefit centers. Next we will address another management challenge to CI centers - that associated with human and technical resources.

Resources Scarcity & CI Innovation

Organizational scholars have spent decades studying the ways in which organizations innovate across a variety of rich traditions (Burns & Stalker, 1961; Cyert & March, 1964; Van de Ven, Polley, Garud, & Venkataraman, 1999). However, many of their findings may not be relevant for the unique case of digital infrastructure innovations, because such innovations are markedly different from traditional technological and administrative innovations (Hanseth & Lyytinen, 2010; Star & Ruhleder, 1996; Tilson et al., 2010). A key set of criteria by which digital infrastructures are different involves the resources that they require, and the long time horizons they involve (Ribes & Finholt, 2009). Therefore, to gain insight into digital infrastructure innovation, we explore the literature relating resource scarcity and abundance to innovative activities, and then look to understand when digital infrastructures may be similar or different than other forms of innovative activity.

When studying the impact of resources on innovative activity it makes sense to investigate *behaviors* commonly associated with innovation. The seminal work of James March (March, 1991) proposes two broad search behaviors commonly associated with innovative activity: exploratory and exploitative search. Exploration involves looking to unfamiliar domains and establishing new competencies to drive innovation; whereas exploitation involves leveraging familiar domains and existing competencies to drive innovation. A rich literature links organizational resources – particularly the availability of abundant resources - to exploration and exploitation behaviors (Cyert & March, 1964; Hoegl, Gibbert, & Mazursky, 2008; Nohria & Gulati, 1996; Voss, Sirdeshmukh, & Voss, 2008). It is the goal of this research to extend this body of work by exploring the relationships between resource levels and digital infrastructure innovation.

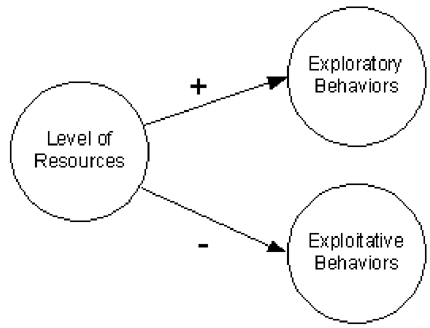

Figure 5: The received view of the impact of resources on innovative behavior

March's original model (1991) implicitly made a number of assumptions, including: (1) funding was gained from successful past innovations; (2) innovations involve products and processes that can be marketed; (3) everything occurring within a firm; and (4) reflected a learning model that emphasized knowledge homogeneity. Cyberinfrastructure contexts are different because: (1) funding based on idea and reputation (more similar to venture capital funding); (2) innovations are generally provided free to the world; (3) boundaries of centers are fluid with the universities they're associated with and across centers; and (4) centers are necessarily cross-disciplinary almost by definition.

Furthermore, recent work argues for a more complicated view of the relationship between resources and innovative behaviors (see Figure 5). Scarce resources have been associated with increased exploration and increased exploitation. Exploitation is typically because of the need to drive efficiency and the constraints that scarce resources put on exploratory modes of search. However, on occasion these scarce resources can impart a certain hunger or desperation in organizations and these organizations can take dramatically bold steps - essentially "swinging for the fence." Resource abundance, on the other hand, is generally associated with increased exploration. This exploration can either lead to fruitful, radical innovation, or may lead to dead-ends. However, it is important to also understand that abundant resources encourage organizations to pour more of these resources into what they do - potentially leading to dramatic gains in optimization or scale, but also potentially to a "competency trap" (Leavitt & March 1988), whereby the skills that have paid off in the past lock-in and become inflexible to alternative and emerging competencies.

	Abundance	Scarcity
Explore	Sandbox	"Swinging for the fence" *Dramatic victory or failure*
Exploit	"Optimization & scale" *Growth or lock-in*	Efficiency

Table 2: The relationship between resource levels and innovation behaviors.

Thus we find a dynamic relationship between the level of resources and an organization's modes of innovative activity. When applying this to the context of cyberinfrastructure, we find that many of these relationships are quite different. First, funding of centers often follows exploratory activity, rather than the other way around. Thus the directionality of the classic relationships do not appear to always apply to the cyberinfrastructural context. Second, resource abundance at one layer of the cyberinfrastructure stack (as discussed in Figure 1) may be accompanied by resource scarcity at another layer. The extreme amount of money spent on "big iron" in the 1980s and 1990s, accompanied by relatively little on software was credited by many center participants with generating some extraordinary innovations - including significant contributions to the open source movement. Thus the interplay of resource scarcity and abundance at different parts of the stack is important for generating innovative activity. Third, consistent with the long-time horizons of cyberinfrastructure, centers undergo periods of relative scarcity and relative abundance in sequence.

Without exception, every CI center that we studied went through some period of resource scarcity. They responded to this resource constraint a variety of ways which included a variety of cutbacks and layoffs. Some centers weathered the resource scarcity through focus and efficiency (exploitation) and in some cases they essentially ran out of money and decided to swing for the fence (exploration). This interplay between resources scarcity and abundance is fundamental to activities in CI centers and our analysis of the impact of the situation has only touched the surface of this important issue. As indicated above, software innovations are key elements of cyberinfrastructure, perhaps *the* key element going forward as hardware becomes increasingly commoditized. Next we will discuss the role of the CI center and scientific software ecosystems.

Scientific Software Ecosystems

Software is the "relational web" upon which distributed collaborative scientific research operates (Spencer et al 2011) and CI centers produce a variety of software artifacts and other standards, the impact of which extends far beyond the project at hand (Spencer et al 2006). While successful management of software development is a challenge in all kinds of organizations, software development in CI centers combines some of the hardest challenges of other environments and adds some unique challenges of its own.

The design of software is challenging because it involves bringing together knowledge of the needs of potential users and the technical skills necessary to build the software. The implementation of software requires drawing together contributions from many participants in a synchronized and coordinated manner. These challenges are compounded in distributed environments (e.g., Herbsleb & Mockus, 2003) where lack of consistent face-to-face contact undermines the coordinating flow of informal communication between developers that seems unavoidably fundamental to the successful development of software. CI centers, particularly when operating in a virtual mode, face these challenges. Further, once built, scientific software must be sustained; CI centers face four particular challenges here:

- First, science-funding cycles have traditionally been focused on discovery, rather than sustainability.
- Second, much software development work in science is undertaken by relatively junior participants (graduate students and post-docs) who are incentivized to move away from software work if they are to advance as scientists (Howison and Herbsleb, 2010).
- Third, many scientific software contributions are written for specific analyses, facing publication deadlines. Such code is quite different from that designed for infrastructural reuse and converting it to be so is difficult (Howison and Herbsleb, 2010).
- Fourth, the scientific reputation economy does not deal well with rewarding software contributions, despite their importance for scientific progress (Howison and Herbsleb, 2011).

As a result of all of these challenges, software has been found to be a consistent point of weakness associated with collaborative research centers, including in crucial areas such as climate models (Merali, 2010; Stodden et al, 2010). In part this stems from the complexity of the software infrastructure, which is composed of many layers written by many different people for many different purposes (Edwards, 2010; Jackson et al, 2007).

The initial enthusiasm for combining computer science researchers with domain specific research projects has dampened in recent years because the need for developing sustainable software environments requires organizational and process capabilities in addition to the technical capabilities typically offered by computer science (Olson et al., 2008, p. 81). The disciplines of software engineering and information systems both look for processes and methods associated with bringing the technology and the domain together. Software engineering tends to emphasize the "requirements analysis” portion of this combination (Hansen et al 2009), whereas information systems literature tends to emphasize the “alignment” or “fit” between the organization and the information technology (Henderson & Venkatraman 1993; Goodhue & Thompson 1995). There is potential to adapt these literatures, with the understanding however that the challenges in the scientific domain are perhaps even more complex given the very dense knowledge required for scientific applications and the exploratory, unfolding nature of research (Segal and Morris, 2008).

CI centers are involved in scientific software production in a small number of different organizational configurations. The first is the software project, the production of a specific piece of software by a relatively cohesive team, over a relatively well understood timeframe. Such software is either for internal use or produced to be released to others, perhaps for use in other centers or by scientists themselves. This form of software development is already complicated, as shown by the very wide range of management techniques, from waterfall to agile, from personal software process to capability maturity models (SEI). Software projects often span multiple physical locations and, through collaborative grants, multiple administrative centers, but retain their planned quality.

The second organizational configuration for software development is the software community. This is epitomized by a popular image of open source projects: spanning multiple organizations and intentions, drawing together participants with multiple agendas, spanning across funding sources and with emergent rather than planned properties, but retaining a focus on a small number of software packages, usually one. CI centers participate in software communities from two perspectives: first as initiators seeking external participation and seeking to sustain the life of a software project beyond its initial funding and second as participants in existing software communities. The management skills required to thrive in software communities are different from those in software projects, requiring skills of gentle persuasion and leadership through transparent action (Colfer and Baldwin, 2010), rather than the structured timetables of requirements and tasks familiar to software projects.

Our research on the conduct of software-intensive science has shown that front-line scientists draw together multiple tools from a range of different production systems and providers to conduct their science (Howison and Herbsleb, 2010; Spencer et al, 2011). Rather than users of a few pieces of code (and thus relatively simply users of software produced in projects or communities) science draws on many pieces. Moreover these pieces themselves rely on many others, rapidly implicating a large sociotechnical web in the support of their science. This raises the question of how is the production of software organized at this higher level.

In domains outside science a third high-level configuration has emerged in response to these complex webs: that of the software ecosystem (Herbsleb et al, 2010). Epitomized by meta-communities like Apache and Eclipse, these concern themselves with the production of many pieces of software that more or less work together. The organizations at their center provide a stewardship role beyond the production and sustenance of a particular piece of software.

While writing new software is important, building any non-trivial software system means relying on software written outside the control of any individual organization. Thus managing software development requires making judgments about the reliability not only of external software but those who write it. This is addressed in literature on the procurement of software and software development services (e.g., Hirschheim et al 2008) and, increasingly, in the management of relationships with heterogeneous open source software development projects (e.g, Agerfalk et al, 2008; Crowston and Howison, 2006). CI centers face a particular challenge here because their needs can be very specialized, but their budgets relatively limited. More than one large scientific research project has found itself unable to get timely support from commercial suppliers focused on more profitable customers, or unable to convince sometimes ornery open source developers to accept very specialized patches (Howison and Herbsleb, 2010).

Our work identifies four key functions for the CI center as ecosystem steward. The first is technical architecture, the establishment of a way for multiple pieces of software to usefully interact; this is typically accomplished through a combination of standards and a core platform of code. The second is the establishment of sociotechnical collaborative infrastructure, including the certification of projects within the community, a template for their organization as well as the provision of (moderately standardized) collaborative tools from source code management, to communication including issue tracking. These tools tend to provide exceptional transparency into not merely the outcomes of the work of others but the work as it unfolds (Dabbish et al, 2011). The third function is ecosystem governance, the challenge of establishing decision-making forums and providing leadership above the level of different software communities, bringing sufficient coherence to the ecosystem overall. Governance includes understanding the performance and resource situation of different projects and how they depend on each other. The fourth function is understanding and influencing incentives for the participants in the ecosystem, including reward systems such as Apache's "Member" designation and ensuring that participants are able to achieve rewards relevant to them within the ecosystem, including financial rewards (it is common for Eclipse participants to build both open and proprietary components).

In the workshop, it was suggested that CI centers might come to see themselves as stewards of particular scientific software ecosystems. The discussion between the participants revealed both possibilities and barriers in this regard. CI centers are well placed as relatively long-lived organizations with broad perspectives and insight into shared challenges in multiple scientific domains. As infrastructure providers they are not viewed as direct competitors with large segments of the domain science world. Yet, as panelists pointed out, CI centers have long competed with each other for software project funding and "one of the things that we've learned about ecosystems stewardship. If you're in the game, it's hard to set rules for the game."

In terms of insight into the activities and needs of the user community participants identified a tension between a role as relatively blind providers of cycles and relatively engaged consultants working directly with users to implement and optimize their scientific computing. The second does provide insight relevant to understanding the strengths and weakness of the ecosystem, while the first does not. CI centers clearly look to funding agencies for action on incentives, not seeing it as part of their current or possible roles. Overall there was a belief that the forthcoming initiatives such as NSF's Scientific Software Institutes program ought to provide appropriate vehicles for ecosystem stewardship.

The question of funding sources raised an interesting distinction between the complexity of software production in science and without. The vast majority of funding for scientific software comes from national grants through organizations like the NSF, DoD and NIH. This can be contrasted with the more diverse sources of resources drawn together in ecosystems like Apache and Eclipse, which attract resources from many companies in different industries as well as individual effort contributions. Meta-projects in science may rapidly find themselves trying to coordinate projects which are directly competing for the same pools of funding. CI centers, however, diversify funding somewhat by drawing together contributions from universities and, in many cases, state economic development funding. In coming years this emphasis on scientific software will only increase, and the role of CI centers is central to scientific software going forward. To date, CI centers have had a strong impact on many forms of software (Internet, open source, scientific computing, etc.), but assessing this impact - as well as all other impacts of CI center activity - is one of the perennial problems with the management of cyberinfrastructure. Next we will address the challenges of evaluating CI center work.

Assessing CI Impact

Technological innovations rooted in CI centers from thirty years ago have resulted in an overwhelming return to the nation in terms of technological innovation in a variety of fields, including laying the foundation for much of the Internet itself (see Tuomi 2002). Further, CI centers enable path-breaking research findings across an ever-increasing number of fields. Some of the most important recent findings in fields as diverse as astrophysics, climate science, and plant biology can be traced to CI centers. However, it is notoriously difficult to assess the impact of the work of CI centers. Scientific impact is often measured through citation networks that result from the publications of sponsored research, or cycles of computation provided. However, these are problematic measures - and can much less easily be applied to CI technologies higher in the stack such as software. Further, qualitative technology and science examples and case studies can do much to show the impact of previous CI efforts, but these examples are far removed from investment or other organizational decision making. The Internet revolution began years and even decades after the base technologies were developed with the help of CI centers (Tuomi 2002). By the time scientific publications are deemed impactful, they are typically years removed from the use of cyberinfrastructural resources. There are no clear cut answers, but our research has addressed some issues around the evaluation of cyberinfrastructure. Next we will characterize infrastructure and then reflect on the challenges to evaluation.

Choosing *infrastructure* as a unit of analysis for evaluation generates a host of difficulties that we will discuss below, and we aim to make clear the reasons for considering CI Centers and their functions a more reasonable and measurable site of evaluation. In its working order, infrastructure is invisible - it is precisely the *infra-* nature of our built systems that make them so useful. They get out of our way and free us to perform more specific types of work. The invisible nature of infrastructure also implies that in order to study the objects, processes, people, and information that compose infrastructures, we must somehow render them visible, and this is the catch. Working infrastructure is invisible: Visible infrastructure is not working (in the ways we need it to.) Thus, the study of infrastructure relies on a sensitive handling of the rendered subject, and cannot be held in gaze for too long without putting undue stress on the work it performs. The consistent focus on studying cyberinfrastructure keeps the structures, the people, and the processes in constant view and may impede its sinking to the hidden and mostly frictionless layer where it performs the way it should.

As an object of discussion, *infrastructure* is relatively young as a concept, with its first contemporary appearance in 1927 as a military description of interconnected roadways, waterways, electrical grids, and the like (Hughes 1983). As an analytic field, infrastructure studies are nascent, gaining ground with the major theoretical works of Thomas Parke Hughes, Susan Leigh Star, and Karen Ruhleder in the 1980s and 1990s (Hughes 1983, Star & Ruhleder 1996). Thus, we are barely a century into the program of articulating what infrastructure is, and much less experienced in gaining control over the mechanisms behind it. However, we do know some properties of infrastructure that make its evaluation difficult. As indicated in our previous articulation of infrastructure, it is clearly accretionary. By this we mean that infrastructure does not assemble or come into being all at once. This makes it clear that the "long now" perspective is core to the concept - infrastructure is realized and made sense of in hindsight. It cannot be planned so much as it can be "grown" or "fostered" (Edwards et al 2007). Further, infrastructure is grown from a combination of top-down and bottom-up. It is not dictated and designed, but negotiated and interacted between a center and the margins of a network in an attempt to aggregate and standardize locally valuable solutions. Infrastructural evolution is continually evolving, unpredictable, path dependent, and sensitive to initial conditions.

As mentioned before, working infrastructure is invisible, particularly when it is making its biggest impact. In general, infrastructure is taken for granted and essentially ignored when users are using it. To an experienced, working scientist, successful infrastructure is backgrounded while the focus of her work is foregrounded. If it becomes visible, then something must have gone wrong or become problematic. The longer "cyberinfrastructure" is held as a visible analytic object to the working scientific community, and domain scientists are continuously recruited into the discourse, the more their attention is taken away from the work with the cyberinfrastructure.

What specific qualities of cyberinfrastructure make it resistant to evaluation? First is the sensitivity to initial conditions. The aggregative and *post hoc* nature of identifying infrastructure means that it is impossible to claim with accuracy when the convergence of systems has precisely achieved the qualities and qualifications of infrastructure. Consequently, If we cannot define where we started, we cannot determine how far we have come. Second, there is a subtle form of paralysis when attempting to base strategic decisions and resource investments on the assumed emergence or development of sprawling, heterogeneous, and complex infrastructure. Infrastructure is large and distributed enough that there is no locus of control, which defies most paradigms of operations and resource management. Given the typical magnitude of resource investments associated with fostering cyberinfrastructure, the irreversibility of decisions makes CI center leaders hesitant and risk-averse. In those situations where they are actively looking to build infrastructure, they must navigate different and often incommensurable epistemologies of goals, outcomes, quality, and success. Also, since infrastructure accretes in part from the margins, it is impossible to keep a reasonable inventory of where relevant contributions are being made, and the effects of resource development and allocation as they move beyond specific projects - the impact of cyberinfrastructure development is lost beyond the first order. Therefore it is equally impossible to reward appropriately and/or sustain the right balance of resource allocations based on performance. Cyberinfrastructure development to date has been based upon an engineering model of “plan & build” which encodes a (complicated, but) linear approach to problem solving. There is no form in which cyberinfrastructure is specifically planned or built in such a straightforward way. All infrastructures - cyberinfrastructure included - are in a constant, unstable process of “becoming”, and thus there is an instability in any evaluative framework of cyberinfrastructure as an analytic object.

Despite the resistance of cyberinfrastructure, *writ large*, as candidate for easy evaluation, CI centers hold more promise. This is not to say that the centers themselves should become the objects of evaluation - the processes they facilitate and resources for which they serve as a gateway or catalyst are what need evaluation. The CI center can serve as a powerful lens through which we see and understand operating parts of infrastructure - a process infrastructure scholar Geoffrey Bowker has termed “infrastructural inversion” - and holding them in the analytic gaze with more control and sensitivity to letting it return to an invisible, working state (Bowker 1994, p 34).

How, then, can we evaluate the scientific impact of CI centers? The answers are equally “becoming”, and as unstable as the cyberinfrastructures they purport to assess. Mindful infrastructural inversion gives us some early answers and points us in promising directions. Given our research, we make the following reflections from the workshop:

1. Third Party Evaluators: Users of cyberinfrastructure do not want to continually attend to the digital infrastructure upon which they do their work - particularly if it’s working. Domain scientists, for example, care about their work, their data, their reputation, their context, and the reward structures associated with scientific production. Scientists that are users should not be responsible for the evaluation of cyberinfrastructure. Evaluation should not involve their explicit attention of scientists to the infrastructure. Further, domain scientists are not equipped with the theoretical and practical expertise to analyze and develop large sociotechnical systems: social and organizational scholars, and technologists who specialize in cyberinfrastructure development are appropriately trained in these areas. CI centers, as sites of convergence and negotiation, can serve as the buffer between scientists who need to do their work, and those of us who need to be considering cyberinfrastructure as an object. As aptly remarked by one of the workshop participants, analogizing the relationship between cyberinfrastructure inquirers and users, “*The way that we’re engaging scientists is basically every time they flush the toilet we’re asking, ‘Did it flush right? Did it flush right? Did it flush right?’ So we’re keeping the infrastructure in front of them every time they use it because we want to know about* *cyberinfrastructure. We're trying to develop evaluation systems, which keep engaging them all the time...which are interruptions to their work*."

2. Reward Systems: Since infrastructures evolve over longer time cycles, and are only identifiable after systems have aggregated and faded into the background, the typical cycles of evaluation born out of a laboratory-based scientific production model - 3 to 5 years typically - are wholly inappropriate to gauge the success or impact of system-building activities that take decades to form. With such a disparate gap in the phenomenal rhythms between domain sciences and cyberinfrastructure studies, those who study the latter are in a precarious position with regard to the reward structure firmly embedded in the academic enterprise (Jackson 2011). First, there is a possible problem of conflation: the success of a cyberinfrastructure is not the same as the rate of successful innovation and new findings in the domain sciences it supports. There is a general tendency to emphasize outcomes - successful new scientific findings and successful new innovations. However, the same practices that lead to successes in both science and technology innovation, also often lead to failure. Evaluations and reward systems for cyberinfrastructure scholarship should emphasize appropriate behaviors to fostering innovation and not so exclusively on scientific outcomes. Behaviors associated with technological innovation would involve fostering trial-and-error critical testing of innovations, new prototypes, tracking of impact in products (databases, code libraries or classes, productive employees, algorithms, etc) at the second- and third-order of application, etc. Remediating evaluation systems to be compatible with cyberinfrastructure and CI center activities involves creating novel measures such as innovation indexes or "long new" sensibilities of scientific impact (such as h-index for infrastructure). Second, as mentioned in the previous section, the incentive and reward structures of domain scientists and cyberinfrastructure scholars are orthogonal. Domain scientists want invisible, working, seamless cyberinfrastructure. Their reward structures require looking past the structure to to the objects of their own scientific inquiry, to publish quickly and often. Cyberinfrastructure scholars are working at odds, looking for each opportunity to render parts of the infrastructure visible, gathering glimpses here and there, often with domain scientists who can get frustrated with interruption, only to publish speculative and incomplete descriptions of a systemic behemoth that defies a unified artifice.

The incentives are weak here, and the reward structures somewhat unforgiving. Although we are very good at measuring and evaluating the known elements of cyberinfrastructure - exploitative, Kuhnian "normal" science and incremental innovation (Kuhn 1996) - we are theoretically and methodologically uncomfortable with evaluating exploratory science, precisely because the metrics born out of exploitative science - the main tools in our evaluation toolbox - make exploratory science look risky, unproductive, and low on ROI. Another pointed comment from a workshop attendee summed up the sentiment quite elegantly: "*People who built the highway got paid with money and they made profits. They're not in an academic reputation economy. The people building infrastructure often are in an academic reputation economy. So, the desire to have the infrastructure disappear seems to be difficult if you also want to earn academic credit from the domain scientists or from the journals."*

3. Measurement Instruments: As discussed, we cannot measure infrastructure directly, so we measure the output of systems and activities associated with the project, facility, or other smaller units of analysis, such as cycles on a supercomputer. This approach has been adequate in the past, but going forward, we cannot continue to measure without equal acknowledgment that infrastructure also involves software, organizational and individual social values, management processes, policies, standards, regulatory and governance regimes, and (importantly) data (Lee 2006, Edwards 2007, Knobel 2011).

Software in particular is a problem because reward systems and measures do not exist for enhancing software, which is an ongoing and dynamic process, outlasting the initial development phase. An important point about evaluation is that it is difficult for static measures to evaluate dynamic phenomena with any reliability. Apples in year one are compared to oranges in year two. The moment context changes, the validity of the measurement instrument is out of sync and we inductively generate a new instrument for this new context. Static instruments have difficulty with dynamic phenomena, but adaptive, learning instruments such as human beings do not. In other areas of academia, we use expert evaluators to judge impact: a similar process can be used for cyberinfrastructure. Again, however, such a human-oriented process can be quite good on average, but is ineffective in assessing the value of new and marginal innovations. Further, we typically downplay the trustworthiness of human evaluation. Human evaluation is not considered to be "objective" and is therefore devalued in preference to hard numbers. Concentrated development in the area of mixed or hybrid methods - combining the strengths of qualitative and quantitative approaches - seems a natural direction of investment for cyberinfrastructure studies and those that provide funding.

In particular participants argued that users of their infrastructure and software don't necessarily provide citations in the scientific publications the users write. The lack of relevant citations limits the applicability of bibliometric analysis, the most common quantitative approach to evaluation of scientific impact. Moreover, since centers are often asked to provide lists of resulting publications during scientific evaluations, centers themselves need to spend considerable time "figuring out" which publications drew on their services. This is a clear area for improvement.

Finally, there are measures of resource scale - such as the top 500 supercomputers - that were never really effective in ranking sustained performance over time, but were valuable as proxies for research infrastructure potential nevertheless. Such ranks are considered increasingly less relevant as the diversity of cyberinfrastrucural resources flourishes - in particular systems (including software) for data manipulation and visualization, as well as platforms for the open sharing of data, pre-publication findings, and open-access publishing.

In summary, the basic concept of *evaluating cyberinfrastructure* defies the historical and *post hoc* nature of what infrastructure is, and the current tools of measurement are ill-equipped to reveal useful insights about assemblages at such magnitude. Still, contemporary administrative life demands accountability, and we must evaluate and measure the progress of our science. Setting the expectations of the infrastructure separate from the processes and products of the sciences it supports is a strong first step. CI centers provide an opportunity to observe activities within cyberinfrastructures at discernable and describable levels of scale, and point to more reasonable units of analysis - objectively and temporally - upon which we can gain traction and evolve new evaluation regimes suited to sciences of complex system and infrastructure emergence.

Agenda Going Forward[2]

It is clear that (at least a subset of) organizational scholars are interested in the particular form of organizing we describe as "CI centers." These centers fill a niche in the scientific ecosystem - stewarding the evolution of cyberinfrastructure. Given their successes, organizational scholars may learn something about particular forms of innovation. Further, although CI centers have

[2] This section benefited greatly from panelist input - discussion during the workshop and panelist notes and presentations. See appendix for the list of panelists.

been incredibly successful in the past, we do not have a good idea as to whether things can be even better, whether the current modes of organizing are the most efficient and effective, and whether the management practices that have worked in the past will continue to work in the future. Thus we have a situation where organizational researchers, CI center managers, and cyberinfrastructure policymakers have a common interest in understanding the management of this unique organizational form. As a result of the panel, following are suggestions for this collaboration going forward.

Continue Interaction through "Research Coordination Network"

The panel was (to our knowledge) the first explicit meeting between organizational researchers and CI center managers to explicitly discuss the organizational form of a CI center. One key to moving forward is to increase the occasions for cross-pollination in the future, consistently over time. A step in this direction is a "Research Coordination Network" recently funded by the NSF to bring these groups together in a series of topical workshops (initial three topics are: virtual organizations, managing cyberinfrastructure, and scientific software see Berente & Howison 2011). In these workshops it is important to include a wider variety of organizational scholars, such as organizational behavior and human resource scholars. Further, it would make sense to connect this work to the work on other cyberinfrastructure-oriented organizations.

Conceptual Clarity about CI Centers as a Category

CI centers, and centers in general, are different than universities, national labs, institutes, government agencies, projects, virtual teams, firms, corporate departments, start-ups, and a host of other organizational forms, but in important ways they resemble or relate to each of the above. Key feedback from the workshop involves being constantly cognizant of the distinctions, similarities, and the relationships going forward. For example, CI centers are not entrepreneurial start-ups, therefore the lessons from entrepreneurial scholarship may apply but they might not. Similarly, CI centers tend to have strong ties to universities and funding agencies, whereas corporate data centers generally serve a single company with a profit motive. CI centers have a broader, even national implications whereas datacenters in firms mainly impact only one organization. When investigating this critical category of organization, it is important to foreground these comparisons.

Sustainability, Longevity and Temporal Issues for CI Centers

The typical funding lifecycle for NSF is 3-5 years, with an upper bound of approximately 10 years. For infrastructure projects, the mindset must be adopted that the developments are expected to have a life beyond the duration of the initial funding. It is quite reasonable for something that starts as a project to grow and turn into something larger, and this distinction should be more integral to the definition of a CI center.

The long-term emergent properties of cyberinfrastructure cannot be overemphasized. Cyberinfrastructure evolves beyond project funding cycles. CI centers are critical organizations for coordinative knowledge between projects over time - acting as stewards of elements of the cyberinfrastructure. Projects come and go, but the centers maintain the knowledge from project-to-project to enable the long term evolution of the cyberinfrastructure. They nurture and guide this evolution. However, funding models are less center- and domain- focused, more project-focused. Cyberinfrastructure-related projects have some conception of the long-term, but these conceptions are different. Going forward it makes sense to critically address how different stakeholders account for the long-term in funding decisions. When does it make sense to maintain an asset as part of the academic environment, and when does it make sense for them to be maintained elsewhere? Which services are better distributed among a broad portfolio of providers (and their segments and contexts)? Decisions about “make or buy” take on a different complexion in the context of a national cyberinfrastructure for science, but these decisions need to be addressed nonetheless. A focus on long-term sustainability of an innovation trajectory needs to incorporate explicit attention to the sustainability of CI centers.

Organizational Design and Benchmarking

Particularly as centers grow, it is important to understand tradeoffs associated with the design of organizations. What works and what doesn’t? When should CI center managers structure one way and when should they structure another way? A critical path forward is to begin benchmarking successful CI center practices and disseminating these successes. Although centers are competitive in many ways, according to the panelists this is a sort of “pre-market” competition that has a variety of cooperative elements.

Internal metrics (not to be confused with scientific impact evaluation that we described above) are critical for prioritizing organizational activity and informing organization design. CI centers have many such internal metrics - one panelist indicated that his center has thirty “home grown” metrics that they regularly report to their board. So such evaluation is being performed, but not in a way that is informed by research across contexts.Critical questions going forward involve how to capture information about what works and what doesn’t; what should be the goals and metrics for CI centers; and which corporate references (i.e., balanced scorecard, ServeQual, Baldrige, etc.) apply and how. This is not an unproblematic issue, however, because CI centers do compete with each other and therefore may not always be willing to disclose key activities.

Further, CI centers generally do not exist in a vacuum, but have tight relationships with universities and institutes, and it is important to understand how they impact the parent institution, and how relevant synergies can be improved. Similar attention should be addressed toward fit with national priorities. It is important to note that CI centers serve functional needs of parent institutions and government agencies, but in addition, they also support symbolic needs of these organizations. To have a CI center gives certain impressions to key stakeholders and legitimizes related institutions in certain ways. Symbolic roles of CI centers should not be dismissed as fluffy and irrelevant because often times it is the symbolic role that enables access to resources that enable success. Finally, with a focus on academic and national contexts, it is important not to ignore commercial firms and infrastructures and avoid focusing too much on open source and government-sponsored phenomena.

Governance

Hand-in-hand with the design of CI centers involves how decision rights should be allocated across a CI center - a concept typically referred to as organizational governance. A particularly important issue associated with the governance of such centers involved the allocation of knowledge workers across different groups and to multiple activities within the center.

The prevailing way that centers (and other technical organizations) allocate knowledge resources generally involves some sort of “mirroring” (Baldwin 2008). The technology architecture is reflected in the organizational structure. Different units work on different aspects of the architecture. SImilarly, CI centers allocate knowledge workers across the digital infrastructure stack. Our findings about stewardship, however, indicate that CI centers span the stack, therefore a straight mirroring perspective is likely not enough. Future research should address the proper governance of CI centers that emphasizes the spanning of domains across the stack.

This attention to the governance of an organization and the infrastructural architecture calls attention to the relationship between platforms and infrastructure. Platforms and infrastructure have a complex relationship. Infrastructure can be a platform, can use platforms and can be part of platforms. CI centers are both participants in platform communities and stewards of platforms and catalysts for the emergence of platforms.

Further, as cyberinfrastructure expands broadly to new user domains, CI centers are dealing with different sorts of users. Panelists indicated the existing, well-established "deep" users of the cyberinfrastructure resources act as partners with the CI centers. Such users are highly knowledgeable, technically competent, and work at a very in-depth level with CI center personnel. "Wide" users, on the other hand, require much different sorts of services. Wide users is a term panelists used to refer to new user communities (i.e., scientific disciplines beginning to use cyberinfrastructure resources). Organizational scholarship indicates that the set of competencies for a high-end user community are typically not the same competencies required to service a low-end user community (e.g. Christensen et al 1998). The two should be decoupled. It is important to understand how this insight applies to deep and wide cyberinfrastructure users and the associated relevant competencies, technologies, and organizational structures.

Finally it is important to understand how commoditization of certain services (i.e., moving computing cycles to Amazon, for example) would affect the national cyberinfrastructure. Iinfrastructures (or ecologies) are heterogeneous systems with deep layers of non-linear and co-adapted subsystems and relationships. Disaggregation and offloading of identified components that are based on potentially superficial categories (computing cycles, knowledge management, data storage, software tool development etc.) can fail to account for the nuanced relationships between what are considered core competencies and support services. Thus, the capacity to engage core work, core competencies, core "science", may be diminished in unpredictable ways when the local ability to manage and adapt scaffolding is taken away.

Human Resources and Education

Human resources are a critical concern for CI center managers. How do you attract, train, and keep highly skilled infrastructural talent? Some panelists reported that attracting talented people around a university setting is not a problem if inexperienced recent graduates are acceptable. However, after they gain experience at the center, they become very valuable to commercial technology organizations. Thus there is a persistent leakage in knowledge that occurs in a CI center context. The extent to which this leakage affects the role of CI center as steward must be further investigated.

Other issues associated with human resources were also raised during the workshop. First, there is a dearth of education options for all levels of CI centers. On the one hand managers, scientists, and technical people come from areas that do train them in some of the skills necessary for CI centers, but these unique organizational forms are distinctive from the contexts they were trained for. It is important to investigate how different this context is and to what degree alternative modes of education can benefit CI centers.

Further, with respect to CI center personnel, it is important to understand inducements and incentive structures. On the one hand such personnel may be part of the "academic reputation economy," but on the other hand they are outside of this economy. What does this hybrid role place them? Further, many of the technologists in CI centers are firmly rooted in broader software and other IT communities. How does their role in CI centers impact their role in these communities differently?

Finally, different as technologies change and as wider users engage with cyberinfrastructure, different, new skills will be required in CI center contexts. For example, with the move to "big data" and as CI centers move to stewards of data, certain skills such as data curation and data science will be in greater demand. When that is filled, there will be new technological innovations that require new skills. Further, wider users need a more "hands-on" customer-service oriented approach. The service-orientation for providing infrastructural services in the future may be quite different from the past. Human resources practices need to remain dynamic to grow with demand for different forms of talent and it is important to understand how organizational research can help.

Conclusion

> *"A vast opportunity exists for creating new research environments based upon cyberinfrastructure, but there are also real dangers of disappointing results and wasted investment for a variety of reasons including … lack of appreciation of social/cultural barriers, lack of appropriate organizational structures, inadequate related educational activities... The opportunity is enormous, but also enormously complex…" (Atkins et al 2003, p.4)*

This excerpt from the Atkins Report emphasizes the way in which ineffective management practices can waste resources and jeopardize large-scale computational research centers and collaborative environments. The management or such research centers should not be taken lightly – especially since the social and organizational elements may be the "hardest part" of these endeavors (David 2006). While the technical issues are critically important, it is equally important not to neglect the organizational and social elements.

Computational science environments are typically run as projects (Karasti et al 2010), and many of the available prescriptive advice and 'lessons learned' emphasize the need for strong project management practices (e.g., Spencer et al 2006). While strong project management is certainly important, these environments represent novel, lasting organizational forms that go beyond contemporary notions of project organization –the scale of some of the larger projects constitute mid-sized organizations in their own right. Further, these research centers produce a variety of software artifacts and other standards, the impact of which extends far beyond the project at hand (Spencer et al 2006).

Teams who win funding for developing the infrastructure for large-scale computational science (i.e. "cyberinfrastructure" Atkins et al 2003) are not simply managing a single project - they are designing and managing persistent organizations and developing long-lasting artifacts. These teams are involved with multiple projects (both NSF-funded and otherwise) and must reconcile competing resource demands while ensuring continuity across projects. They engage in a variety of activities beyond the design and delivery of research, including human resource management, customer service, quality improvement, strategic planning, accounting, and often software engineering. Most research center management teams, however, have little formal training in organizational management.

Therefore, just as the scholarship into this form of innovative organization is promising, so is the opportunity for valuable, actionable findings to help guide this important type of organization.

References

Adizes, I. (2004) Managing Corporate Lifecycles, The Adizes Institute Publishing, Santa Barbara, CA.

Atkins, D. E., Droegemeier, K. K., Feldman, S. I., Garcia-Molina, H., Klein, M. L., Messerschmitt, D. G., Messina, P., Ostriker, J. P. & Wright, M. H. (2003) Revolutionizing science and engineering through cyberinfrastructure. Report of the National Science Foundation blue-ribbon advisory panel on cyberinfrastructure. Arlington, VA: US National Science Foundation.

Baldwin, C.Y. 2008. "Where do transactions come from? Modularity, transactions, and the boundaries of firms," *Industrial and Corporate Change*, 17(1): 155-195.

Benner, M., & Tushman, M. (2003). Exploitation, exploration, and process management: The productivity dilemma revisited. *Academy of Management Review, 28*(2), 238-256.

Berente, N. (2010) Supporting Successful Design and Management of Research Centers, NSF VOSS project: http://www.nsf.gov/awardsearch/showAward.do?AwardNumber=1059153&WT.z_pims_id=503256

Berente, N. and Howison, J. (2011) Research Collaboration Network for Managing Collaborative Research Centers, NSF VOSS Project:
http://www.nsf.gov/awardsearch/showAward.do?AwardNumber=1148996&WT.z_pims_id=503256

Bowker, G.C., Baker, K., Millerand, F., and Ribes, D. (2010) "Towards Information Infrastructure Studies: Ways of Knowing in a Networked Environment," (in J.Hunsinger, M.Allen, L.Klasrup eds) International Handbook of Internet Research, Springer

Bowker, G.C., and Star, S.L. (1994). Sorting Things Out: Classification and Its Consequences. MIT Press: Cambridge, MA.

Burns, T., & Stalker, G. M. (1961). *Management of Innovation*. Routledge Kegan & Paul.

Burt, R.S. (1995). *Structural Holes: The Social Structure of Competition* (2nd Ed.). Cambridge, MA: Harvard University Press.

Christensen, C. M., Suarez, F. F., Utterback, J. M., (1998) Strategies for survival in fast-changing
industries." Management Science 44, S207-S220.

Cummings, J. and Keisler, S. (2011) "Organization Theory and New Ways of Working in Science," Atlanta Conference on Science and Innovation Policy, 2011

Colfer, L., & Baldwin, C. Y. (2010). *The Mirroring Hypothesis: Theory, Evidence and Exceptions*.

Cyert, R. M., & March, J. G. (1964). *Behavioral Theory of the Firm*. Prentice Hall.

Dabbish, L., Stewart, C., Tsay, J., & Herbsleb, J. D. (2011). Social Coding in GitHub: Transparency and Collaboration in an Open Software Repository. *CSCW*.

David, P.A. (2006) "Towards a cyberinfrastructure for enhanced scientific collaboration: providing its 'soft" foundations may be the hardest part." in D. Foray and B. Kahin (eds), Advancing Knowledge and the Knowledge-Economy, MIT Press, Cambridge, Mass.

Davis, J.H., Schoorman, F.D. and Donaldson, L. (1997). Toward a stewardship theory of management. Academy of Management Review, 22, 1, 20-47.

DeSanctis, G. and Monge, (1999) "Communication Processes for Virtual Organizations," Organization Science, 10(6), 698-703.

Deutsch, J. A., & Deutsch, D. (1963). Attention: Some Theoretical Considerations. Psychological Review, 70(1), 80-90. doi: 10.1037/h0039515

Dicke, L. (2002). Ensuring accountability in human services contracting: can Stewardship Theory fit the bill? *The American Review of Public Administration*, 32, 4, 455-470.

Edwards, P., Jackson, S., Bowker, G. and Knobel, C. (2007). "Report to the NSF of a Workshop on 'History and Theory of Infrastructures: Lessons for new scientific infrastructures'," University of Michigan, School of Information.

Gao, L.S. and Iyer, B. (2006). Analyzing complementarities using software stacks for software industry acquisitions. *Journal of Management Information Systems*, 23, 2, 119-147.

Garud, R. & Kumaraswamy, A. (2005) "Vicious and Virtuous Circles in the Management of Knowledge: The Case of Infosys Technologies," MIS Quarterly, 29(1), 9-33.

Hansen, M. T., & Hass. (2001). Competing for attention in knowledge markets: Electronic document dissemination in a management consulting company. Administrative Science Quarterly.

Hanseth, O., & Lyytinen, K. (2010). Design theory for dynamic complexity in informationinfrastructures: the case of building internet. *Journal of Information Technology*, *25*, 1-19.

Hedberg, B.; Dahlgren, G.; and Olve, N. (2000) Virtual Organizations and beyond: Discover Imaginary Systems, John Wiley & Sons, Inc.

Herbsleb, J. & Howison, J. (2009) Socio-technical ecosystems for scientific software development NSF VOSS Project:
http://www.nsf.gov/awardsearch/showAward.do?AwardNumber=0943168&WT.z_pims_id=503256

Herbsleb, J. D., Muller-Birn, C., & Towne, W. B. (2010). The VistA Ecosystem: Current Status and Future Directions (No. CMU-ISR-10-124). Pittsburgh, PA: Carnegie Mellon University.

Hoegl, M., Gibbert, M., & Mazursky, D. (2008). Financial constraints in innovation projects: When is less more? *Research Policy*.

Hughes, T.P. (1983) Networks of Power: Electrification in Western Society 1990-1930, Johns Hopkins University Press

Jackson, S.J., Ribes, D., Buyuktur, A.G., Bowker, G.C. (2011). “Collaborative rhythm: temporal dissonance and alignment in collaborative scientific work.” Proceedings of CSCW 2011: 245-254.

Karasti, H., Baker, K.S., and Millerand, F. (2010) “Infrastructure Time: Long-term Matters in Collaborative Development,” Computer Supported Cooperative Work, May 30, 2010

Knobel, C.P., and Bowker, G.C. (2011). “Values in Design.” Communications of the ACM. Vol. 54, No. 7: 26-28.

Kuhn, T.S. (1996). The Structure of Scientific Revolutions. University of Chicago Press, Chicago, IL, third edition.

Lee, C., Dourish, P. and Mark, G. (2006): 'The Human Infrastructure of Cyberinfrastructure', In Proceedings of the Conference on Computer Supported Cooperative Work (CSCW '06), ACM Press, New York, NY, pp. 483-492.

Levitt, B. and March, J. (1988) “Organizational Learning,” American Review of Sociology 14:319-40.

March, J. G. (1991). Exploration and Exploitation in Organizational Learning. *Organization Science*, *2*(1), 71-87.

Nicolini, D., Mengis, J., and Swan, J. (2012) “Understanding the Role of Objects in Cross-Disciplinary Collaboration,” Organization Science, forthcoming.

Nohria, N., & Gulati, R. (1996). Is slack good or bad for innovation? *ACADEMY OF MANAGEMENT JOURNAL*, *39*(5), 1245-1264.

Obstfeld, D. (2005). Social networks, the *tertius iungens* orientation, and involvement in innovation. *Administrative Science Quarterly*, 50, 100-130.

Ribes, D. and Finholt, T.A. (2009). “The Long Now of Technology Infrastructure: Articulating Tensions in Development,” Journal of the Association for Information Systems, 10(5): 375-398.

Ribes, D. and Lee, C.P. (2010). Sociotechnical studies of cyberinfrastructure and e-research: current themes and future trajectories. *Computer Supported Cooperative Work*, 19, 231-244.

Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. Psychological Review, 84(1), 1-66. doi: 10.1037/0033-295X.84.1.1

Simon, H. A. (1957). Administrative Behavior, 2nd Edition,. The Free Press.

Spencer Jr, B.; Butler, R.; Ricker, K.; Marcusiu, D.; Finholt, T.; Foster, I. & Kesselman, C. (2006),'Cyberenvironment project management: lessons learned’. Technical report, NSF Grant CMS-0117853.

Spencer, D., Zimmerman, A., and Abramson, D. (2011) “Project Management in E-Science: Challenges and Opportunities,” Computer Supported Cooperative Work, Vol.20, No.3, 155-163.

Star, S., & Ruhleder, K. (1996). Steps toward an ecology of infrastructure: Design and access for large information spaces. *INFORMATION SYSTEMS RESEARCH*, *7*(1), 111-134.

Tilson, D., Lyytinen, K. and Sørensen, C. (2010). Digital infrastructures: the missing IS research agenda. Information Systems Research, 21, 4, 748-759.

Ven, A. H. V. de, Polley, D. E., Garud, R., & Venkataraman, S. (1999). *The Innovation Journey*. Oxford University Press, USA.

Vargo, S.L. and Lusch, R.F. (2004). Evolving to a new dominant logic for marketing. *Journal of Marketing*, 68, 1-17.

Vedres, B. and Stark, D. (2010). Structural folds: generative disruption in overlapping groups. *American Journal of Sociology*, 115, 4, 1150-1190.

Voss, G., Sirdeshmukh, D., & Voss, Z. (2008). The effects of slack resources and environmental threat on product exploration and exploitation. *ACADEMY OF MANAGEMENT JOURNAL*, *51*(1), 147-164.

Yoo, Y., Henfridsson, O. and Lyytinen, K. (2010). The new organizing logic of digital innovation: an agenda for information systems research. *Information Systems Research*, 21, 4, 724-735.

Zimmerman, A. and Finholt, T. A. (2007): 'Growing an Infrastructure: The Role of Gateway Organizations in Cultivating New Communities of Users,' in Proceedings of the International ACM SIGGROUP Conference on Supporting Group Work (GROUP '07), ACM Press, New York, NY.

Appendix - Panelists:

Stan Ahalt, RENCI

Stan Ahalt is the director of the Renaissance Computing Institute (RENCI) and professor of computer science at the University of North Carolina at Chapel Hill. Previously he was director of the Ohio Supercomputer Center (OSC).

Nicholas Berente, University of Georgia

Nick Berente is assistant professor with the University of Georgia's Terry College of Business. His research focuses on the management of scientific research centers and infrastructural innovation.

Jay Boisseau, TACC

Jay Boisseau is the director of the Texas Advanced Computing Center (TACC) and a co-principal investigator for XSEDE. Previously he was associate director at the San Diego Supercomputer Center (SDSC).

Jennifer Claggett, University of Georgia

Jennifer Claggett is a doctoral candidate with the University of Georgia's Terry College of Business. Her research includes work on infrastructural innovation and the diffusion of health care IT innovations.

Debbie Crawford, Drexel University

Debbie Crawford is Senior Vice Provost for Research at Drexel University. Previously she was deputy assistant director for CISE at NSF.

Bill Feiereisen, Intel Corporation

Bill Feiereisen is a Senior Scientist and Corporate Strategist in High Performance Computing at Intel Corporation. He also holds a research appointment in computer science at the University of New Mexico. He has previously served as the Director of High Performance Computing for Lockheed Martin Corporation, the Chief Technologist and Division Director of the Computer and Computational Sciences Division at Los Alamos, and before that as the director of the NASA Advanced Supercomputing Facility at Ames Research Center.

James Howison, University of Texas, Austin

James Howison is assistant professor with the University of Texas at Austin School of Information. His research focuses on the organization of scientific software development.

John Leslie King, University of Michigan

John King is William Warner Bishop Professor in the School of Information at the University of Michigan. Previously he was Dean of the School of Information and Vice Provost for Strategy at the University of Michigan, and Professor and head of Information and Computer Science at UC Irvine.

Laurie Kirsch, University of Pittsburgh

Laurie Kirsch is Senior Associate Dean and Professor of Business Administration at the University of Pittsburgh Katz School of Business. Her research focuses on the governance of complex distributed projects.

Cory Knobel, University of California, Irvine
Cory Knobel is co-director of the Values in Design Laboratory and assistant adjunct professor at the Department of Informatics at the Donald Bren School of Information and Computer Sciences, University of California, Irvine. His research focuses on cyberinfrastructure, cyberscholarship, values in sociotechnical design, and emerging forms of scholarly expression.

Patricia Kovatch, Mt. Sinai
Patricia Kovatch is the Associate Dean for Scientific Computing at the Mt. Sinai Medical School. She is the former director of the National Institute for Computational Sciences (NICS) and has also worked at the San Diego Supercomputer Center (SDSC).

David Moses, PSC
David Moses is executive director of the Pittsburgh Supercomputer Center (PSC). He is the former chief operating officer and co-founder of Gaussain, Inc.

Nicole Radziwill, James Mason University
Nicole Radziwill is assistant professor in the Department of Integrated Science and Technology at James Madison University. Previously she was assistant director at the National Radio Astronomy Observatory headquarters. She has served multiple times as an examiner for the Malcolm Baldridge National Quality Award, is Past Chair of the American Society for Quality (ASQ) Sofware Division, and was recently named one of ASQ's New Voices of Quality for the upcoming decade.

Joe Rubleske, University of Georgia
Joe Rubleske is a postdoctoral researcher with the University of Georgia's Terry College of Business. His research focuses on practices of organizing around service and digital infrastructure innovation.

Barry Schneider, NSF
Barry Schneider is program director for NSF Office of Cyberinfrastructure (OCI) TeraGrid/XD (XSEDE) program. Before coming to OCI he served as Program Director for Theoretical Physics in the Physics Division of NSF. Barry spent twenty years in the Theoretical Division of Los Alamos National Laboratory before coming to the NSF in 1991.

Sandra Slaughter, Georgia Tech
Sandra Slaughter is professor and the Alton M. Costley Chair in Information Technology Management at the Georgia Tech College of Management. Her research focuses on open source software development, software process improvement, and cyberinfrastructure.

Phil Westmoreland, ICSE
Phil Westmoreland is executive director of the Institute for Computational Science & Engineering (ICSE) and professor of Chemical & Biomolecular Engineering at North Carolina State University. Previously he served as program director in NSF/ENG/CBET, co-leading CBET's Cyberinfrastructure activities.

Susan Winter, NSF
Susan Winter is acting deputy office director for the NSF Office of Cyberinfrastructure (OCI) and lead program director for OCI's Virtual Organizations as Sociotechnical Systems (VOSS) program.