Foregrounding the Cyberinfrastructure Center as Cyberinfrastructure Steward

Joe Rubleske  
Department of MIS  
Terry College of Business  
University of Georgia  
rubleske@uga.edu

Nicholas Berente  
Department of MIS  
Terry College of Business  
University of Georgia  
berente@uga.edu

ABSTRACT
A recent uptick in studies of cyberinfrastructure-enabled research projects has advanced our limited understanding of the use, development and management of cyberinfrastructure (CI) resources. Conspicuously absent from most of these studies, though, is any consideration of the key roles played by supercomputer centers, or what we term ‘CI centers’, in CI-enabled research networks. This position paper begins to fill this sizable gap by identifying the three major functions served by CI centers and by framing the CI center as a potential steward of CI.

Author Keywords
Cyberinfrastructure; cyberinfrastructure centers; stewardship theory; cyberinfrastructure stewardship

ACM Classification Keywords
H.4.2; K.6.1

General Terms
Human factors; management; theory

INTRODUCTION
As cyberinfrastructures become more integral to scientific discovery across a growing number of domains, demand for cyberinfrastructure (CI) resources increases. Accordingly, knowledge about how CI resources are effectively used, developed and managed is crucial to the success of CI’s mission. While a growing number of studies of CI-enabled research projects have advanced our knowledge in this regard, very few CI studies have focused on or even considered the supercomputing center, or what we term the ‘CI center’. This gap is conspicuous given the important roles played by CI centers in CI-enabled research networks.

This position paper is organized as follows. After making the case that the CI center has been overlooked in CI studies, we identify (in the second section) the three major functions served by CI centers. In the third section we introduce stewardship theory [1] and then, in the fourth section, identify five ways in which a CI center can act as a CI steward. The paper concludes by outlining future research needs.

This is not to say that studies of CI-enabled research projects have not advanced our knowledge. Indeed, a great deal has been learned from them, including insights on:

- Overcoming challenges related to sharing data and knowledge across domain boundaries [4];
- Balancing the tension between the need for sustainable CI and the short-term nature of CI funding [3, 5];
- Facilitating data reusability [17];
- Recognizing the divergence of temporal orientations across project roles [2];
- Balancing various tensions intrinsic to CI growth [6];
- Explicating the role of the project manager in promoting the quality of interactions between project members [7]; and
- Developing domain-specific CI resources [8].

---

1 For our purposes, a research network comprises multiple research centers (or sites) and is organized around a common research problem requiring high-performance computing and shared digital resources. The Long Term Ecological Research (LTER) Network, which comprises 26 separate research sites, is one example of a research network. Research networks have also been referred to as collaboratories, science gateways and virtual organizations.

2 One notable exception is Kee and Browning [9], which included CI centers in its study of funding for CI projects.
For the most part, these studies examine one or more CI-enabled research networks (as cases) but do not consider the CI center; where it is considered, it plays at most a background part. Accordingly, we cannot help but wonder whether the findings from these studies would have differed had the CI center been included.

FUNCTIONS OF THE CYBERINFRASTRUCTURE CENTERS

Given that cyberinfrastructure (CI) centers have largely been overlooked in CI studies, much of what we know about them comes from reports produced on behalf of the National Science Foundation (NSF) and the Coalition for Academic Scientific Computation (CASC). In one widely referenced report, for example, an NSF advisory panel identified some of the major resources managed by CI centers:

High-performance computers, large data archives, sophisticated visualization systems, collaboration services, licensed application packages, software libraries, digital libraries, high-speed connections to a national research backbone, and a cadre of skilled support personnel helping users take advantage of the facilities. [10, p. 74]

At a broad level, CI centers serve three functions. The first function, and the one with which they are most often identified, is to provide infrastructure-as-a-service (IaaS). IaaS includes automated services such as job scheduling, algorithm execution, data visualization and data storage. The delivery of IaaS depends on the effective implementation and maintenance of IaaS resources (e.g., high-performance processors, networking and storage hardware, firmware, systems software).

The second major function served by CI centers is to provide support and training to users. CI centers routinely help users secure an account, navigate file spaces, prepare and submit job files, and manage data from executed jobs. Training workshops and more personalized support may be offered to users having difficulty parallelizing code and building applications that enable the sharing (with project and/or network members) of instruments, data, software and knowledge.

The third and final major function served by CI centers is to develop new CI resources. In some CI centers, such as the Pittsburgh Supercomputing Center (PSC), there is a dedicated organizational unit for this function (e.g., PSC’s Advanced Systems Group). While CI centers have traditionally focused their development efforts on the bottom layers of the CI ‘stack’ (i.e., hardware, systems software and middleware), a growing portion of development efforts are being directed at the facilitation and support of CI resource sharing [9]. This trend is due in large part to two related factors. First, NSF has made explicit its goal of promoting CI adoption by scientists from domains (such as computational linguistics and economics) that historically have not made use of CI [11, 12]. As Stewart et al. [13, p. 3] have noted, achieving this goal depends on the development of tools that “allow scientists to use computation at scale via intuitive Web interfaces, without requiring scientists to first become computational experts.” Second, while advances in CI hardware have mostly been incremental over this period, advances in application and portal software for CI have been more exceptional [14, 15]. Thus, by paying more attention to software development, CI centers realize greater returns on their efforts to grow and improve CI.

STEWARDSHIP THEORY

Stewardship theory was introduced more than 20 years ago in response to agency theory's assumptions about the motivations that drive an agent's decision making [16]. More specifically, agency theory assumes that the decisions made by agents – who are hired by a firm's principals to manage the firm – will invariably serve their own interests, even in cases where the decision reduces principals' utility. Drawing from criticisms of agency theory, Donaldson and Davis [16] argued that some agents derive more utility from collectivistic behavior. Accordingly, the decisions made by these agents (stewards) will always serve the interests of the firm's principals, even if the decision reduces the agent's utility in the short term.

Davis et al. [1] argued that this conception of organizational stewardship was flawed in its assumption that an agent is either opportunistic or steward-like in all situations. For Davis et al., an agent's decision to act as a steward (or not) in any given situation is a function of certain psychological and situational factors. For example, an agent who derives relatively more utility from extrinsic rewards – and is thus a personal opportunist according to agency theory – may still act as an organizational steward where management philosophy emphasizes “involvement” (versus control) and/or when the organization's steward is characterized by a “low power distance” (i.e., non-acceptance by less powerful members of privileges for more powerful members).

While stewardship theory explains behavior at the individual level, we believe that organizations can also behave as stewards (and/or opportunists). With respect to CI centers, we propose that CI centers, as agents charged with managing CI, will strive to act as stewards (and not as opportunists) in most situations. Given the dearth of knowledge about CI centers, we defer the task of proposing the conditions under which CI centers will not act as a steward until we have analyzed interview data that we recently collected. In the meantime we draw from the CI
literature to determine the needs of CI’s principals – namely, researchers organized into project teams – in order to identify some of the ways in which CI centers can act as stewards of CI.

**Dimensions of Cyberinfrastructure Stewardship**

In what ways can cyberinfrastructure (CI) centers act as CI stewards? What, exactly, does CI stewardship entail? In order to answer these questions, we make two assumptions. First, we assume that CI centers are agents charged with effectively managing CI. Second, we assume that CI management is undertaken to help achieve the superseding goal of scientific discovery (or knowledge creation). Thus, for our purposes, project-based researchers – who are responsible for goal achievement – are the principals whose interests are served by CI centers.

Accordingly, we have drawn from studies of CI-enabled research projects to identify five needs of principals (i.e., project-based researchers) that – given the functions currently served by CI centers (see above) – could conceivably be met through the efforts of CI centers.

**Data preservation.** The NSF has called for new resources aimed at managing and exploiting the ‘data tsunami’ [12]. While it is typically the research community that develops metadata standards and contextualizes and describes data [2], CI centers possess the expertise to develop new tools for preserving scientific data in perpetuity.

**Knowledge sharing across boundaries.** Pennington [4] has articulated the urgent need for new approaches to sharing knowledge across projects and domains. Given that CI centers work closely on numerous projects across multiple domains and possess extensive CI “know-how,” they are well positioned in this regard.

**CI service quality improvement.** While most CI users are generally satisfied with the quality of CI services [11], areas for improvement can always be found. For example, Stewart et al. [13] recently noted that “approval of larger requests [for the services of a certain CI] can easily take a quarter of a year or longer.” CI centers are counted on to address these concerns.

**Adaptation to change while maintaining stability.** Ribes and Finholt [18, p. 376] recently asked, “How can the perseverance of the infrastructure project be ensured in the face of changing technologies, emerging standards, and uncertain trajectories?” The CI center has a more comprehensive view of CI than any other CI stakeholder, and should thus play a major role in resolving this problem. One possibility is to build extensibility features into the CI stack, where extensibility is a design principle that takes future growth into consideration [3].

**CI growth.** The emergent and unpredictable nature of CI growth [19] makes it “impossible to predict perfectly in advance what the infrastructure will be or how it will be used” [8, p. 246]. Nevertheless, CI centers can promote CI growth in a number of ways, including (but not limited to) the maintenance of a stable platform and the implementation of modular components in the CI stack [20].

**FUTURE RESEARCH NEEDS**

The conceptualization of the CI-enabled research network and all its parts – the CI center included – is a large, ongoing project. This paper takes a first step toward integrating the overlooked CI center into this project. That being said, three research needs follow from this paper. First, and as noted, there is a need to propose the conditions under which CI centers will not act as stewards. Which factors – organizational, network-based or other – are salient? Second, our assumptions about project-based researchers as principals should be reconsidered. Third, and finally, there is a need to extend our initial list of principals’ needs.

**ACKNOWLEDGMENTS**

We thank and gratefully acknowledge the Office of Cyberinfrastructure at the National Science Foundation for the grant (#OCI-1059153) that supported this study.

**REFERENCES**


