Human-Centered e-Science: A Group-Theoretic Perspective on Cyberinfrastructure Design

Tyler Pace, Shaowen Bardzell, Geoffrey Fox
School of Informatics and Computing, Indiana University
tympace, selu, gcf @indiana.edu

ABSTRACT

Cyberinfrastructure (CI) is a rapidly growing area of global research and funding with a history for emphasizing the role technology will play in changing scientific work practices. This paper proposes a group-theoretic perspective on small group collaboration that is informative to CI research and the design of CI environments. To illustrate the relevancy of a group-theoretic perspective to CI, this paper presents a critical review of over 150 CI research papers published in the last decade through the lens of McGrath's seminal Time, Interaction and Performance theory of groups. After relating common CI research themes through a focus on groups, we propose a series of early design principles aimed at helping interaction designers and technologists to better design and implement group collaboration tools in CI applications.

KEYWORDS: cyberinfrastructure, design, group theory, HCI, CSCW

1. INTRODUCTION

Cyberinfrastructure (CI) is a rapidly growing area of global research and funding [5]. Essentially, CI refers to the "infrastructure of distributed computer, information, and communication technologies" [24] supporting a transformation in the sciences towards large scale, collaborative data driven enterprises. Recent vision statements of CI highlight the potential for CI to serve as a "harbinger of a broader impact … on the conduct of knowledge-based activities" [4]. Furthermore, CI based knowledge communities "offer the potential for a new wave of global-scale collaboration across multiple disciplines, geography, and institutions" that serve to "empower a revolution in what science explores, how it is done, and who participates" [4].

As noted in the earliest vision statements, collaboration is at the heart of CI [5]. The many complexities of large scale data sharing and the growing mandate to design and build more usable, accessible and diverse CI tools are subjugated to the overarching goal of enabling, promoting and increasing effective collaboration in the sciences. At its core, scientific collaboration is primarily a social and organizational issue rather than a technological one [11, 26, 35]. Therefore, research in CI is in no small part concerned with the social and organizational issues of scientific collaboration and the ways in which CI tools may alleviate, burden or otherwise influence the work of scientists.

Lee et al [26] recently explored the "human infrastructure" of cyberinfrastructure to reveal how "human and organizational structures share properties with technological infrastructures." For Lee et al, human infrastructure constitutes the "people, organizations, networks" and arrangements that constitute them. This is primarily an institutional and organizational view of the human side of cyberinfrastructure reflecting the "synergistic collaboration of hundreds of researchers, programmers, software developers, tool builders and others" who develop "applications and software for complex, distributed, and dynamic" CI environments. As such, the primary emphasis of Lee et al is on the broad relations between institutional and organizational actors spanning many groups working on many projects and the human infrastructure (e.g., policies, standards, norms, power structures, tools) that connects them. Informed by the work of Lee et al, our work focuses on small group interaction and how it relates back to the larger human infrastructures of CI. The work presented in this paper reviews the breadth of CI research and reveals both the centrality of groups and the utility of group theory in understanding how groups operate within the larger context of CI's human infrastructure [26]. We use McGrath's seminal theory of groups [2, 28] as a critical lens for analyzing a large body of CI research through the lens of small group collaboration.

1.1. McGrath's Theory of Groups

McGrath developed the Time, Interaction and Performance (TIP) theory of groups in the 1980-1990s after decades of social psychology research on the nature of group work [27, 28]. The TIP theory has been well received in studies of computer supported cooperative work (CSCW) [23] and continues to hold a prominent position in multi-disciplinary group research nearly 20 years after its inception [2].

In the broadest sense, TIP theory is about groups and what they do, but with special emphasis given to the group and member interactions that facilitate group performance. TIP theory attempts to conceptualize group and group activity at a "level of molarity and complexity that reflects, to some degree, the nature of groups in everyday life" [28]. Accordingly, TIP theory is concerned with the natural complexity of groups, small and large, homogenous and intermingled, ad hoc and permanent. Additionally, TIP was one of the earliest group theories to highlight the growing role of computer-mediated communications (CMC) in the management and performance of group tasks [28]. The emphasis on complex professional work groups and computer-mediated communication in TIP theory make it particularly well suited for analyzing CI research.

Our use of TIP theory in this paper is not primarily meant to serve as a vehicle for importing the work of McGrath into the research on CI. Instead, we believe that applying TIP theory to CI research will help us to develop a group centric perspective on the both the history and future of CI research. This group oriented perspective of CI research is intended to be the major contribution of this paper. In order to develop this group perspective of CI research, we will use a handful of the central constructs of TIP theory to illustrate the nature of CI research conducted over the last 10 years. Through the lens of TIP theory, we will present a group or "human infrastructure" [26] centric view of CI research broadly conducted within the human-computer interaction (HCI), CSCW and computer engineering domains. By positioning the history of CI research as fundamentally concerned about groups, we aim to shed further light on the centrality of the social and the human in our investigations of the advanced technological systems of CI.

2. METHOD

The study presented in this paper represents the culmination of a six month review of over 150 academic books, articles, and technical reports on cyberinfrastructure and e-science published since the year 2000. Articles were pulled from conferences and journals in HCI (e.g., ACM CHI, ACM GROUP), CSCW (e.g., ACM CSCW, ECSCW, JCSCW), computer-mediated communication (e.g., JCMC), and computer engineering (e.g., IEEE E-Science). It is not our goal to fully encapsulate the last 10 years of CI research from these multiple, related domains (such a task would be foolish in the space provided), but rather we present a cross-sectional representation of the common CI research questions and outcomes of the last decade as seen from a group theory point of view. Similar approaches to conducting cross-sectional literature reviews have been used to great success within HCI [9].

This work does not represent an externally validated content analysis, taxonomy or other type of categorization of the CI literature. Rather, we are using an interpretive and critical approach to analyze the literature based on constructs from TIP theory in addition to utilizing the expertise of the researchers in the CI, CSCW and HCI domains. Bos et al [13] used a similar approach for evaluating CI research when developing a taxonomy of CI collaboratories. While the contributions of Bos et al's taxonomy of collaboratories has been invaluable to our understanding of the breadth of CI in practice, we aim to complement their technological and functionally oriented taxonomy of the types of CI with a group focused analysis of CI research that reveals less about the different types of CI projects and more about the ways in which CI is influenced by groups and vice versa.

3. A GROUP CENTRIC PERSPECTIVE ON CYBERINFRASTRUCTURE RESEARCH

In the following sub-sections, we offer a synthetic perspective on the group centric nature of CI research via a partial lens of TIP theory. We emphasize the use of TIP theory's group "functions" to analyze our body of CI research. Functions in TIP theory are the overarching categorizations of tasks undertaken by all groups in order to facilitate positive interactions. The three group functions in TIP theory are *production*, *group well-being*, and *member-support* [28].

3.1. Managing Group Production

The *production* function of TIP theory concerns itself with the relation between groups as functional units and the environment (technological and organizational) within which the groups operate [28]. Major tasks of the production function include the initial choice of the project and project goals, technical problem-solving (choice of the most appropriates means to carry out the project), and political preference-resolving (choice of policies to resolve potential conflicts of value and interest). Our review of the CI literature through the lens of the production function reveals recurring questions surrounding the choice of an appropriate CI *environment* (project choice), obtaining and maintaining sufficient *access* to the CI environment (technical problem-solving), and resolving *data management* conflicts (political preference-solving) within the CI environment.

As interest in CI projects and research continues to climb, scientists will have increasing opportunity to choose from multiple CI environments. The choice of a technological and organizational environment is fundamental to the success of collaborative work because every environment forecloses some group interactions while promoting others. In terms of CI research, there is a growing mandate for the use of open, accessible and modifiable CI environments rather than the continued development of closed technologies [3, 5, 10]. As such, a key research area in CI is the development, distribution and promotion of open middleware technologies and layered architectures that support the rapid building of tailored (i.e., group or community specific) CI environments that better support the varying specificity of work practices among scientific work groups [4, 10, 22].

In addition to choosing the right CI environment, groups must obtain and maintain a sufficient level of access to the environment. The issue of access in CI research includes concerns surrounding the ways in which CI tools facilitate communication with and among users when they are not actively engaged with the CI environment [6, 7, 10, 16, 17, 18]. Common technologies such as email, RSS and instant messaging have all been used to facilitate the outward communicative capacity of CI applications [16]. These technologies all enable groups to maintain an appropriate level of access to the CI environment when they are operating within other, non-CI contexts during the course of their work.

The technical-problem solving task of the production function suggests that groups benefit from a wide range of choice in terms of the technological and communication tools used to fulfill project goals. Research into communication needs of CI users strengthens this claim, showing the need to support diverse technological and communicative needs of groups operating within CI environments [23, 26, 38]. As such, CI tools are developed with a range of synchronous and asynchronous communication and collaboration tools to match the needs of groups of various sizes and types [23, 26]. However, balancing the type and amount of telecommunication services offered is especially difficult in the area of scientific collaboration [38].

Early organizational and collaboration research identifies the value of informal, direct, two-way communication when dealing with the complex and uncertain environment of scientific work [38]. However, recent work demonstrates that synchronous communication tools sometimes requires such high coordination costs that it may result in a loss of collaboration effectiveness [38]. As such, there is growing interest in developing communication technologies for CI environments that provide a range of technical-problem solving options including automated updates of work progress, off-line processing of work, and the ability to answer quick questions without demanding immediate responses [23, 38].

The management (i.e., collection, storage, processing, sharing) of vast amounts of data is fundamental to the development of CI [4, 5]. However, early CI reports warn of the difficulty of resolving the political and disciplinary interests surrounding data management in the sciences (a political-preference resolving task). Accordingly, research into the technical and organizational solutions of the management of data in CI environments tends to fall into two broad categories. First, there are many concerns about how to meaningfully collect, share and analyze data in situations ranging from small work groups to multi-disciplinary and multi-institutional collaborations. Second, CI environments reveal or reinforce the fundamental role data plays in shaping scientific communities and practice.

Data management in CI environments presents a range of group work issues. Large CI projects often require the ability to store, annotate and share large amounts of multimodal data both within and beyond the original research team [1]. As one of the major proposed benefits of CI tools is collaboration within and across disciplines, data management in CI environments necessitates supporting several data models, re-analysis of data under differing assumptions, validation of many computational models, and the exploration of multiple measures of validity and analysis [31]. Furthermore, when sharing data across large distributed teams, there arise issues with how to develop and store standardized annotations (without sacrificing individual researcher ability) [1] and how to develop CI environments that support separate collaborative and individual perspectives on stored data [19].

While sharing large amounts of data creates issues of data integrity and storage, perhaps the most commonly reported issue in regards to data sharing is the problem of metadata. Creating and maintaining data for personal use is very different from creating data for group use due to the need to share not only data but context [1, 17]. Without context, data is practically meaningless [7, 29].

As individuals, scientists develop a tacit understanding of their data [15]. However, when asked to share their data with diverse and remote colleagues it becomes a significant collaboration and technical challenge to determine what contextual information is required to maintain data value [1]. Additionally, sharing data at larger institutional and organizational levels entails the policy problems of determining and attributing intellectual property [14] as well as abiding by informed consent and other ethical guidelines for anonymization of research data collected from human subjects [15, 29].

The production function brings forth questions of how groups actively position themselves with regard to the technological and organizational environments within which they operate. CI research highlights issues of environmental posturing through repeated concerns of how and why groups choose CI environments, the means by which they maintain sufficient access to the environment, and the myriad issues of managing data in remote multi-disciplinary collaborations. Overall, our review brings to bare the focal ways in which groups position themselves relative to the opportunities and demands of different CI environments.

3.2. Facilitating Group Well-being

The *group well-being* function of TIP theory describes activities that deal with the development and maintenance of the group as a system. Major tasks of the well-being function include managing relations among group members, carrying out interpersonal activities involved in the completion of group work, and defining member roles within the group [28]. Reviewing the CI literature through the lens of the well-being function highlights issues of *disciplinarity* (managing relations), *trust* in the CI environment (carrying out interpersonal activities), and the development of *context* and *awareness* within work groups (defining member roles).

Scientific work is deeply rooted in the epistemological and ontological practices of its many and varied disciplines. However, a shared interest in the potential of CI environments to transform or elevate scientific work brings scientists together from many disciplines. Incorporating the varied work practices of scientists from many disciplines into functional groups entails the development of CI environments that are mindful of differing disciplinary goals [8] and requirements [36].

Virtual and conceptual organizations are commonly presented as codified structures designed to overcome disciplinary barriers in CI environments [3, 10, 12, 26, 33, 35]. The ideal qualities of a virtual organization include the following:

- Ability to facilitate collaboration across disciplines and institutions [35].
- Enable frequent data and resource sharing [3, 10].
- Embrace fluid command structures that support rapid formation and dissolution of temporary and long term task-driven working groups [26].
- Establish clear lines of group and role membership [26].
- Permit flexible, dynamic workflows and scheduling across disciplines with differing work styles and practices [33].
- Enable group structures that espouse the roles and responsibilities of members across political, disciplinary and academic boundaries [33].

However, despite the lofty goals and early success of some virtual organizations, recent research into virtual organizations reveals that while the promise and potential of virtual organizations continues to predicate their use, there is in fact little understanding of the complex social and political relationships required to build and operate successful virtual organizations and CI environments [12, 26].

For many virtual organizations and CI environments, trust in the organization and the technology is a major roadblock to the interpersonal activities that support group well-being [17, 25, 29, 34]. A significant issue of trust in CI environments is determining who or what is considered to be credible in an environment so richly composed of human, organizational, and technological actors [25]. The development of a “common ground” (chiefly, context and awareness) fosters trust in scientific work with vast cultural and professional differences [34].

Common ground is an established prerequisite for trust in scientific work [34]. Developing common ground entails issues of context and awareness that underpin the ability of groups to successfully define and execute their roles within CI environments. Unfortunately, the specification of context of work creates problems for remote collaboration due to the difficulties of fully specifying cultural norms and intricate task interdependence. Understanding context is more central to the issue of remote collaboration than the disciplinary status of group members [38]. In other words, group members of the same discipline with similar professional backgrounds readily and often fail to develop a context of work in collaboration.

The development of context and awareness in remote scientific collaboration is promoted through both the information sharing characteristics of groups as well as certain technological features of CI environments [35]. Groups that are adept at sharing task and process information (e.g., info about current and relevant tasks and work processes) and socio-emotional information (e.g., interpersonal information about collaborators) are more likely to establish a shared context of work. Technological features that facilitate awareness are control (e.g., modifiability of the CI environment), sensory richness (e.g., multimodal presentation, degree of perception), level of distraction (e.g., extent of isolation) and the overall realism (e.g., consistency with the "objective" world) of the CI environment [35].

The group well-being function emphasizes the importance of maintaining and supporting the group itself as a functional unit. CI research embodies the importance of the whole group as an actor in CI environments through its continued emphasis on negotiating disciplinary conflicts, developing fluid organizational structures, promoting trust in social and technological agents, and discovering novel ways of enabling shared context and situational awareness in remote collaboration. While these issues have been popular in CSCW research for many years, the new contexts and sheer scope of CI environments offer new and unique opportunities for researchers to reexamine the important role played by groups as groups (rather than groups as a summation of individual actors) play in creating engaging, rewarding and productive collaborative experiences via technology.

3.3. Supporting Group Members

The *member-support* function of TIP theory highlights activities that consider the ways in which the individuals are embedded within a group, thereby reflecting the relations between individual members and the group. Major tasks of the member support function include the assignment of policies concerning member advancement, the individual's participation in group activities, and negotiation of the individual's expected contribution to and payoff from the group [28]. Reviewing the CI literature through the lens of the member support function brings forth issues of the value of the member's role with respect to the group (policies for member advancement), professional identity (individual participation in the group), and motivation/incentives for the use of the CI environment (negotiation of the member's contribution and payoff).

It is well established that data is the foundation of scientific collaboration [7, 11, 12]. As such, members of scientific disciplines orient themselves and their policies for professional advancement around data itself. CI environments are bringing a new level of data management to the sciences and have a major impact on how scientists determine value and participation within work groups [11, 12]. However, data is not created or treated equally across scientific disciplines. In order to develop policies (explicit and implicit) for valuing data, it is critical to understand how data itself is constructed and used in different disciplines and work groups [7, 15]. In scientific communities, data can serve as a point of differentiation between sub-disciplines (e.g., between theoretical and experimental physics), as a point of entry for communities of practice, and as a means for obtaining and maintaining power and status [7].

The ability to access data is a critical point of entry into a scientific community [7, 11, 12]. In some fields, like earthquake engineering, data is generated in local laboratories as a result of conducting experiments. However, in space physics, scientists rely on data from remotely located instruments, other scientists, or public sources. Obtaining access to data not only provides access to a scientific community, but it also positions a scientist within the socio-political hierarchy of their discipline [7, 15]. For many disciplines, having one's "own" data is "better" than relying on public or borrowed data. Access to data (whether creating it or gathering it from a shared instrument) entails a certain amount of political and/or financial resource. Epistemological, political and pedagogical (i.e., training of graduate students and their use of data) practices all come to bear on how data is used and interpreted in the sciences. However, CI research often reduces the question of data to issues of storage or representation thereby neglecting the complex and sometimes volatile social practices surrounding data creation and use in the sciences [7, 12, 15]. Data is fundamental to the means by which scientists value CI environments, relate themselves to their peers and a broader discipline. Developing a better understanding of how scientists relate to data and educating scientists on the value of CI environments is crucial to the issue of developing or maintaining professional identity in CI environments.

Scientist's professional identity is regularly challenged in professional collaborations and when confronted with new and foreign technologies like CI environments. Providing education to scientists about the benefits of using CI environments can help to ameliorate concerns over diminishing their professional identity in the face of rapid CI growth [8]. In order to overcome the barriers to the adoption of CI environments, we must consider the technical, social, and identity challenges faced by scientists as they move to new tools, groups and organizational structures in the future [11, 12, 17].

The growth of a scientist's professional identity through the use of a CI environment is in part predicated on the successful negotiation of the scientist's contribution to and payoff from the CI application. Incentives are becoming common place in CI environments as a means to codify contribution and payoff negotiations. Any incentive system should include all potential users and not just those with cutting edge research interests [3, 10]. As many scientists are only peripherally aware of the potential benefits of CI applications, the burden remains on CI developers and virtual organizations to educate potential users of the value of CI environments [8, 32]. Frequent approaches to encouraging CI use include developing CI based tools that mimic the functionality of pre-existing tools [17], carefully designing for disciplinary goals [8], and providing examples of how CI environments can enhance current work practices and abilities [17, 20].

The member-support function reveals the myriad ways in which individuals as politically situated and socially entrained actors negotiate their position, role, contribution and payoff from professional collaborations. Research in CI reveals the frequency of these concerns within the context of collaboration via CI environments. CI researchers routinely note the frequent interpersonal negotiations of the value of the member's role within the group, issues of maintaining and performing professional identity, and offering sufficient political, intellectual or social incentives for using CI environments.

4. IMPLICATIONS FOR DESIGN

Our group centric review of CI research highlights the equal, if not fundamental, role of group structure and operation in facilitating the success of CI research and projects. Given the centrality of groups to CI, we developed a set of group oriented implications for design based on our group centric perspective on CI environments. The following design implications are offered to inspire and inform designers and developers of CI environments to consider the complex ways group determine what constitutes their environment, use the environment to support the interactions of the group, and how individual members of a group appropriate the environment as part of their socio-political positioning within the group.

4.1. Improve the Usability of CI Environments

The growth of CI environments offers a great opportunity to explore new realms of application and system design (e.g., grid computing). Researchers rapidly create, merge, and share new tools as more disciplines and scientists explore the possibilities for CI in their own work. However, in the fervor to develop CI tools, many applications are produced with limited or no usability testing and virtually no commitment to a user centered design process [3, 10, 39].

As noted in a key report on the potential and pitfalls of virtual research communities and e-science in the UK [10], one of the biggest challenges in the adoption of CI "will remain ease of use … and poor human-computer interfaces." The report highlights that attempts to "bolt on" usability after technical development are unlikely to be successful. However, preferencing technical over "human factor" development is the "current norm … in this field" which will require a "paradigm shift in attitudes" and the employment of "user-focused design methodologies if there is to be a breakthrough in this area." This report reinforces the difficult decisions groups face when choosing, developing and operating within CI environments. Unusable CI environments are likely to become undesirable thereby diminishing the willingness of collaborators to choose the environment for their work. Moreover, unusable CI environments present considerable problems when users need to obtain and maintain access to their CI environment. In short, unusable CI presents some of the most immediate and significant barriers to entry for the selection or adoption of CI environments.

4.2. Enable External Access to CI Environments

The CI environment is but one of many tools used by scientists. Developing CI environments that support external access (e.g., access to CI resources outside of the primary application channel) facilitate and improve the use of CI environments [7, 10]. Additionally, offering multiples means for system notifications to be received by users is highly desired [18]. Using common communication technologies such as IM, email and RSS promotes more user integration with and access to the CI environment [16].

Farooq et al [18] provide a recent example of designing notification systems for the CiteSeer scientific collaboration platform. Based on a survey of CiteSeer users, Farooq et al determined that lightweight and flexible notification systems help alleviate the problem of limited attention in scientific work. Scientists are highly focused on work related to their primary research interests and can only find limited time for general scientific "awareness activities" such as tracking citations or finding new papers [18]. Notification systems supported by CI platforms can help improve a scientist's awareness of a broader range of scientific activities while also permitting more flexible and appropriate means for obtaining and maintaining access to the CI environment, especially when not actively engaged in work conducted in or through the CI application. Systems that enable external access to the CI environment are an important technological step in permitting groups to manage access to their CI resources.

4.3. Support Telecommunication Diversity within CI Environments

For all the benefits in computational speed and massive data analysis afforded by CI, the original vision guiding the development and funding of CI environments is that of global multi-disciplinary scientific collaboration [5]. Research into the telecommunication needs of scientists engaged in remote collaboration reveals that providing a wide range of synchronous and asynchronous communication and collaboration tools improves the likelihood of successful collaboration and involvement with the CI environment [23, 26, 38]. Telecommunication tools that support scientific collaboration are quite varied and include tools for identifying peers and time scheduling as well as (virtual) spaces for work and social interaction. A sufficient diversity in telecommunication services is more likely to meet the varying collaboration needs of the many types of working groups that use any given CI environment [26].

Supporting a diversity of telecommunication services in a CI environment addresses several of the group theory based research issues previously identified in this paper. Telecommunication diversity is intrinsically related to obtaining and maintaining access to a CI environment. Without sufficient means and opportunity to communicate, collaboration is stifled. Additionally, offering an array of telecommunication services may assist potential CI users when choosing or constructing their CI environments. Finally, a variety of telecommunication services fosters the development of context and awareness as different telecommunication technologies each embody their own strengths and weaknesses with regard to the transmission of contextual information.

4.4. Provide Open Ended Data and Metadata Structures

The success and value of many CI environments depends on the ability to collect and share data in ways that are both highly transportable and disseminative [1, 11, 15]. However, current efforts to create metadata structures that meet the demanding and sometimes conflicting goals of data transportation and dissemination have fallen short [15, 29]. As such, there is still a need to further explore the social practices around data use, to move beyond our reliance on metadata, and to improve our understanding of data abstractions more broadly [7].

As previously discussed in this paper, data and their abstractions are the crux of both scientific work and communities. Designers have a responsibility to consider the ways that CI environments interact with and promote certain types of data over others. Recent research into data use in CI highlights a preference for already digital quantitative data (e.g., output from a computerized lab instrument) over non-digital qualitative data (e.g., a physical sculpture in a university museum) [15]. Data types that lend themselves to technically and financially inexpensive incorporation into CI tools may become the preferred types of data on the grounds of usability rather than intellectual desirability [15]. Human collaborators and not technologies should remain the primary agents in the socio-political process of reconciling the complex issues of data management in CI environments.

4.5. Facilitate Social Entrainment During CI Adoption

Research into successful CI collaborations has determined that for collaboration to be successful collaborators must be ready to use the infrastructure [34]. The issue of collaborator readiness is a problem with the entrainment process of CI applications and virtual organizations. Without proper training and motivation of group members, many attempts to utilize CI end in failure [3, 8, 10, 34]. Developing adequate educational and motivational structures may help new CI users to overcome the technical, social and organizational challenges faced by adopters of CI [17].

In terms of a group perspective on CI, issues of education and motivation of new members (users) are, in part, issues of social entrainment. Social entrainment refers to the ways in which group members synchronize their behavioral processes (such as the use of CI environments) to facilitate positive group interaction and performance [28]. Entrainment can be fostered through the use of external signals (e.g., educational materials, political incentives). In CI environments, these signals might include education about the value of CI work (e.g., profiling successful uses of the CI environments from other groups or projects) or providing social incentives for participating in support processes (e.g., community FAQs, developer chats, public documentation).

4.6. Support the Development and Sharing of Multiple Contexts and Hierarchies of Work

The necessity of supporting a shared context and awareness in collaboration is a well established point of CI research [3, 4, 10, 33, 35, 38]. However, the turn towards a group centric view of CI environments presented in this paper presents the need to support and develop for multiple contexts and hierarchies of work which may or may not be shared among all group members. Groups are complex social systems that engage in multiple ways with multiple concurrent projects [28]. As such, group members are embedded in multiple simultaneous work contexts and must prioritize between each context [28]. Not all members of a group are involved in all possible projects which leads to conflicts when group members with different sets of work contexts and hierarchies assert and act on different task priorities. Issues of multiple contexts and hierarchies of work are heightened in scientific collaboration due to ensuing issues of disciplinarity and trust [33, 38].

Designing for multiple and simultaneous contexts is a recent and developing idea in CSCW and CI research. However, early work in the area suggests that flexible and dynamic calendar systems and improved activity awareness systems may help transmit meaningful information regarding the complexity of contexts in scientific work [33]. Offering flexible calendaring and awareness systems allows collaborators to become more actively engaged in the reconciliation and/or construction of shared contexts and hierarchies of work that foster more efficient and productive group interactions [28].

4.7. Offer Familiar Tools and Assistance in Use

Recent research on the perceived value of CI environments notes that scientists depend largely on word-of-mouth and peer referrals when choosing CI applications [37]. Positive referrals of CI applications are largely based on the ability of the CI environment to closely reflect existing work practices (i.e., those practices that were in place before the move to CI) and require little new training [17, 29, 37]. Additionally, the amount and accessibility of documentation and support staff related to the CI environment and its use is a major factor that influences the choice to explore a CI opportunity [37].

However, a recent report [37], warns that documentation prepared in a domain specific language (usually computer science) may be perceived as unfamiliar or entirely incomprehensible to researchers from other disciplines. Potential disconnects between the language and availability of support systems and the domain knowledge of the scientist user can create situations where scientists feel that they are treated with contempt (by the system or the organization) because of their “lack of knowledge” [37] in regards to the CI environment. Providing tools and support systems that encourage or motivate a familiar and understandable type of use (e.g., by mimicking tools already used outside of CI) helps scientists to overcome the threat to their professional identity posed by daunting disciplinary changes like the uptake of CI based research.

5. CONCLUSION

The work presented in this paper represents the conclusion of a six month study of a decade of CI literature published in HCI, CSCW, CMC and computer engineering disciplines. However, this paper does not represent an attempt to summarize the state of the art in CI research by developing technologically rooted or functionally descriptive taxonomies of CI applications and technologies. Instead, we conducted a critical analysis of CI research through the perspective of McGrath’s seminal theory of group interaction and performance [MCGR/91]. By using a group theory as a critical lens for analyzing CI research, we aim to focus the discussion of CI environments as fundamentally about group collaboration and performance. This approach is valuable to CI researchers and designers because it serves to decenter the technology laden discussion of CI research in order to reveal the complex and nuanced ways in which CI and groups influence each other.

Our group centric perspective of CI research highlights the following recurring questions and problems with regard to the adoption and use CI environments.

- How do groups choose appropriate CI environments?
- How do groups obtain and maintain access (social and technological) to their CI environments?
- How do groups resolve socio-politically rooted data management (i.e., collection, storage, processing, sharing) conflicts?
- In what ways do the differing epistemological and ontological practices of scientific disciplines influence group well-being?
- What role does trust play in fostering group well-being with regard to the use of CI environments?
- How are technologically rooted contexts of work used to maintain group well-being?
- How do group members negotiate their value within the group with respect to the use of CI environments?

- How do scientists overcome the challenges to their professional identity presented by the adoption of CI tools?
- How can individual motivation and incentives be used to facilitate the participation of group members in CI oriented work?

After presenting our group oriented perspective of CI research, we offered a series of design implications for CI researchers and developers. These design implications highlight the ways in which groups construct their CI environment, use the environment to facilitate group interaction, and how members of a group appropriate CI as part of their socio-political positioning process with regard to the group.

We can already see our design implications in use in groups like the Meta Institute for Computational Physics (MICA) [30]. MICA is a small virtual world based collaboration among astrophysicists interested in developing research applications for virtual world. For MICA, virtual worlds are used as a fully functional CI for collaboration purposes as broad as data visualization, dissemination of research findings, and group administration. Throughout their many uses of virtual worlds as CI, MICA maintains a commitment to external access (data for simulations and visualizations is transferred into and out of several virtual worlds), support telecommunication diversity (use of wiki, website, virtual worlds, and email listservs), and facilitate social entrainment (offer frequent meetings that intersect intellectual interests with aptitude for working in virtual worlds to ensure group members are surrounded with peers of similar interest and ability). While MICA's success is not solely based on their embodiment of our proposed design principles, we do believe that the design principles illustrate in part the continued operational success of distributed workgroups like MICA. Our future research will serve to validate the proposed design principles across multiple groups and work contexts.

ACKNOWLEDGEMENTS

Work on this project was sponsored in part by NSF OCI grant #OCI-0928542.

REFERENCES

[1] Adolphs, S., Bertenthal, B., Boker, S., Carter, R., Greenhalgh, C., Hereld, M., Kenny, S., Levow, G., and Papka, M.E., "Integrating cyberinfrastructure into existing e-Social Science research," In Proc. of the Intl. Conf. on e-Social Science, Ann Arbor, USA, 2007.

[2] Arrow, H., McGrath, J.E., Berdahl, J., SMALL GROUPS AS COMPLEX SYSTEMS: FORMATION, COORDINATION, DEVELOPMENT, AND ADAPTATION, 1st ed., Sage Publications, London, UK, 2000.

[3] Assunção, M. D., & Buyya, R., "A Case for the World Wide Grid: Interlinking Islands of Grids to Create an Evolvable Global Cyberinfrastructure," IEEE Intl. Conf. on e-Science and Grid Computing, Melbourne, Australia, 1-18, 2005.

[4] Atkins, D.E. "Cyberinfrastructure and the Next Wave of Collaboration," EDUCAUSE Australia 2005, Auckland, New Zealand, 2005.

[5] Atkins, D. E., Droegemeier, K. K., Feldman, S. I., Garcia-Molina, H., Klein, M. L., & Messina, P., "Revolutionizing Science and Engineering through Cyberinfrastructure: Report of the National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure," National Science Foundation, Washington, DC., USA, 2003.

[6] Berman, F., "Making Research and Education Cyberinfrastructure Real," *EDUCAUSE Review*, 43, 4, 18-32, 2008.

[7] Birnholtz, J. P. and Bietz, M. J., "Data at work: supporting sharing in science and engineering," In Proc. of ACM Conference on Supporting Group Work, Sanibel Island, USA, 339-348, 2003.

[8] Blanke, T. and Dunn, S., "The Arts and Humanities e-Science Initiative in the UK," In Proc. of the IEEE Intl. Conf. on E-Science and Grid Computing, Washington, DC, USA, 2006.

[9] Boehner, K., Vertesi, J., Sengers, P., and Dourish, P., "How HCI interprets the probes," In Proc. of ACM CHI 2007, San Jose, USA, 1077-1086, 2007.

[10] Borda, A., Careless, J., Dimitrova, M.F., Michael, F., Jeremy, H., Paul, G., Stéphane, P., Caroline, S., Michele and Wiseman, N., "Report of the Working Group on Virtual Research Communities for the OST e-Infrastructure Steering Group," Office of Science and Technology, London, UK, 2006.

[11] Borgman, C., SCHOLARSHIP IN THE DIGITAL AGE, 1st ed., The MIT Press, Boston, USA, 2007.

[12] Borgman, C. L., Bowker, G. C., Finholt, T. A., & Wallis, J. C., "Towards a virtual organization for data cyberinfrastructure," Intl. Conf. on Digital Libraries, Austin, TX, USA, 353-356, 2009.

[13] Bos, N., Zimmerman, A., Olson, J., Yew, J., Yerkie, J., Dahl, E., and Olson, G., "From Shared Databases to Communities of Practice: A Taxonomy of Collaboratories," *Journal of Computer-Mediated Communication*, 12, 2, 2007.

[14] Burk, D. L., "Intellectual Property in the Context of e-Science," *Journal of Computer-Mediated Communication*,

12, 2, 600-617, 2007.

[15] Carlson, S., and Anderson, B.," What Are Data? The Many Kinds of Data and Their Implications for Data Re-Use," *Journal of Computer-Mediated Communication*, 12, 2, 635-651, 2007.

[16] Dolby, P., Pearce, N., Fish, A., Van Ark, T., Crouchley, R., Allan, R., "Supporting the Uptake of Cyberinfrastructure in the Social Sciences and the Challenges Faced," In Proc. of the Intl. Conf. on e-Social Science, Ann Arbor, USA, 2007.

[17] Faniel, I., "Unrealized Potential: The Socio-Technical Challenges of Large Scale Cyberinfrastructure Initiative," University of Michigan School of Information, Ann Arbor, USA, 2009.

[18] Farooq, U., Ganoe, C. H., Carroll, J. M., & Giles, C. L., "Designing for e-science: Requirements gathering for collaboration in CiteSeer," *International Journal of Human-Computer Studies*, 67, 4, 297-312, 2009.

[19] Fraser, M., Hindmarsh, J., Best, K., Heath, C., Biegel, G., Greenhalgh, C., and Reeves, S., "Remote Collaboration Over Video Data: Towards Real-Time e-Social Science," *Computer Supported Cooperative Work*, 15, 4, 257-279, 2006.

[20] Gong, J. and Lin, H., "A Collaborative Virtual Geographic Environment: Design and Development," in (Eds.), Balram, S. and Dragicevic, S., COLLABORATIVE GEOGRAPHIC INFORMATIVE SYSTEMS, 1st ed., Idea Group, London, UK, 2006.

[21] Grudin, J. 2002., "Group dynamics and ubiquitous computing," *Commun. ACM*, 45, 12, 74-78, 2002.

[22] Han, D. and Hwang, S., "e-Science Workbench: an Approach to Build Domain-Specific Problem Solving Environments," In Proc. of IEEE Intl. Conf. on E-science, Washington, D.C., USA, 749-755, 2008.

[23] Hinz-Hoare, V., "Designing a Collaborative Research Environment for Students and their Supervisors (CRESS),"August 12, 2007, Available: arXiv:0708.1624/

[24] Jankowski, N. W., "Exploring e-Science: An Introduction," *Journal of Computer-Mediated Communication*, 12, 2, 549-562, 2007.

[25] Jirotka, M., Procter, R., Hartswood, M., Slack, R., Simpson, A., Coopmans, C., Hinds, C., and Voss, A., "Collaboration and trust in healthcare innovation: The eDiaMoND case study," *Computer Supported Cooperative Work*, 14, 4, 369–398, 2005.

[26] Lee, C., Dourish, P., & Mark, G., "The human infrastructure of cyberinfrastructure," In Proc. of ACM CSCW 2006, Alberta, Canada, 483–492, 2006.

[27] McGrath, J.E., GROUPS: INTERACTION AND PERFORMANCE, 1st ed., Prentice Hall, New Jersey, USA, 1984.

[28] McGrath, J.E., "Time, Interaction, and Performance (TIP): A Theory of Groups," *Small Group Research*, 22, 147-174, 1991.

[29] Meyer, E.T., Schroeder, R., and Dutton, W.H., "The Role of e-Infrastructures in the Transformation of Research Practices and Outcomes," In Proc. of iConference 2008, Los Angeles, USA, 2008.

[30] MICA. "MICA: Meta Institute for Computational Astrophysics," MICA [website], January 21, 2009, Available: http://www.mica-vw.org.

[31] Myers, J.D., and Dunning, T. H., "Cyberenvrionments and Cyberinfrastructure: Powering Cyber-research in the 21st Century," In. Proc. of the Foundations of Molecular Modeling and Simulation Tools, Blaine, WA, USA, 2006.

[32] Neeman, H., Severini, H., and Wu, D., "Supercomputing in plain english: teaching cyberinfrastructure to computing novices," *SIGCSE Bull*., 40, 2, 27-30, 2008.

[33] Nomura, S., Birnholtz, J., Rieger, O., Leshed, G., Trumbull, D., and Gay, G., "Cutting into collaboration: understanding coordination in distributed and interdisciplinary medical research," In Proc. of ACM CSCW 2008, 427–436, 2008.

[34] Olson, G., Teasley, S., Bietz, M., & Cogburn, D., "Collaboratories to support distributed science: the example of international HIV/AIDS research," Proc. of SAICSIT 2002, South Africa, 44-51, 2002.

[35] Sonnenwald, D. H., "Collaborative Virtual Environments for Scientific Collaboration: Technical and Organizational Design Frameworks," in (Eds.) Schroeder, R., and Axelson, A.S., AVATARS AT WORK AND PLAY, 1st ed., Springer, Dordrecht, Netherlands, 2006.

[36] Strömqvist, S., "Educating the Humanities for e-Science," In Proc. of the IEEE Intl. Conf. on E-Science and Grid Computing, Washington, D.C., USA, 2006.

[37] Voss, A., Mascord, M., Fraser, M., Jirotka, M., Procter, R., Halfpenny, P., Fergusson, D., Atkinson, M., Dunn, S., Blanke, T., Hughes, L., and Anderson, S., "e-Research infrastructure development and community engagement, UK e-Science All Hands Meeting, Nottingham, UK, 2007.

[38] Walsh, J. P., & Maloney, N. G., "Collaboration Structure, Communication Media, and Problems in Scientific Work Teams," *Journal of Computer-Mediated Communication*, 12, 2, 712-732, 2007.

[39] Zimmerman, A., & Nardi, B. A., "Whither or whether HCI: requirements analysis for multi-sited, multi-user cyberinfrastructures," In Proc. of ACM CHI 2006, Quebec, Canada, 160-1606, 2006.