TRACING THE EVOLUTION OF COLLABORATIVE VIRTUAL RESEARCH ENVIRONMENTS: A CRITICAL EVENTS-BASED PERSPECTIVE

Ashley B. Trudeau

Thesis Prepared for the Degree of

MASTER OF SCIENCE

UNIVERSITY OF NORTH TEXAS

August 2016

APPROVED:

Iftekhar Ahmed, Major Professor
Brian Richardson, Committee Member and Chair of the Department of Communication Studies
Zuoming Wang, Committee Member
David Holdeman, Dean of the College of Arts and Sciences
Victor Prybutok, Vice Provost of the Toulouse Graduate School

Trudeau, Ashley B. *Tracing the evolution of collaborative virtual research environments: A critical events-based perspective*. Master of Science (Communication Studies), August 2016, 91 pp., 15 tables, references, 79 titles.

A significant number of scientific projects pursuing large scale, complex investigations involve dispersed research teams, which conduct a large part or their work virtually. Virtual Research Environments (VREs), cyberinfrastructure that facilitates coordinated activities amongst dispersed scientists, thus provide a rich context to study organizational evolution. Due to the constantly evolving nature of technologies, it is important to understand how teams of scientists, system developers, and managers respond to critical incidents. Critical events are organizational situations that trigger strategic decision making to adjust structure or redirect processes in order to maintain balance or improve an already functioning system. This study examines two prominent VREs: The United States Virtual Astronomical Observatory (US-VAO) and the HathiTrust Research Center (HTRC) in order to understand how these environments evolve through critical events and strategic choices. Communication perspectives lend themselves well to a study of VRE development and evolution because of the central role occupied by communication technologies in both the functionality and management of VREs. Using the grounded theory approach, this study uses organizational reports to trace how critical events and their resulting strategic choices shape these organizations over time. The study also explores how disciplinary demands influence critical events.

ACKNOWLEDGEMENTS

I would like to dedicate this work to my amazing parents, Tom and Cindy. Your constant support and encouragement has been key to my success throughout my life. Thank you for instilling within me a dogged motivation to achieve my goals. Thanks also to my brothers Justin and Paul for setting the bar high.

Dr. Ahmed, I do not know how to begin to thank you for all of the opportunities you have given me over the past year. I owe much of my success to your guidance and support. It has been a pleasure to work with you. Thank you to my brilliant committee members, Dr. Richardson and Dr. Wang, for your patience and support over the past year. Thanks also to Dr. Scott Poole for reading this thesis countless times and providing invaluable feedback and advice throughout this process.

I would like to thank the Department of Communication Studies at UNT for giving me such a supportive academic home. I cannot begin to describe how much I have grown as both a scholar and human being over the past two years. Thank you also to my fellow graduate students—especially my cohort—for everything. I am a better person because of each and every one of you.

Thanks to Hilary and Kelly for keeping me company over the past two years from over 500 miles away. Thanks also to Gryff, my furry companion, without whom I would not have survived grad school.

Last, but not least, thank you to Rob. You kept me sane during the last leg of thesis writing. I am so happy you’re in my life.

This research is supported by the U.S. National Science Foundation’s (NSF) Virtual Organizations as Sociotechnical Systems (VOSS) Grant Award # 1308176.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER 1

INTRODUCTION

Recent developments of high performance computing technologies and communication and coordination tools are enabling scientists to conduct their research in ways never before possible. While research collaborators were once limited by geographic distance, recent technological developments now empower dispersed scientists to collaborate, develop functional scientific teams, and to produce higher quality research (Carusi & Reimer, 2010). Some disciplines such as physics and astronomy started using dispersed collaboration as early as mid-1900. Many other disciplines, including the social sciences have already taken notice of the need to collaborate virtually, and initiatives in North America, Europe, Australia, and Asia have been instituted to explore the possibilities of virtual scientific collaboration (Jankowski, 2009). Due to the growing prevalence of collaborative research utilizing virtual technologies, it is important to understand the developmental and evolutionary processes of these research environments and how they function within various disciplines.

Throughout the past century, research communities have become increasingly more collaborative. The volume of multi-authored publications in recent years testifies to this trend (Duque, et al., 2005; Lee & Bozeman, 2005; Lukkonen, Persson, & Sivertsen, 1992; Sonnenwald, 2007; Figg, et al., 2006). Collaborative scientific group formation is often situated within organizational contexts. Therefore, it is important to consider structural elements and their interdependencies when examining how scientific collabroations develop, operate, and change. In the past few decades, technology became an important structural element that transformed scholarly activity (Sahu, 2013).

The National Science Foundation (NSF) funded the first academic computing centers and began taking action in computational science in the 1960s and 1970s (Atkins, 2003). By the 1990s, multiple supercomputing centers emerged as means to engage in large-scale data processing for scientific research in the United States. This integrated grid structure of the supercomputing centers for scientific activities was termed as *cyberinfrastructure*. The Atkins Report (2003) stated, "[t]he newer term *cyberinfrastructure* refers to infrastructure based on distributed computer, information, and communication technology. If *infrastructure* is required for an *industrial* economy, then we could say that *cyberinfrastructure* is required for a *knowledge* economy" (p. 5). Virtual Research Environments (VRE) are situated within the cyberinfrastructure that allow coordinated scientific activities amongst dispersed researchers. In addition to sharing instruments and data, VREs enable researchers with different expertise to collaborate with one another in ways that were never possible before. Advancements in technology continue to enhance research capabilities, allowing researchers to perform their activities in new and creative ways (Atkins, 2003).

Although scientific VREs across disciplines contain many structural similarities, there may be some differences between VREs in the "hard science" (i.e. natural and physical sciences) and the social science disciplines. Few differences between these disciplines are rooted in the history of VRE activity within each area. While cyberinfrastructure for the hard sciences experienced epic success, the developments largely did not meet the operational needs of the social science disciplines (Beaulieu & Wouters, 2009). Although some scholars (Borgman, 2007; Contractor, 2007) claimed that virtual research does not isolate any disciplines from growth, the undeniable emphasis centers on the hard sciences (Jankowski, 2009). Aside from the historical differences in the progression of both social and hard science VREs, there are differences due to scientific needs and practices across different scientific domains. These differences are often reflected in the operation of their VREs.

It is important to investigate how differences between social and hard sciences influence their VRE development, sustainment, and maintenance. In order to gain a rich understanding of how VREs function over a period of time, comparisons of multiple VREs in different disciplines must be drawn.

This study will compare the developmental processes of hard and social science VREs. The study will examine two prominent cases one from each domain: The United States Virtual Astronomical Observatory (VAO) and the HathiTrust Research Center (HTRC). VAO represents one of the most successful and older initiative of the hard sciences. HTRC is one of the most comprehensive social sciences VRE developed in the recent years. Communication perspectives lend themselves well to a study of VRE development and evolution because of the central role occupied by communication technologies in both the functionality and management of VREs.

Using the grounded theory approach, this study will perform a longitudinal process analysis of critical events in order to understand how these environments evolve over time. Critical events are organizational situations that trigger strategic decision making to adjust structure or redirect processes in order to maintain balance or improve an already functioning system (Gaddis, 2002; Gersick, 1991; Mintzberg, Raisinghani, & Theoret, 1976; Poole, 2004). Critical events influence strategic decision-making that guides organizational change and adaptation. This study will examine how critical events related to (i) scientific needs and requirements and (ii) disciplinary habits reflect in the VRE design process and affect organizational evolution.

Chapter 2, review of literature, reviews the existing literature and reports related to strategic decision-making and organizational evolution, virtual collaboration initiatives around the world, virtual research environments, virtual research organizations, and scientific collaboration practices of various disciplines. This review enables us to understand the history and present situation of different scientific domains, dynamics of scientific collaboration, and group and organizational processes related to dispersed collaboration. Rationale for research questions, developed based on the discussion, follows the review. Chapter 3 focuses on the method of investigation. Chapter 4 provides the results of the study following analysis. Chapter 5 discusses the significance and theoretical implications of the study. Chapter 6 summarizes the lessons learned and discusses the limitations of the study.

CHAPTER 2

RELATED LITERATURE

The process of the development and evolution of virtual infrastructure into a full-fledged research environment and then into a virtual research organization (VRO) could be unfolded by analyzing organizational critical events. Strategic decision-making through strategic choice that guides organizational change and adaptation is the theoretical base of this study. In the study, the events or sequences of events are related to the technological development, nature of individuals and groups embedded into the organizational context, outside influence, and need satisfactions. The following discussion therefore sheds light upon the strategic decision-making process, strategic choice perspective, technological evolution and its effect on cyberinfrastructure development, discussion of initiatives that addresses change in scientific practice, and the development of VREs and VROs. Based on the discussion, the final section of the chapter introduces specific research questions that the study will try to answer.

Strategic Decision-Making through Strategic Choices

Strategic decision-making is a topic often researched in connection to organizational change and adaptation (Mintzberg, 1978; Fredrickson & Mitchell, 1984; Dean & Sharfman, 1996). Mintzberg (1978) defined strategy as “a pattern in a stream of decisions” (p. 935). Mintzberg (1978) argued that strategy is the product of both formulation and evolution. In general, there are three forces that drive strategy formation (Mintzberg, 1978). First, the environment in which the organization is situated is constantly changing, requiring organizations to adjust. Second, bureaucratic forces act to stabilize the organization despite this changing environment. Finally, organizational leadership and decision makers function to mediate between the competing forces of the environment and bureaucracy (Mintzberg, 1978). This perspective of organizational strategy posits that "strategy formulation is primarily a decision making process and that a more productive research approach may be to study how organizations make individual (i.e. single) strategic decisions and whether they attempt to integrate those decisions into some overall strategy" (Fredrickson & Mitchell, 1984, p. 400).

Strategy formation does not work on a schedule. Organizations may remain in a stable, unwavering environment for a number of years before an event throws the system off balance, requiring organizational leaders to adjust their strategy (Mintzberg, 1978). Critical events are organizational occasions that trigger strategic decision making to adjust structure or redirect processes in order to maintain balance or improve an already functioning system (Ahmed & Poole, 2012; Gaddis, 2002; Gersick, 1991; Mintzberg, Raisinghani, & Theoret, 1976; Poole, 2004). Critical events have several characteristics. First, while these disruptions demand the attention of organizational decision makers, critical events range in severity (Mintzberg et al., 1976) and impact on organizational activities (Gersick, 1991). Second, they may be anticipated or unanticipated and are not temporally bound. Third, critical events may be a sudden crisis requiring immediate action or a slow recognition of a problem (Mintzberg et al., 1976). Finally, the decisions made because of critical events involve a change in structure, policy, or practice (Ahmed & Poole, 2012).

In their model of strategic decision-making, Mintzberg et al. (1976) identified three types of decisions in their decision recognition routine: opportunity decisions, crisis decisions, and problem decisions. These decisions are triggered by events that bring attention to issues that require or inspire change. On the less severe end of the spectrum, opportunity decisions are often triggered by an organization member's ideas. These decisions are purely voluntary and usually made to improve an already stable environment. This idea may be dormant in their mind for a period, but the individual may take action on this opportunity when they are in a position of leadership. On the opposite end of the spectrum, crisis decisions are triggered by single stimuli that require immediate action. In between opportunity and crisis, decisions in terms of severity are problem decisions. Problem decisions are usually triggered by multiple stimuli and do not necessarily require the immediate action of decision makers. As organizational problems often involve many interdependent factors, decision makers may need to take their time to understand the problem in its complexity (Mintzberg et al., 1976) and enact their decisions.

'Strategic choice' is the process of action in which organizational power-holders enact their decisions (Child, 1972). Dissatisfaction with previous understandings of organizational structure inspired the development of the strategic choice perspective. Prior organizational research was largely positivistic and cross-sectional, positing that "the way in which organizations are designed and structured is determined by their operational contingencies" (Child, 1997, p. 43; Child, 1972). Child (1972) argued that the choice to use statistical methods to discover relationships between organizational characteristics leaves emergent processes to be inferred. Thus, Child (1972) aimed to introduce the strategic choice of organizational power-holders as a key influence on the context, standards of performance, and structural design of the organization. Child (1972) stated, "When incorporating strategic choice in a theory of organization, one is recognizing the operation of an essentially political process in which constraints and opportunities are functions of the power exercised by decision makers in the light of ideological values" (p. 16).

Child (1972) argued that prior theories of organization were insufficient because they promoted organizational structure as the orienting and intervening feature of organizations. In contrast, Child (1972) introduced the idea of strategic choices made by decision makers as the dominant factor. The strategic choice perspective accounted “for organizational variation directly through reference to its sources rather than indirectly through reference to its supposed consequences” (Child, 1972, p. 14). Childs (1972) echoed arguments made by Weick (1969), recognizing the importance of choice when organization members enact their environments by making decisions such as the location of the organization, hiring decisions, and relationships to formulate. While prior theories suggest that the structural constraint of network proximity determined the organizational relationships formed between people of power, the strategic choice perspective positions these relationships as the products of strategic and political decisions.

Child (1972) referred to those in the position of power as the “dominant coalition.” This term does not necessarily suggest a formal position within the organization and, as such, those within the dominant coalition may change over time. Child (1972) stated, “The dominant coalition concept draws attention to the question of who is making the choice. It thus provides a useful antidote to the sociologically unsatisfactory notion that a given organizational structure can be understood in relation to the functional imperative of ‘system needs’ which somehow transcend the objectives of any group of organizational members” (p. 14).

Based on the model of strategic choice, Miles, Snow, Meyer, and Coleman (1978) developed the adaptive cycle, consisting of three “problems” of organizational adaptation that power-holders must solve: the entrepreneurial problem, the engineering problem, and the administrative problem. The entrepreneurial problem is focused on defining the organization and establishing goals. The engineering problem “involves the creation of a system which operationalizes management’s solution to the entrepreneurial problem” (Miles et al, 1978, p. 549). Solutions to engineering problem involve selecting specific resources in order to attain some outcome. Finally, the administrative problem involves rationalization of the system in order to maintain stability and reduce uncertainty. While the administrative problem is relevant when reducing uncertainty for the existing system, this problem is also pertinent during the process of evolution and innovation. Miles et al. (1978) also developed strategic typology consisting of three types of organizations: defenders, prospectors, and analyzers. Defenders are responsible for enacting and sustaining a stable environment by choosing to "carve out and maintain a small niche within the industry which is difficult for competitors to penetrate" (Miles et al., 1978). While defenders and prospectors represent two opposite extremes, analyzers represent a balance between the two. Analyzers are organizations that aim to seek opportunity while also minimizing risk and maintaining stability.

Technological Development and the Beginning of VRE Ideation

Technological advancements over the past few decades have transformed the ways researchers conduct their research, improving efficiency in cost, time, and quality of research (Sahu, 2013). According to Sahu (2013), research technologies have facilitated examinations of large-scale populations with a massive amount of data with excellent accuracy. Technology has also enabled scholars to easily access large bodies of literature and has enhanced distance collaboration capabilities (Sahu, 2013). Reilly (2003) identified 10 milestones of development in computer science and technology. Of these milestones, the last three are the most significant to the development of technology today: the Personal Computer (PC), Graphical User Interface (GUI), and—most importantly—the World Wide Web (WWW) (Reilly, 2003; Sahu, 2013). The PC converted more environments into places of computing as it allowed users to access computing technologies from virtually anywhere. The GUI extended the functionality of the Apple Macintosh followed by the PC and made computing more accessible and user friendly. Finally, the WWW drew massive appeal as it allowed users to access information and connect with eachother across distance almost instantaneously. The WWW gained 50 million users in less than four years (Sahu, 2013). Web 2.0—an evolution of the WWW—involved "the deployment of applications which leverage the Internet and improve as more people use the application" (Sahu, 2013, p. 3). The development of Web 2.0 marked a massive shift in society's understanding of information accessibility and interaction. While the original conception of the WWW allowed web developers to provide content for users, Web 2.0 empowered users to take an active role in the development of content through tools such as blogs, wikis, social networks, and virtual worlds (Sahu, 2013). These Web 2.0 capabilities also enabled scientists to work in ways never before possible.

Web 2.0 technologies intervened in research in several ways. First, the Internet changed the way researchers identify research problems. Researchers are now able to garner up-to-date information about potential research problems immediately through on the Internet. Researchers can also gather ideas through social media networking websites such as Facebook and Twitter. Online databases allow researchers to bypass libraries entirely by providing digital access to relevant literature. Data collection technologies help researchers to conduct virtual interviews through various voice, video, and messaging technologies, allowing researchers to conduct interviews with far away recipients without the time and cost of travel. Similarly, online survey collection technologies allowed researchers to collect a wider range of data without associated costs. Advancements in online data collection technologies have also reducted daa entry time significantly (Sahu, 2013). Perhaps the most significant capability afforded by Web technologies is the ability for researchers to collaborate with one another in new and meaningful ways.

The technology that researchers use to collaborate with one another is as varied as the work styles of the scholars themselves (Datta, Rzadca, Ang, & Hong, 2010). According to Datta et al. (2010), there are two general types of collaboration software that research groups use: commit-based and real-time. Commit-based systems work as a database in which users deposit shared data after they believe their work is complete and then collaborators can view and update it. Real-time software—as the name suggests—allows users to view changes to the data as they are being made. Some collaboration technologies allow users to maintain local databases and allow access for input and collaboration from peers. Personal Bibliographic Data Management System (PBDMS) allows users to manage, share, and annotate bibliographies and collaborate virtually with peers. This technology enables users to subscribe to other bibliographies and chat functions facilitate collaboration among collaborators (Datta, et al., 2010). In addition to technologies allowing researchers to access shared materials, software has also been designed to assist in specific stages of a research project (Brunvand & Duran, 2010). The institution of Web 2.0 inspired the development of many tools to aid researchers in collaborating with one another on specific projects. MyNetResearch is one of the most inclusive VREs available as it allows researchers to communicate and share projects, files, and resources within a completely web-based environment. Amongst its many capabilities, MyNetResearch tools facilitate in locating funding, data collection, and literature search (Brunvand & Duran, 2010).

The innovations in communication technology have afforded many opportunities to scientists that marked a shift in scholarly activity. As powerful scientific research organizations became aware of the increasing capabilities available in Web technologies, they began to focus their attention to the development of shared research technologies to facilitate collaborative research.

Maintaining National Significance through Sustained Scientific Collaboration: VRE Initiatives across the Globe

Development and sustenance of present day national economies depends upon scientific innovation and production. It cannot be achieved without a significant amount of effective collaboration among scientists. Scientific collaboration consists of social interaction between scientists in which they share knowledge and tasks in order to achieve some mutually conceived goal (Sonnenwald, 2007). Scientific collaboration has become a seemingly permanent research practice throughout the last century of scholarly activity (Duque, et al., 2005; Lee & Bozeman, 2005; Lukkonen, Persson, & Sivertsen, 1992; Sonnenwald, 2007; Figg, et al., 2006). Scientific collaboration increased due to the professionalization and specialization of scientific practice and it allowed scientists to share various resources such as knowledge and information, equiptment, and networking contacts (Duque, et al., 2005; Lukkonen, et al., 1992). Scientists have also trended towards collaboration due to the encouragement of funding agencies due to high costs of technology and other associated costs of research. Because of this, collaborative research has grown more interdisciplinary (Duque, et al., 2005; Lee & Bozeman, 2005).

Atkins Report (2003) mentions that the NSF funded the first academic computing centers and began taking action in computational science in the 1960s and 1970s. By the 1990s, supercomputing centers emerged as a means to engage in large-scale data processing for scientific research. This was termed “cyberinfrastructure” in the United States. The Atkins Report (2003) stated, “The newer term *cyberinfrastructure* refers to infrastructure based on distributed computer, information, and communication technology. If *infrastructure* is required for an *industrial* economy, then we could say that *cyberinfrastructure* is required for a *knowledge* economy” (p. 5).

In order to maintain leaderships in innovation, the advancement of virtual research has largely been driven by national initiatives. In 2001, the United Kingdom instituted the 'e-Science Programme', which "aimed to give UK researchers a leading position in the development of 'grid' technologies, research software and other e-infrastructures, which facilitate and enable research across all disciplines" (Research Councils UK, 2014). The grid technology provided access to large collections of data and technological resources that could be accessed remotely (Brunvand & Duran, 2010). In 2004, the National Centre for e-Social Science (NCeSS) was created by the U.K. Economic and Social Research Council (ESRC) in order to "develop leading-edge methodological tools and techniques to enhance the U.K. social science research community's capacity to collect, discover, access, manipulate, link, share, analyze and visualize both quantitative and qualitative data" (Halfpenny, Procter, Lin, & Voss, 2009, p. 73). In its initial phases, NCeSS identified two research goals: understanding applications and social nature of virtual research. First, researchers sought to understand how web applications could be used to assist in various research methods. Second, researchers examined how VREs were created, its role in research communities, and its implications for social science practice and research (Halfpenny, et al., 2009).

The National Science Foundation (NSF) played a massive role in the development of virtual research in the United States. The NSF established the Office of Cyberinfrastructure "to provide user-friendly, reliable information technology and knowledge management to all researchers and educators to catalyze discovery at the frontiers of all science and engineering disciplines" (National Science Foundation, 2005). The development of a virtual research initiative was promoted by a 2003 report of the NSF Blue-Ribbon Advisory Panel on

Cyberinfrastructure, which framed virtual research as the key component to a “knowledge economy” (Atkins Report, 2003).

In 2005, Australia’s e-Research Coordinating Committee released a report stressing “the need for Australia to develop a world-class e-research capability across all research disciplines” (Paterson, Lindsay, Monotti, & Chin, 2007, p. 123). Unlike the major initiatives in the U.K. and U.S., Australian initiatives generally use the more inclusive term “e-research” and their priorities reflect this terminology (Genoni, Merrick, & Wilson, 2009). The National Collaborative Research Infrastructure Strategy (NCRIS) is a program developed by the Australian government to provide cyberinfrastructure capabilities to Australian researchers (Australian Government, Department of Education and Training, 2015). The NCRIS emphasizes “collaboration from the outset, the strategic identification of capabilities through the consultative road mapping process, the facilitation process to develop capability plans and the provision of funding for skilled staff and operating costs” (National Collaborative Research Infrastructure Strategy, n.d. , p. 1). The Systemic Infrastructure Initiative (SII) funded the Managed Environment for Research Repository (MERRI) projects, which were “designed to enhance the ability of Australian researchers to tap into new resources and to better share their results with the wider community” (Paterson et al., 2007, p. 123). One of these projects—DART—aimed to give dispersed researchers virtual access to an infrastructure of shared tools and resources. DART also provided support in communication, project management, data sharing, and literature searching. These national and international initiatives transformed the landscape of scientific collaboration and practice through VREs. In the past, collaboration amongst groups of researchers took place face-to-face in collocated laboratories, which often restricted access to potentially interested collaborators working in different locations. Developments in technology throughout the past few decades enabled dispersed scholars to collaborate in ways that were never possible before. Anandarajan and Anandarajan (2010) stated, “Without the physical boundaries of traditional social networks, online social networks replicate and enhance the benefits of traditional social networks across time and space and accelerate and globalize the process” (p. 7). In the beginning of virtual collaboration, e-mail facilitated faster communication and file sharing between researchers. Tools developed out of the Web 2.0 era—such as instant messaging, blogs, and Wikis—facilitated social networking in which researchers could share their interests and ideas with one another (Anandarajan & Anandarajan, 2010). Research practices began to evolve and intertwine with technology as virtual interaction became more accessible and now collaborators are able to perform nearly all of their collaborative work virtually.

Virtual Research Environments

Virtual research environments (VRE) are cyberinfrastructures that enable researchers to collaborate across any distance and boundary (Carusi & Reimer, 2010; Fraser, 2005). There are many alternative terms for VREs such as collaborative e-Research communities, collaborative virtual environments, collaboratories, cyberinfrastructure, and e-science, but for the purposes of this study, VREs will be the term of choice (Carusi & Reimer, 2010; Jankowski, 2009). According to Fraser (2005), “The VRE, for the most part is the result of joining together new and exciting components to support as much of the research process as appropriate for any given activity or role” (n.p.).

In their survey of VREs, Carusi and Reimer (2010) discovered three trends about the development of VREs. First, VREs seemed to be the product of collaboration between researchers and software developers, with researchers (as users) frequently directing the development and emphasizing the needs of the users. Before developing the VRE, researchers often conceptualized the system and defined the requirements necessary to its functionality. Along with this first finding, the development process is most successful when it is user-centered and feedback-driven. Iterative development is incredibly important, especially if there is variation in the technical knowledge and expertise of users. One of the main challenges presented by VREs is the lack of technical knowledge necessary to operate the tools of the system, so it is important for developers to gain user feedback throughout the development process. Finally, research libraries often view VRE development as an opportunity to support the work of researchers. Carusi and Reimer (2010) stated that the research community occasionally approaches research libraries to propose a partnership in developing a VRE.

Collaborative virtual research is also becoming a more prevalent research practice for another major reason. Funding agencies invest money in VREs because they see many benefits in geographically dispersed research groups including the possibility of international and interdisciplinary collaboration, increased researcher productivity, increased access to expensive technological resources, faster communication and dissemination of research outcomes, and higher overall quality of research findings (Carusi & Reimer, 2010). In general, funding bodies usually view collaboration as a positive attribute when providing funding, especially when the project is interdisciplinary or multi-institutional (Newell & Swan, 2000).

Bos et al. (2007) identified a typology of VREs based on their technical and organizational attributes. First, some VREs function to provide multiple researchers access to a shared instrument. Shared instrument VREs have enhanced the scientific discovery process as they “have often pushed the envelope of synchronous (real-time) communications and remote-access technology” (Bos, et al., 2007, p. 660). When managing large amounts of data outputs from these instruments, users must carefully manage security issues. Organizational issues arise in large research groups using a shared instrument because the system may become oversubscribed. In these situations, VREs may appoint a committee to oversee access to the instrument (Bos et al., 2007).

Second, a Community Data System is an "information resource that is created, maintained, or improved by a geographically-distributed community" (Bos, et al., 2007, p. 660). These semi-public VREs usually contain large amounts of data accounting for a general interest area. Technically, these VREs require communities to standardize their data and these communities often draw users seeking large data sets to develop modeling techniques. Organizationally, these communities often face challenges in motivating contributors (Bos et al., 2007). Third, an Open Community Contribution System is a VRE in which geographically dispersed contributors collaborate on a research problem within an open system. These VREs are more open to contributors than the Distributed Research Center and, unlike the Community Data System, it often involves the contribution of work rather than data. An example of this kind of community is a project by astronomers in which non-scientist volunteers performed two scientifically important tasks: detecting and classifying craters (Kanefsky, Barlow, & Gulick, 2001). This kind of research is particularly pertinent to the astronomy field, which has been known to seek out collaboration with amateurs for data analysis work. Kanefsky et al. (2001) revealed that "Important contributions have been made by amateur astronomers in several areas of research, including monitoring dust storms on Mars, timing asteroid occultation, and discovering comets" (p. 1). Open Source Community Systems often face the challenge of ease of usability and accessibility across multiple platforms, especially in collaborations between scientists and non-scientists. In addition, quality control of product is a common issue in Open

Source Community Systems due to the large number geographically diverse contributors (Bos, et al., 2007).

Fourth, a Virtual Community of Practice is a VRE in which researchers with similar interests communicate online. In these VREs, scholars can solicit information from their colleagues such as advice, new techniques, and ideas, but the goal is not necessarily to initiate research projects (Bos et al., 2007). These environments may operate similarly to social-networking sites in that researchers can engage with one another somewhat informally about mutual interests (Brunvand & Duran, 2010). The technological resources used in Virtual Communities of Practice can vary, but the main technological factor that influences collaboration is usability. Bos et al. (2007) posited that prosperous Virtual Communities of Practice often employ tools such as "listserv, bulletin boards, and accessible web technology" (p. 663). Like Open Community Contribution Systems, Virtual Communities of Practice also experience fluctuations in participation and membership as these VREs are in competition with other interest areas online (Bos, et al., 2007). Fifth, Virtual Learning Communities aim to "increase the knowledge of participants but not necessarily to conduct original research" (Bos et al., 2007, p. 663). These environments are common to online degree programs and other formal education programs (Bos et al, 2007).

Sixth, Distributed Research Centers function similar to university research centers but they are geographically distributed. This VRE is marked by increased interpersonal interaction about a research area of interest. Distributed Research Centers often experience the same technological issues as other types of VREs such as data standardization and security, but these VREs are unique in their organizational complexity (Bos et al, 2007). Distributed Research Centers "must gain and maintain participation among diverse contributors, work to standardize protocols over distance, facilitate distributed decision-making, and provide long-distance administrative support” (Bos et al., 2007, p. 665). Researchers must also consider issues of intellectual property when working with more than one institution as well as the ethical dynamics of collaborating and co-authoring with younger scholars (Bos, et al., 2007).

Finally, Community Infrastructure Projects are VREs that “seek to develop infrastructure to further work in a particular domain” (Bos, et al., 2007, p. 666). Community Infrastructure Projects are often interdisciplinary as researchers from different backgrounds share scientific tools, protocols, and methods. These environments require the institution of standards and protocols for collecting and handling data. Community Infrastructure Projects require the careful consideration of organizational issues when working in this kind of VRE. Interdisciplinarity may pose an organizational challenge within these environments because researchers must negotiate research goals and agendas with others. Collaborations between computer scientists and researchers within other disciplines, for example, may experience challenges in negotiating technical advancement—the focus of the computer scientist—and practical usability—the focus of the disciplinary other (Bødker & Grønbæk, 1991; Weedman, 1998). Researchers must also decide between academic and private sector management. Finally, younger scholars often face the issue of how their work in building this infrastructure furthers their academic careers and if this work is considered as reputable as publication. If not, these scholars must decide if they should avoid this work in favor of pursuing publication (Bos, et al., 2007). Each type of VRE Bos and colleagues identified in their typology serves a unique purpose that facilitates different types of work, however research has shown that researchers who work in VREs share similar motivations for engaging in these communities.

Scientific Research Groups within VREs

Scientific research groups are collections of scientists who collaborate with one another in the development of research and sharing of funding, data, and technical resources. However, these groups do not necessarily function under the same formalized organizational structure of the institution in which it operates (Zulueta & Bordons, 1999). Research groups often form through network connections between scholars and their collaborators. Past studies of networks have examined close ties between people based on subjective data and—while they revealed important structural trends—the data collected was frequently both unsubstantial and uncontrolled (Newton, 2001). Newton (2001) advocated for the study of affiliation networks, which comprise a collection of individuals who are connected by a common participation in some sort of group. Researchers can measure membership of groups more easily than ties of friendship and, thus, conduct more precise research (Newman, 2001). One type of affiliation network specific pertinent to this study is a co-authorship network. Co-authorship networks are made up of the relationships signified by the collaboration of a group of researchers on a single paper (Newman, 2001; Perianes-Rodriguez, Olmeda-Gomez, & Moya-Anegon, 2010). Perianes-Rodriguez et al. (2010) championed co-authorship as an excellent unit of measurement of scholarly networks of collaboration. Co-authorship is indicative of a much stronger tie than other connections between scholars such as citation, which simply indicates a referential rather than social relationship (Perianes-Rodriguez et al, 2010). Networks of scholars within the sciences are usually highly clustered. In other words, two scientists are significantly more likely to collaborate with one another if they have another collaborator in common (Newman, 2001).

Actors within the research group often vary in involvement. Price and Gürsey (1976) identified four categories that characterized the nature of co-authoring patterns: continuants, transients, newcomers, and terminators. Continuants are the researchers who publish work before and after the given collaboration (Braun, Glanzel, & Schubert, 2001; Price & Gürsey, 1976; Wagner & Leydesdorff, 2005). Braun et al. (2001) reported that continuants rarely appear single authored works but produce more than 10 published papers per year. Transients are those who publish in the given year of the project, but not before nor after. Newcomers publish in the given year of the project and after, but not before. Finally, terminators publish before and in the given year of the project but never again. Braun et al. (2001) found continuants frequently appear as co-authors and seem to serve as mediators in collaborations including the other categories of actors. Thus, continuants seem to determine the degree of collaboration the project will entail.

In their exploration of networks in co-authorships, Wagner and Leysdorff (2005) argued that continuants operate in two ways: as "nodes," to whom less productive scholars attach themselves or—in the event that the scholar is extremely well connected—"hubs," to whom many collaborators attach themselves. Hubs are the more powerful entities who no longer compete in order to achieve a higher status reputation, but focus on building their network of intellectual contacts. High status continuants attract transient and newcomer individuals who seek to build a reputation through collaboration and co-authorship with highly regarded scholars (Wagner & Leydesdorff, Network structure, self-organization, and the growth of international collaboration in science, 2005). Networks studies also show that some laboratory directors have a very large network of collaborators because they attach their names to all of the papers resulting from studies within their laboratories (Newman, 2001).

As groups continued to collaborate within VREs, many eventually become organizations in their own rite. The structure of these virtual research organizations has, in many ways, mirrored collocated organizations. As in collocated research organizations, virtual organizations typically have designated individuals responsible for the management and coordination of stakeholders as well as scholars who execute the scientific activities of the project. That being said, the scientific processes and the virtuality of communication represent the key differences between collocated and virtual organizations.

From VRE to Virtual Research Organization (VRO)

The Institute of Medicine National Research Council (IMNRC) (2002) described scientific research environments as dynamic social organizations consisting of many interdependent parts such as researchers, administrators, and funders. The IMNRC (2002) developed an open-systems model of research organizations consisting of five interrelated parts: (1) the external conditions of the research environment that affect the outputs of organizational work, (2) inputs and resources such as funding and human resources activities, (3) the internal environment consisting of organizational structure, processes, and ethics, (4) outputs and outcomes of organizational work, and (5) feedback. According to Storer (1973), researchers have been studying the scientific organization since the early 1950's and many of these studies focused on the issues of morality and productivity in research environments. In their study of scientists in organizations, Pelz and Andrews (1976) argued that, while a researcher may desire independence and autonomy, it does not mean that he or she has total freedom in his or her work. In fact, Pelz and Andrews (1976) found that highly autonomous scholars in loosely coordinated organizations experience less stimulation, withdraw from colleagues, and grow somewhat disinterested in their research.

Structural elements are the more enduring aspects of scientific organizations (IMNRC, 2002). Many different elements of organizational structure are integral to function and success of an organization. Collaborative groups within the organization influence the structure of the organization and will be discussed further in the following section. In recent history, collaborating groups have become increasingly interorganizational due to the accessibility of virtual communication technologies (Stohl & Walker, 2002). The use of technological resources in scientific research has become incredibly prevalent and will be discussed further in this section. An understanding of status and hierarchy within the research environment provides a "blueprint for researcher behavior" (IMNRC, 2002, p. 55), which embodies the culture of the organization. In addition to these more formal structures, organizations also contain:

> informal epistemic community norms and conventions, which will be recognized (if not always adhered to) by members of the various scientific and technological professional groupings, as well as some particular 'local social norms' that are likely to emerge among colleagues engaged in recurring or extended research collaborations (David, 2004, p. 9).

Another structural element is the organization's collection of missions, goals, plans, and practices (IMNRC, 2002). Identifying specific goals and visions for a research project can be difficult in the early phases of research, but it is important for researchers to communicate these to the stakeholders involved (Sonnenwald, 2007). Sonnenwald (2007) stated, "Articulating clear visions and goals that multiple individuals and groups can understand and support is a skill scientists' need when initiating large and complex scientific collaborations" (p. 657). Finally, scientific organizations must possess clearly defined policies and procedures for conducting responsible research (IMNRC, 2002). Research organizations often must follow specific legal regulations regarding responsible handling of "contracts, liability, privacy and intellectual property" (David, 2004, p. 9). While the above structural elements involved in the research organization that influence the conduction of research, scientific research groups themselves do not necessarily function in the same ways as their parent organizations (Zulueta & Bordons, 1999).

In summary, the structural elements of an organization are critical to the process of scientific collaboration. While many of the above features of organizational structures are generally present across most research organizations, different research disciplines can exhibit some variation in collaboration practices.

Scientific Collaboration and Disciplinary Differences

Sonnenwald (2007) argues that scientific collaboration “has the potential to solve complex scientific problems and promote various political, economic, and social agendas, such as democracy, sustainable development, and cultural understanding and integration” (p. 643). However, scientific collaboration does not necessarily mean the same for different scientific domains. The nature of scientific collaboration changes with the nature of scientific practice. There are several factors that influence the process of scientific collaboration (see Table 1 for details). The following section discusses few significant aspects that differentiate collaborative patterns across disciplines.

Difference in Scholarly Output

Collaboration has become an important aspect of all scholarly domains of today but research has shown that there are some disciplinary differences in collaboration patterns. While natural science and engineering disciplines published more than humanities in general, studies show that disciplines such as biology, physics, and medicine publish significantly more collaborative work than the humanities or social sciences (Babchuk, Keith, & Peters, 1999; Lariviere, Gingras, & Archambault, 2006; Newman, 2001).

Theoretical and Methodological Orientation

Shin and Cummings (2010) argued that disciplinary consensus differences determine the degree to which scholars in a given discipline collaborate, stating:

> In the hard-disciplines, researchers apply standardized research methodology while in the soft-sciences the methodology is diverse depending on perspectives on which the research is based. In the soft-sciences, research publication takes longer and may require greater effort, and it is therefore difficult to get a consensus on theory and perspectives" (p. 591).

Mentoring Practices

Mentoring within certain disciplines may also play a role in collaborative practices. Babchuk et al. (1999) argued that—in fields such as chemistry—many publications based on doctoral dissertations are co-authored by an advisor because graduate students seem to have less autonomy in pursing research interests in those disciplines.

Existing Collaboration within the Community of Practice

In certain disciplines, clustering significantly affects how researchers collaborate. Famous for being highly collaborative, network analysis of the high-energy physics discipline has shown that scholars are more likely to collaborate with scholars with whom they have a mutual collaborator. Biomedical research, on the other hand, shows lower degrees of clustering amongst researchers, perhaps due to more rigid organizational structures (Newman, 2001).

Interdisciplinarity

Interdisciplinary research involves incorporating more than one discipline into a research environment. Interaction within an interdisciplinary environment consists of "interactions among individual scientists, between individual scientists and their organizations, and among different disciplines involved in the research" (Qln, Lancaster, & Allen, Types and levels of collaboration in interdisciplinary research in the sciences, 1997). Interdisciplinary collaboration is becoming more common as certain projects require certain expert knowledge and skill beyond that of one scientific discipline (Hagstrom, 1965). Not to mention, interdisciplinary collaborations have produced important fields such as oceanography and cognitive science (Cummings & Keisler, Collaborative research across disciplinary and organizational boundaries, 2005).

Changing Need of Collaboration Patterns and Community Development

Research collaboration in the sciences has contributed to the evolution of scientific practice significantly and research in community has become even more interconnected as society continues to shift towards virtuality. However, interdisciplinary nature of collaboration did not take away the disciplinary culture of VROs as majority of the VROs are developed through disciplinary community collaboration. Collaboratories at a glance project lists a significant amount of available VREs/VROs that support this argument.

Impact of Technology

Grid technology, a contribution of the particle physics field, allowed researchers to move beyond information sharing on the Internet to computational collaboration via the Grid, a groundbreaking and transformational development for cyberinfrastructure (Beaulieu & Wouters, 2009). Grid technologies enabled the collaborative work of many computers—either collocated or distributed—to function as one superior system (Dutton & Jeffreys, 2010a). While Grid technologies immediately benefitted research processes in natural and physical science fields, they experienced slower uptake in the social sciences and humanities, even in the more data-rich fields where such technologies seem appropriate (Dutton & Meyer, 2010b).

Research Output/Goal of Research

In general, hard science and social science differ in research objectives, which often translate to their perceived relevance in the eyes of funders. Hard science research is often focused on new discoveries, which allows scientists to create new products or processes, while social science research is often focused on identifying new insights about particular social phenomena (Lewis, Funding social science research in academia, 2000).

External Influence/Resources

While cyberinfrastructure for the hard sciences was experiencing epic success, the developments largely did not meet the operational needs of the social science disciplines (Beaulieu & Wouters, 2009), nor were funding organizations providing support for the development of such technologies. There are four types of bodies that provide funding for academic research. First, the Higher Education Funding Councils provide funding for research based on Research Assessment Exercises, which assess performance. Second, non-profit research councils such as the Social Science Research Council (SSRC) and the Economic and Social Research Council (ESRC) generally fund academic programs that study particular areas. Charitable organizations—while considered small players in providing social research funding—often provide funding for research that benefits the greater good. Finally, government organizations—such as the NSF—provide funding perceived as “relevant” to particular agendas (Lewis, Funding social science research in academia, 2000). Social science researchers meet challenges in attempting to gain funding because “there is pressure on the traditional “academic” funders of research to be supporting more relevant work. This inevitably leads on to questions of ‘relevant for whom?’ and the need to consider the role of potential research users and beneficiaries” (Lewis, 2000, p. 375). Funding bodies concerned primarily with funding research relevant to current, short-term concerns are limiting the impact that social science research could have in the long-term. Lewis (2000) further argued that truly relevant research seeks to gain a deeper understanding of some phenomenon beyond the output of information.

While there is an undeniable emphasis on hard science with regards to virtual research (Jankowski, 2009), social sciences have received more support for the development of research technologies in the effort to sustain competitiveness with international players (Dutton & Meyer, 2010b). In an NSF report on Cyberinfrastructure and the Social Sciences, Berman and Brady (2005) asserted that "Cyberinfrastructure has the potential to be a fundamental enabler of innovations and new discoveries, and it is just as critical for the advancement of the social, behavioral, and economic (SBE) sciences as it is for engineering and the physical, natural, biological, and computer sciences" (p. 4). Berman and Brady (2005) also stressed the capabilities of social science researchers to collaborate with computer scientists just as successfully as researchers within hard science disciplines did.

Rationale and Research Questions

In consideration of the available literature, this study aims to understand VREs structural development and evolutionary processes. VREs are technology-mediated organizations embedded in cyberinfrastructure aiding dispersed scientific groups to collaborate and engage in research and innovation. Majority of the VREs are either embedded in larger organizations (i.e., universities or research institutes) or funded by external entities (i.e., NSF) or both. For extensive technology mediation, beside scientists and managers, a significant number of technologists are present within the organizational environment whose roles are crucial in organizational success. It is also important to note that the majority of the VRE development initiatives are at a relatively early stage. Proper guidelines for VRE development are still unavailable and most of the projects are constantly being shaped by tackling organizational crises or problems. Therefore, it is very important for us to understand the nature of critical events that emerge during VRE operations as these critical events are shaping the evolution of VREs. Understanding structural and processual problems and decision-making within VREs will provide insight into how VREs operate on a larger scale and thus, provides a foundation upon which theoretical arguments can be built.

RQ 1: What critical events shaped VAO and HTRC evolution?

Social and hard science researchers require different capabilities to meet their needs and it is important to investigate their larger differences and similarities. One of the objectives of this study is to uncover the influence of scientific needs and requirements, and disciplinary habits in the VRE design process and VRO development. The discussion shows us that the conceptualization of VREs does not necessarily differentiate scientific domains. A majority of the current classifications emphasizes structural elements (i.e., instruments and data structure) or group composition (i.e., open community contribution) while conceptualizing VREs (Bos et al.'s discussion is an example of that). Although scientific VREs contain many structural similarities, there may be some differences between VREs in the "hard science" (i.e. natural and physical science) disciplines and the social science disciplines. The differences between these two groups are easily observed in the history of virtual research. We can also see differences in work habit, problem conceptualization, innovation, and the concept of accomplishment among scientists of different domains. How these disciplinary differences translate into organizational needs and requirements and thus, create critical events, is an important question to consider in understanding VRE design process and organizational evolution.

RQ2: How do disciplinary-specific demands and requirements influence critical events?

CHAPTER 3

METHODOLOGY

This study is a comparative analysis of two VREs from the social sciences and hard sciences, respectively. First, the United States Virtual Astronomical Observatory (VAO) is a large-scale virtual environment that provides astronomers access to the tools and resources they need to conduct productive research. Second, the HathiTrust Research Center (HTRC) is a VRE created by Indiana University and the University of Illinois that facilitates collaborative research through the development of research technologies to correct the challenges researchers face when collecting large amounts of academic literature. VAO and HTRC have been chosen as sample cases as they were the two of the most significant cases of a larger project aiming to understand how VREs develop and change over time. Both organizations focus solely on the development of tools to support a VRE that meets the needs of a user community. In addition, both organizations maintain significant records of information about the development and maintenance of the organization that helped in the construction of a working timeline of the history of each organization.

This study follows a process research method to identify event sequences that produce some outcome (Poole, Van de Ven, Dooley, & Holmes, 2000). Poole et al. (2000) argue that the process method reveals the development, turning points, and contextual interdependencies of elements in the organizational process. Following that argument, this study will utilize activity tracks to examine the process of VRE development. Activity tracks (also termed “events”) are developmental markers that assist in mapping development and finding interdependencies (Ahmed & Poole, 2012; Poole et al., 2000; Ahmed et al, 2016). Activity tracks are useful when examining multiple cases because they are “distinguished primarily for analytical purposes, so that there are similar constructs across cases and within the same case and so that complex processes can be analytically decomposed" (Ahmed & Poole, 2012, p. 5-6).

A process method yields insight into how a VRE develops by observing critical turning points and how multiple contextual elements interact as the process of organizational change unfolds. It is possible to view organizational developmental process as a whole and relate the holistic process to its ultimate effectiveness as judged by various criteria (e.g., Poole & Holmes, 1995). However, it is often more useful to consider the critical turning points or "ups and downs" of the effectiveness of the developing entity at various points in its process. Moreover, this process method perspective identifies event that looks like a roaring success at one point but turns out to be a failure at another or vice versa.

This study utilizes five activity tracks for examining the development. These tracks include: technological design and implementation, scientific work, community of VRE users, developers, funders, and stakeholders, managerial and organizational system, and critical events. The analysis of technological attributes of the VRE seeks to identify the types of technologies used in the VRE, the interdependencies between technologies, and the processes used to "create, integrate, deploy, and maintain the VRE" (Ahmed & Poole, 2012; Ahmed et al, 2016). The science track identifies the social and technical processes in which individuals 'do' science in VREs. The community track examines the usability (task allocation, user nature, and task-user interaction) and sociability (purpose of community, role definition, and distribution of roles) of the VRE community. The management and organization track identifies two structures within VREs: (1) the structure surrounding the relationship between scientists and external stakeholders such as funders and (2) the management of research processes. Finally, the critical events track isolates one-time events that affect the entire organizational development (Ahmed & Poole, 2012; Ahmed et al, 2016).

Identification of Critical Events

Critical events are organizational situations that trigger strategic decision making to adjust structure or redirect processes in order to maintain balance or improve an already functioning system (Gaddis, 2002; Gersick, 1991; Mintzberg, Raisinghani, & Theoret, 1976; Poole, 2004). Critical events are disruptions that demand the attention of organizational decision makers. This study identifies an event as critical if it has one or more of the following characteristics:

- Range in severity and impact on organization
- Anticipated or unanticipated
- Not dependent on time
- Sudden crisis or slow recognition of a problem
- Trigger change in structure, policy, or practice

Data Collection

This study utilizes two different types of data: organizational documents and interviews. Periodic reports released by each VRE will be used as the primary data source. The periodic reports provided by each organization will be used to construct a developmental timeline for the VRE by providing information such as technological status reports, milestones, and forecasted developments. Annual reports are a good way to analyze from a process perspective because they help to create a temporal vision of the organization's events, decisions, and milestones. In addition to written documentation, in-depth interviews with key players will be used to provide supplemental information and clarification. Interviews of key organizational personnel will be conducted to (i) get more insight and/or into already identified events, (ii) to identify missing events, and (iii) to understand the consequences of events.

Organizational Documents

Five annual reports from the VAO between the years of 2011 and 2015 will be analyzed in this study. These reports provide rich information about the developmental activities within every leg of the organization. The reports vary slightly throughout the years and this reflects the many changes that the VAO underwent during this period of time. Major sections are added or absorbed by other sections as needed. Four sections of the annual reports are consistent across the board: Executive Summary, Management, Operations, and Standards and Infrastructure. The Executive Summary provides an overview of the activities throughout the year. This section acts as a snapshot of the rest of the report, primarily highlighting important events, major milestones, and major ongoing efforts within each acting arm of the organization. Management sections discuss the construction and delivery of reports, meetings, milestones, and financial status. The foundation of this section revolves around reviewing accomplishments and important events, constructing reports, and planning. The Operations sections address the technological maintenance arm of the VAO. These sections highlight the development of scientific tools, the monitoring and validating of VAO services, and the internal operations of the VAO. The Operations section provides a report of the functionality of the various tools offered by the VAO. Finally, the Standards and Infrastructure section provides an overview of the development of international standards in collaboration with the IVOA as well as the development of an infrastructure "that supports both VAO science tools and community developers and researchers at large" (VAO Annual Report, 2012).

The following sections are generally present in each report, but they vary in location throughout the years of reports. User Support sections detail the quality assurance arm of the VAO. These activities include testing of products for user functionality and providing training and education for users. Science Applications sections discuss the innovation and development of VAO products. In this section, the reports detail progress made throughout the year for each tool as well as plans for future innovation. Annual reports also contained appendices with information about VAO participants and important presentations and publications related to the VAO throughout the year.

Information about the HTRC is addressed in the larger HathiTrust Digital Library annual reports. This study will focus on annual reports between 2011 and 2014, during the development of the HTRC. HathiTrust annual reports generally cover seven categories of information. First, the report details news and announcements. This may include events, position announcements, and awards. Second, the reports list the new universities that collaborated with the HathiTrust during the year. Third, new content is reported. This includes the addition of any new texts added to the repository, including both public domain texts and texts to which the HathiTrust gained permission to add. Fourth, the reports detail the major technological activities throughout the year. This category focuses primarily on the registry and the tools that support it. Fifth, the report highlights the activities of work groups, committees, and governing bodies. This section provides useful information about organizational structure as well as the policy and standard development within the organization. Sixth, the reports present the activities of the various special projects and initiatives of the HathiTrust. This section includes progress reports about the HTRC, amongst other projects. Finally, annual reports reviewed the various HathiTrust-related publications and presentations throughout the year.

Interviews

Interviews were conducted with key personnel in both the VAO and HTRC. These interviews were designed to garner the first hand experiences of management, policymaking, and technology leaders of each organization. Following Child's (1972) idea of dominant coalition, interview sample includes thirteen individuals (eight individuals for VAO and seven from HTRC) responsible for evaluation, planning, and implementation of strategic choices. VAO interview sample includes five management/policy making personnel, two technologists, and one responsible for managing scientific user supports. HTRC interview sample includes four management/policy making personnel, one technologist, and two responsible for coordinating projects. The interview questions were designed to measure the participants' involvement history in the organization, technology development processes, community interaction, and the nature, structure, and practices of the organization. In addition, the participants were provided the definition of critical events and were asked about their experiences in dealing with these events when they happened. Appendix A provides a list of interview questions. This list was developed to guide the discussion.

Analysis of Data

The analysis of collected data will be performed using grounded theory procedures. Grounded theory provides data-centered methods for the collection and analysis for qualitative data in order to build theories based on the data itself (Charmaz, 2006). When using grounded theory, one must first gather rich data and then analyze it using pointed coding procedures that allow for the emergence of theoretical significance. Coding is the central process of using grounded theory, as it is the "pivotal link between collecting data and developing an emergent theory to explain these data" (Charmaz, 2006, p. 46). Charmaz (2006) presents four phases of coding in grounded theory. Initial coding requires theorists to be open and flexible to the possibilities presented in the data. In this phase, researchers identify potential themes that emerge from the data in order to pursue them in later analysis. When performing initial coding, researchers might use word-by-word, line-by-line, incident-to-incident, or In Vivo (coding of unique terms identified by participants) coding procedures. Focused coding—the second phase—is a more pointed process of selecting significant trends that emerged in initial coding and using them to process larger quantities of data. Axial coding entails the construction of a framework to apply to the data based on the outcomes of focused coding. This phase "relates categories to subcategories" (p. 61) as it provides more structured analysis (Charmaz, 2006). Finally, theoretical coding phase requires theorists to make connections between the categories identified in focused coding. Theoretical codes "not only conceptualize how your substantive codes are related, but also move your analytic story in a theoretical direction" (Charmaz, 2006, p. 63). The analytic story construction will help the study to examine the research questions. Finally, the two cases will be compared based on the findings.

Events are an important element of every organization. Some events are related to day-to-day or month-to-month operations within an organization and can be predicted by key personnel while others may occur in the external environment and cause a major, unexpected shift in organizational structure, processes, or practices. However, while some events may be hugely influential for the functioning and success of an organization, they are not necessarily critical. An event becomes critical when it permeates its effects across multiple work areas in the organization, requiring accommodating efforts beyond the work area responsible for that functionality. Critical events vary in severity. Based on the event's impact on the organization, this analysis will identify two levels of severity: critical and severe impact. Organizational leaders are able to accommodate critical impact events with relative ease due to experience. While organizations can accommodate for them, serious impact critical events require intensive activities in several working parts of the organization.

In order to establish a timeline of critical events, the annual reports from the VAO and the HathiTrust were examined. Combined, the reports from the VAO included 358 pages total. Available documents for analysis of the HTRC throughout its years of operation consisted primarily of annual reports from the HathiTrust Digital Library, its parent organization. As the research arm of the HathiTrust Digital Library, the HTRC does not release independent reports. Combined, the HathiTrust reports totaled 29 pages. The procedure for identifying critical events was as follows. First, each annual report was reviewed to gain a general understanding of the sequence of events during each year. During this review, events that appeared to be important were briefly noted. Next, a spreadsheet was formulated and each noted event, including beginning and ending dates and any supporting information, was entered into the sheet. Then, transcriptions of interviews with key players in each organization were reviewed for supporting information that would help in evaluating each event to determine if it was critical. Finally, the definition and characteristics of critical events were used to evaluate each event. At this stage, critical events were labelled in order to group them and evaluated each event based on severity. Each critical event fell into one of two labels: critical or severe. While the remaining entries were significant events in the history of each organization, they were deleted if they did not meet the critical event criteria.

In-depth interviews with key personnel in the VAO and HTRC were used in the critical event identification process as well as to provide additional insights into the organizational values, processes, and procedures of each organization. The process for identifying critical events in the interviews followed many of the same steps as the analysis of the annual reports. First, transcripts of all interviews were gathered and carefully reviewed. Second, transcripts were coded to identify critical events across all of the interviews. Then, critical events were labelled in order to develop a categorization system. Finally, critical events were evaluated based on severity or impact upon the organization.

In addition, this study looked into the relationship among events from a temporal perspective in order to establish relationship among events and their influence on organizational change over time. Two types of inter-event relationships are relevant in this case, nested and entangled relationships. Events are nested when an earlier event heavily influences a later event. Events are entangled when they influence each other but are not tightly linked into a single, coherent process. In either nested or entangled relationships, events may have a positive relationship in which they reinforce each other or a negative one in which they disturb or dampen another activity.

CHAPTER 4

RESULTS

This study identified 57 critical events from the VAO and HTRC. Analysis of the history of the VAO yielded 37 critical events and the HTRC yielded 20 critical events. Following the grounded theory procedure, identified critical events were placed into groups based on similarities. A similar analysis was carried out to cluster groups based on their similarities which provided a more profound explication of the nature of these events. The first grouping was labeled as ‘sub-groups’ and the latter ‘critical event groups’. The following section details the findings of the analysis of critical events from the VAO and HTRC.

Virtual Astronomical Observatory

Six groups of critical events emerged from the analysis of the VAO history. The groups are technological failures, personnel changes, funding changes, organizational technologies, re-scoping activities, and close-out activities. The following sections provide further explication of each group and the subgroups within them.

Technological Failure

The first type that emerged from the data was that of technological failure. An unsurprising finding for a virtual organization, critical events in this group involved major errors and outages experienced by the VAO during these years. There were 11 technology failure critical events in total and these events were placed into four subgroups: internal service failure, external service failure, delay, and circumstantial interruption. Table 4 provides a list of technology failure subgroups, events within each subgroup, and brief descriptions of each event.

Personnel Changes

Throughout the duration of the VAO project, there were several resignations and replacements of personnel in key positions within the organization. Thus, the second type, personnel changes, included resignations and replacements of key personnel in each work area. There were 10 personnel change critical events and these events were placed into two subgroups: technical and management/administration. Table 5 provides a list of personnel change subgroups, events within each subgroup, and brief descriptions of each event.

Funding

The third type involved changes in funding. In early 2011, the NSF and NASA required the VAO to re-scope the project. In response to this event, the VAO had to consolidate and eliminate two major work areas. Later in 2011, the NSF and NASA notified the VAO that funding for the project would be ending earlier than anticipated, requiring the VAO to construct a close out plan. About the cuts, one respondent stated:

> We received this year a 50% budget cut because of financial limitations and we were directed by the agencies to do a couple of things, one was limit application development – instead we purely had infrastructure projects. We disagreed with this and fought this vociferously, we thought it was the wrong decision. And still do, in fact.

Two major critical events fell into this group and these events were placed in two subgroups: re-scope and close-out. Table 6 provides a list of funding subgroups, events within each subgroup, and brief descriptions of each event.

Organizational Technologies

The fourth type involved the adoption of organizational technologies and major updates made to these key technologies. The most important event in this type was the institution of an issue tracking service called JIRA, which organized and monitored VAO services. There were five organizational technology critical events and these events were placed in two subgroups: adoption and technology change. Table 7 provides a list of organizational technology subgroups, events within each subgroup, and brief descriptions of each.

Re-Scoping Activities

The fifth group of critical events includes the initial re-scoping activity of the VAO following the budget cuts initiated by the NSF and NASA. During this time, the VAO was required to make changes and cuts to the work areas. The re-scoping of the organization required the VAO to make several structural changes, which meant, "repurposing the people that you already had involved rather than bringing on new people – so some contributions changed and became greater or fewer." This group includes one critical event and this event was placed in one subgroup: reorganization. Table 8 provides further explication of the subgroup as well as the event within that subgroup.

Close-Out Activities

The close-out activities group includes all critical events that resulted from the aforementioned de-funding of the VAO. When the funding agencies notified the VAO of the close out, they requested a close out proposal from the VAO to determine where the various technological outputs would be maintained following close out. One major aspect of the close-out activity was the last major technological development for the VAO: data cubes. One participant stated:

> It really started with previous VAO board of directors…and they had defined a strategy for us, which involves one final significant development effort, which is to accommodate –-data cubes – within the (standard) data access protocol that the observatory manages and a data cube is basically – you think of an image which also has a third dimension -

> wavelength, a spectrally or time resolved image, or both, you could have a 4D structure. This is an increasingly common way of representing astronomical data, and up to this point we had plans for doing this, but had not implemented it. So this kind of formed the basis of the closeout plan – if there's a gap in the capabilities of the virtual observatory, we should fill that gap while having the technical resources available to do so."

In response to the close-out, NASA initially requested proposals from three data centers for a plan of action for supporting the VAO products, essentially asking them compete with each other. Eventually, they were asked to work together "but with some tension because one of the things that headquarters has asked them to do is to take on this activity with no increase in their budgets… So it's far from an ideal process." There are five technology transfer critical events and these events were placed in two subgroups: management and technology transfer. Table 9 provides a list of close-out activity subgroups, the events within the subgroups, and brief descriptions of each.

HathiTrust Research Center

Six groups of critical events emerged from the HTRC data. These groups are settlement, proposals, funding, personnel change, management, and technical. The following sections provide further explications of each group and the subgroups within them.

Settlements

In its early days, the HathiTrust Digital Library faced two lawsuits. The first, a 2008 suit between Authors Guild and Google in which independent authors accused Google of copyright infringement for the unauthorized scanning of books, is considered the inciting event for the research center. Second, in 2014, Authors Guild sued HathiTrust for infringement upon copyright for their use of Google Books' scanned books. There are four critical events involving settlements and these events were placed in two subgroups: suit and court ruling. Table 10 provides a list of settlement subgroups, the events within each subgroup, and brief descriptions of each event.

Proposals

The second group of critical events involves the process of submitting proposals for the HTRC. These events involve the process of legitimizing the research center by gaining approval, access, and funding for operations. About the significance of proposals, one participant stated,

> If there's some new source of funding that's become available, which may cause a project to either spin off in a different direction or take a new approach to the thing or a new view of things, then that could also trigger a change into what the future of the project will morph into.

There are three critical events in this group and these events were placed in two subgroups: decision to write proposal and accepted proposal. Table 11 provides a list of proposal subgroups, the events within each subgroup, and brief descriptions of each event.

Funding

The third group of critical events involves the receipt of funding for the research center. This group includes the funding provided by both the larger HathiTrust organization as well as external funding agencies. In academic institutions, in general, funding always plays a major role in each work area, project, and goal. One participant confirmed this stating:

> I think that you perhaps may find that as a general theme across these kinds of academic projects, you would find that the funding kind of dictates the timeline, and the funding also dictates these critical events that you mentioned. Because if the project is not well planned and well run during its course, then you will find that you have these sort of fire

drills that you get into towards – when people start realizing that, hey, we're not so far along as we would have hoped to be by this time. And, there's a deadline looming. We have to create a project report and give it to our funding agencies to justify further funding that we would like to request."

The HTRC faced the challenge of coordinating funding and balancing the contributions of each academic institution involved. One participant stated:

> We split a third now and the idea is for us to run it for five years and show what we come up with and how we're serving people and then, eventually show to HathiTrust that, you're probably gonna want to pay for more of this, then just a third. If you want to keep a viable research center because it's expensive. I mean, like even before we started getting money from them, you know, there was at least a million dollars a year or more from each institution going into making the project work.

This group includes four critical events and these events are placed in two subgroups. Table 12 provides a list of funding subgroups, the events within each subgroup, and brief descriptions of each event

Personnel Changes

The departure and hiring of key personnel is a considerable event for the HTRC. Thus, the fourth group of critical events includes changes in key HTRC personnel. Several of the participants identified significant departures and replacements as a significant impact on the organization. One participant explained the impact of hiring new personnel stating:

> If this person will truly be qualified for the role and do a good job, then you'll find that prior to this person being hired, things were sort of, kind of haphazard maybe. But, then, if this person does a good job keeping everybody united and organized and rallying

> towards the same goal, then you'll find that from that point on, things are much more coherent, and in theory, they're getting better.

This group includes four critical events and these events are placed in two subgroups: management and technical. Table 13 provides a list of personnel change subgroups, the events within each subgroup, and brief descriptions of each event

Management

Closely tied with changes in personnel, the fifth group includes changes in management practices and structure. This includes events in which changes were made to the HTRC itself and the ways in which the HTRC interacted with the HathiTrust at large. In the beginning of the project, the HathiTrust governance experienced some shifting by installing a Board of Governors and an Executive Director, which changed the way that the HTRC interacted with higher up leadership in the organization. Regarding this, one participant stated,

> That's all been – I think pretty good because I think they needed a full-time executive director previously and they didn't have it. And so it was hard to get attention of the Board and make them understand what we were doing and sometimes I think there were mixed signals on things of people not thinking the communications were going well. I think from both sides, but now that we have much more direct connection with the executive director and he's really, you know, got good management with the Board. It makes it much easier to talk about what we're doing and getting it across to people.

This group includes four critical events and these events are placed within two subgroups: decision and structural change. Table 14 provides a list of management subgroups, the events within each subgroup, and brief descriptions of each event

Technical

The fourth group involves technical aspects of the development and protection of HTRC infrastructure. This group includes the technical and security accommodations made in the shifting of in-copyright data to the HTRC. According to one participant, the shift of in-copyright data to the system marked "a new phase for the center – the center, both internally in terms of how it's structured and also in terms of its usefulness to you scholars, more generally." The process required a great deal of technical coordination in the HTRC. Another participant stated:

> We had to rethink our, not only our physical tech infrastructure but our – the roles and security. Roles really with HTRC as the pertain to security and we had to make a lot of changes about how we classify people. A lot of changes to our infrastructure in our system to make sure that we had, you know if copyright data is gonna be on this machine, this machine needs to be drastically limited as to how you can access it.

This group includes one critical event and this event is placed within one subgroup: copyright data. Table 15 provides further explication of the technical subgroup as well as the event within it.

Despite the fact that the VAO and the HTRC are disciplinarily different organizations, both are marked by several similar critical events that shape the evolution of each entity. Since both organizations are academic institutions, funding is a constant focus and struggle in each organization. Since the VAO project was sent into a close-out at such an early stage, the critical events surrounding the two main funding events (re-scoping and close-out) brought forth several other critical events focused on shifting structure and downsizing work areas to accommodate for the decrease in funding. This is the main area of divergence between the two organizations. Despite the important role funding played in each organization, the fact that the VAO went through a close-out marks a significant deviation between the two in terms of critical events that emerged. The HTRC has not experienced a close-out, so the activities of the organization are largely focused on attaining grants and awards from outside organizations for their work. Like most, if not all, organizations, both the VAO and the HTRC experienced personnel changes in key areas of the organization. Turnover is to be expected in academic organizations such as these. Academic professionals who hold positions in outside academic institutions largely comprise both organizations and, thus, often move in and out of part-time projects as priorities change.

Identified Domain Specific Needs and Priorities

The review of literature on scientific collaboration and disciplinary differences posits that each scientific domain should have its own needs and priorities. The section below presents the findings identified form document analysis and interviews.

Virtual Astronomical Observatory

Technological needs. VAO developers De Young, Hanisch, Szalay, Berriman, and Fabiano (2010) produced a paper with the history and vision of the VAO project. After the new Millennium, the astronomical community was rife with discussions of developing a virtual observatory. A conceptualization of the United States National Virtual Observatory (NVO) was produced in 2000 and, by 2002 the NVO was initiated. The NVO laid the groundwork for the eventual development of a functional virtual observatory, providing both the environment for the technical infrastructure of the virtual observatory as well as the organizational structure for developing, testing, and implementing the technologies. The final months of the NVO project funding were used to move toward an operational observatory, the product of which was the VAO (De Young et al., 2010).

Keeping the technological needs of the astronomical community in mind, the VAO developed several science applications: Data Discovery Tool, Interoperable SED Access and Analysis tool (also known as Iris), the Catalog Cross Comparison Service, and a prototype of the Time Series Search tool. The VAO infrastructure included the following: the JIRA ticketing system, testing services, the Subversion (SVN) code repository, YouTube channel, blog, and mailing lists. The JIRA issue tracking service was key in VAO operations as they helped to track and monitor the progress of VAO activities. JIRA was central to the development of scientific applications, as it helped to track the project life cycle and software issues discovered after testing. The SVN repository is a versioning system that provided a place for the code and data developed by the VAO. Finally, the VAO developed a close-out repository using Google cloud services to store VAO resources for the continuation of work after the close-out. In addition to JIRA, the VAO used several communication technologies for their work including e-mail, wiki, and video/teleconference capabilities such as WebX and Skype.

The technological needs of the VAO highlighted the organization's focus on collaboration and the ability to track the progress being made throughout the project. Tools like JIRA and wikis help organizational members to stay up to date on their own progress as well as the work of others.

Community needs. De Young et al. (2010) detailed the guiding principles of the VAO. First, the VAO would aim to enable efficient and effective research, allowing researchers to produce better research in both breadth and depth. Second, the VAO would serve as a tool for both education for and recruitment of interested students. Third, addressing concerns about scientific awareness in larger society, developers hoped the VAO would reach a large public audience. De Young et al. (2010) stated, "Because astronomy is a science with widespread public interest and support, it is an ideal venue for the creation of scientific programs that will have public appeal and that will convey concepts of scientific methodology in addition to specific scientific information" (p. 3). Fourth, the VAO would open doors to international collaboration and leadership. The development of a virtual observatory had been an international goal since the conception of the idea. In fact, the International Virtual Astronomical Observatory (IVOA) acted as "the international forum for establishing standards, maintaining communication, and fostering collaboration among the various VO projects around the world" (De Young et al., 2010, p. 3). Thus, the VAO was developed to participate and cooperate with the IVOA. The final goal of the VAO was to facilitate activities between other initiatives.

The VAO itself was inherently community-focused as it kept the astronomical community involved even in the development of an infrastructure. According to Evans et al. (2010), "As an ecosystem, the implied architecture is one that is open, loosely-coupled and yet smoothly connected, bringing together VAO-provided tools with top-quality tools developed by the wider Virtual Observatory (VO) community. In particular, VAO science tools provide key end-user capabilities that make the environment usable but also pave the way for community-contributed tools" (n.p.). Interaction and engagement with the astronomical community was a key aspect of the project and, when the funding agencies suggested cutting areas of the organization that facilitated the relationship between the VAO and the community, VAO leaders pushed back. One participant stated:

> We felt that this was totally inappropriate and was going to be a real detriment to any follow up organization to basically have to start all over again in establishing contact with the research community. So, in consultation with the board, we responded to things by making some revisions to the PEP [Project Execution Plan]. For example, we relabeled

> professional engagement as Documentation Services…and then we resubmitted the PEP to the agencies.

In order to maintain engagement with the community, the VAO personnel shifted work area labels to maintain the engagement set forth in their organizational goals. VAO personnel criticized astronomical entities that did not seek out the input from the broader community of astronomers when making decisions. One individual stated:

> They [IVOA] have a closed community and they haven't engaged the end users who are the world's astronomers and asked them what their requirements and specifications are. Ask them to the meeting. There is a big data provider, a new project coming around, ask them to a meeting. Say, "Give us a talk on your project. What do you need?" And they can sit down and say, "We can do this and this. Should we pursue this further?" Or we can say, "Oh, I had no idea you guys needed to do this. We need to work on this." There has been very little of that...But as a result of them not engaging users the IVOA has got a reputation of this massive bureaucracy that tries to solve complex data access problems through pure thought. And which tries to decide for a community how am I going to represent their data? And that does not sit well with a lot of people.

Since its inception, the VAO was always focused on communicating and collaborating with the astronomical community. As this organization was largely made up of astronomers, it was vital to the leaders of the organization to be able to establish connection and engagement with the user community. When this value was threatened due to funding cuts, the organization members pushed back and worked to shift work areas to accommodate the decrease in funding while still maintaining relationships with the user community.

From VAO to IVOA. Despite the defunding of the VAO, the close-out allowed the organization to formulate a plan for the future of the products developed during the project. In the close-out plan, the various entities of the VAO were distributed to different groups for long-term curation and the preservation of the legacy of the VAO. VAO assets were transferred to the NASA archived, which also assumed the VAO role in the development of international standards for virtual astronomy. Many of the tools and services were maintained by the VAO partnered organizations that developed them. In many instances, the organizations responsible for these entities continued to develop them after funding ended. One participant stated:

> The first thing is you; it won't be done by the time the funding terminates. You hope to be started, so what you want to do is put to a useable point where things are in production use, you have a product and you're doing things, but that will only be in the beginning. With any kind of science software when you start using it, you're only getting started, if it's actually used and starts to get used for science then there will be lots of features and stuff that you'll want to add. Support for new data and so forth, so if we're successful and the goal is to get it in use enough with all this seed money and effort and so forth. That it continues on and was worth doing what you want to carry along far enough so that the projects successful and is piece of software that they get to use and maybe hopefully grows into something more successful. That's how all these projects are…We're successful if it's done enough and deployed, when it's in production use and continues to be developed after the VAO funding ends. It would probably be developed by the institutions that are using it.

Even though the VAO closed after only a few years of operation, the legacy of the organization has been maintained through the continued development of the various tools and applications developed by the major groups in the organization. In addition, VAO tools continue to be used by astronomers all over the world.

HathiTrust Research Center

Technological needs. As previously stated, the HTRC grew out of the Google Books lawsuit and was pursued by the HathiTrust Digital Library in order to create a repository of published content that will both serve the needs of traditional research as well as computing-intensive research in the humanities and social sciences. One participant described the technological aim of the organization, stating,

> One-third of the collection is in the public domain…The remainder, let's call it 70% or 67%, is under copyright, which means that while you can build indexes against it, we cannot distribute this text…What it [the HTRC] is designed to do is to provide what is known as non-consumptive research access to the remaining text, to that 70% [that] is under lock and key because of copyright issues. By non-consumptive, it means we're building systems to allow researchers to create algorithms, analytic programs that they would submit to us at the HTRC to run against those big data…We will run the analysis for them, and they will get the answers back…They never read the text, they never get the text, they just get their answers.

The HTRC currently has three tools available to do non-consumptive research. First, the HTRC Portal and Workset Builder allow researchers to gather and analyze collections of electronic texts. Second, HathiTrust+Bookworm is a tool that enables the visualization and analysis of trends in word usage in texts within the HathiTrust Digital Library. Finally, the HTRC Data Capsule is a secure research environment for conducting researcher-driven analysis on texts within the HathiTrust Digital Library. The HTRC personnel use several technologies to communicate and collaborate with one another including video and teleconferencing, screen sharing, data sharing applications like Dropbox, wiki, listservs, and messaging applications like Slack.

Community needs. When discussions of developing to HTRC began, the working group spoke with researchers from different disciplines and institutions to garner the needs of the potential user community. In their proposal, the Executive Committee of the HTRC identified three major categories of research needs. The first need of the community was the ability to assemble a large amount of texts on a particular topic and to extract data in order to do some particular task. The Executive Committee provided several examples of this need including the ability to assemble "references to people, places, and things across time, languages, and locations to study 'big' questions of history." The second major need of the community was the development of research tools to be able to accomplish the first need. Finally, due to the increasing availability of electronic data, researchers needed the ability to collaborate with one another more than ever. An example scenario identified by the community included the ability to find out "how other scholars have used the Research Center data, including what tools they employed, what texts they searched, what secondary data they created, and whether this data is accessible."

The HTRC has continued to work closely with the user community in their continued development of the research center. The HTRC prioritizes interaction with researchers by "fostering collaboration and connection and going out and working with researchers to see what they want." The HTRC holds an annual event called UnCamp geared toward community engagement and outreach. One participant explained UnCamps stating, "We've invited a number of external researchers and pretty much anybody that was interested to sign up and come and learn about the HTRC and the effort that we're putting forward. And through such venues, we've been able to engage in discussions."

HathiTrust investment in a research center. The HathiTrust partners sought to develop a research center in order to assist with the growing need for big data research in the humanities and social sciences. HathiTrust Digital Library had a particular advantage in developing this type of research center because, while the process for awarding organizations a research center was not yet determined, the HathiTrust leaders ensured that a research center would be awarded to them due to the large body of scanned text that was provided by the HathiTrust partners. According to the proposal for the development of a research center, the Executive Committee argued:

> By positioning ourselves to take on the responsibility of managing a Settlement-enabled Research Center, in addition to supporting research capabilities across the HathiTrust Corpus itself, the partners can ensure the vitality of HathiTrust as a data provider in this new research environment and further our efforts to ensure, through cooperative means, the efficient management, preservation and accessibility of the scholarly record over time.

These findings provide interesting insights into the organizational processes and practices of the VAO and HTRC within their given disciplines. The main difference in technological needs between the VAO and the HTRC is the use of technologies for development. The VAO used JIRA as a tool for tracking progress throughout the various stages of development, while the HTRC has not adopted a similar technology as of yet. There is talk among the organization members of utilizing tracking technologies like JIRA in the HTRC, but it has not been established at the current time. Both the VAO and the HTRC greatly valued community and engagement and emphasized how fostering these relationships was vital to producing quality products. Finally, it comes as no surprise that the futures of each organization at their current stage are vastly different. Due to the close-out, VAO assets were distributed to various partner organizations for maintenance and continued development. After the close-out, the VAO itself no longer operates as a whole, but rather as a grouping of interconnected partners operating under the same set of standards set forth by the IVOA. The HTRC is in a strikingly different phase of development as they work towards establishing themselves as a viable resource for the research community. The HTRC personnel continue to venture forward in pursuit of providing a space for researchers in the humanities and social sciences to perform innovative computational research. While both organizations use different procedures, practices, and tools to perform their work, both are driven and motivated by meeting the needs and expectations of their respective research communities, even in the midst of financial hardship.

CHAPTER 5

DISCUSSION

The objective of this study was to identify how critical events shape the evolutionary processes of two virtual research organizations. This study also explored the relationship between discipline and the nature of critical events that emerged throughout the lifetime of the organizations by analyzing the histories of a hard science and a social science VRE. Mintzberg's (1978) strategic decision making and Child's (1972) strategic choice perspectives provided a useful theoretical framework for analyzing organization evolution through critical events because it promotes the dominant coalition's role in enacting change in structure, process, or policy to accommodate for critical events.

In this study, two research questions were explored to understand the nature of critical events and how they function as a key element of organizational evolution. Findings of the study provided insight into both the questions. In addition, the theoretical framework of strategic choice in decision-making was also supported by the results.

The Role of Critical Events in Organizational Evolution

Child (1972) argued that existing models of organizational change that identified characteristics of organizations such as structure or environment with the evolution of organizations. Child's answer to the weaknesses of prior models of organizational change was the strategic choice perspective. Child (1972) argued that organizations evolve due to the strategic choices of the dominant coalition. The strategic choice does not deny the importance of structure or environmental events, but rather emphasizes the agency of the dominant coalition in making choices about these things. An analysis of critical events as an integral element of organizational change without examining the implications of strategic choices would be insufficient in its lack of acknowledgement of the complexity of organizational change. Thus, the strategic choice perspective supports the notion that critical events along with the strategic decisions that organizational leaders make to accommodate for them are the key elements of organizational change and evolution.

The findings of this study indicated that there are two major types of critical events in organizations: critical events and severe critical events. Critical events emerge as a natural process of organizational operation. While these events require the dominant coalition to make strategic choices about how to proceed, members understand that these types of disruptions happen due to the nature of the organization and are equipped to act upon them with relative ease. An example of a critical event from the VAO would be the decline in compliance from a VO data provider. The operations team noted the decline and notified the data provider of the issue. The data provider checked into the problem and resolved it by correcting their software. Since the VAO is an organization based on producing technological tools, organization members understand and anticipate technological errors. The strategic choice to formulate a work area to monitor operations of the VAO system from the outset indicates the anticipation of errors such as this. Because this work area was in place, the issue was resolved without major disruption to the entire organization.

Unlike critical events, severe critical events emerge due to a unique situation that is not necessarily expected by organizational members. These events usually require considerable accommodation, especially if the organization is not well equipped to handle the nature of the event. The participants identified the legal events surrounding the HTRC as a significant and unique event. The original settlement of the suit between Authors Guild and Google provided funds for the development of a research center. The settlement was considered the inciting event of the HTRC, but the settlement was dismissed by the court. The dominant coalition of the HTRC made the decision to proceed with the research center despite the fallout of the settlement and seek out funding from other sources. The original lawsuit continued to influence the HTRC despite being somewhat inconsequential in providing actual resources to the organization. The activities of the HTRC were allowed to continue because of the ongoing debate about the value of executing research on both public domain and in-copyright data.

Similarities between VAO and HTRC in Critical Events

It was evident from the findings that both organizations had faced critical events regularly. The nature of these events were sometimes similar experiences mainly due to the academic nature of both the organizations.

The results showed several major personnel changes in both the VAO and HTRC. As in any organization, sometimes these resignations were natural, expected, or unsurprising. For example, after serving the VAO for some time, the individual in the Project Scientist role decided to retire. His retirement was announced some time before he left the organization, giving the dominant coalition a significant period to fill his role. While some of these departures tended to be critical events, others were be extremely severe and disruptive to the organization, especially when a key member of the organization unexpectedly resigned in the midst of an important project. This occurred when the senior architect of the HTRC abruptly resigned during the discussion about shifting in-copyright data into the HTRC system. Due to the unexpected departure of an individual in an integral role, HTRC was forced to find a temporary replacement until a more permanent solution could be made. This departure affected the entire team working with copyright issues. While both organizations experienced personnel changes throughout their tenure, there were varying levels of impact on the organization due to the nature of the departure.

While the circumstances surrounding funding for the VAO and HTRC were vastly different, the overall importance of funding was another similarity between the nature of critical events of both organizations. In conversations with personnel from both organizations, regardless of the individual's specialty or work area, it became apparent that funding was the thread that ran through every aspect of each organization. In both organizations, funding affected most decisions made by the dominant coalition. Many of the major evolutions of the VAO and HTRC could be observed by tracing funding events in and of themselves. Despite the similarity in the general importance of funding on the operations of the organizations, the ways in which each organization approached funding was very different due to situational events.

Differences between VAO and HTRC in Critical Events

While many organizations will experience similar critical events throughout their evolution, there will be differences that will depend on the nature of the organization. This study sought to understand how two academic virtual organizations evolve in different disciplines. The results of the study showed significant differences in the critical events that emerged from the VAO and HTRC.

While the critical events of both organizations showed the high importance of funding, the circumstances surrounding the organizations were vastly different, influencing the nature of the funding events. The VAO was supported by and responsible to two major funding agencies: NSF and NASA. Unlike the HTRC, organization members in the VAO were not seeking out grants and awards from smaller private funding agencies. In addition, the NSF and NASA maintained a level of ownership over the products that the VAO developed and the VAO sent the funding agencies Project Execution Plans (PEP) every year to maintain support. The HathiTrust

Digital Library both houses and provides some funding to the HTRC, but members of the HTRC constantly wrote and submitted proposals for funding from private organizations.

The VAO and HTRC faced two major critical events related to funding that signified a massive divide between the two. First, the VAO was issued a decision by the funding agencies to close the program before the end of the funding period. This critical event influenced all of the critical events that followed because the VAO had to make several strategic choices about how to proceed and divide VAO assets in a way that would best maintain the integrity and usability of the tools. The choices VAO made were domain specific in the way that organizational responsibilities were divided up among the astronomical communities within the United States and European countries. Strategic choices that kept scientific computational resources operational during the shutdown period were purely based on the configuration and strengths of astronomical community members. The existing collaboration within the astronomical community was very high and this was highlighted by the ease with which the dominant coalition of astronomers distributed VAO assets to various partner organizations. This level of synergy was not an emergent factor of HTRC operations. Second, as previously discussed, the HTRC was massively influenced by the two lawsuits instigated by the Authors Guild. The HTRC was developed as a result of the settlement between the Authors Guild and Google and, when the settlement was not realized, the dominant coalition realized the need for computational tools for the humanities and social sciences and decided to proceed anyway.

The second major difference in the critical events that emerged from the data was the nature of technical events. The VAO annual reports provided significant detail about both the major and minor technical problems that emerged during the given time period. While it is highly likely that the HTRC has experienced technical hiccups throughout the tool development process, the reports showed far less detail about technical progress. There are two reasons that may explain this differential. First, the VAO was significantly more structured in their tool development processes and placed more importance on tracking and reporting issues through the JIRA service. It was due to the long collaboration and technology use history of the discipline. According to the HTRC personnel, the tool development process was somewhat disorganized and still very much in flux and, while the adoption of tools like JIRA was on the horizon, they have not been utilized yet. In the given data, there was no record of specific technical issues during the tool development process. Second, since the HTRC updates were situated within the larger HathiTrust annual reports, perhaps technical information was withheld because it may not be of interest to the audience of readers within humanities and social science disciplines.

The impact of the technologies developed by each organization reflected additional disciplinary differences between the VAO and HTRC. Federal funding organizations—NSF and NASA—played a massive role in the ways in which the VAO planned and executed their technologies. Additionally, the technologies developed by the VAO were calibrated based on international standards instituted by the IVOA. The IVOA developed these standards to allow for the development of their data depository that astronomers around the world could both access and analyze simultaneously. The HTRC, on the other hand, was largely responsible to the HathiTrust and developed their technologies based on their own evaluations of the interests and needs of the community without having concern for standardization. The HTRC produced technologies for the non-consumptive analysis of textual data. Thus, the HTRC was also responsible for maintaining awareness of the legal ramifications of developing tools to analyze in-copyright data.

The findings provided significant insight into the ways two disciplinarily different organizations evolved through time and how critical events that occurred both within and outside of the organizational environment influenced evolution. While some similarities between the organizations emerged in the study, many of the critical events that emerged were specific to the organization because of the nature of the discipline and the circumstances surrounding the organization at the time.

CHAPTER 6

CONCLUSION

This study examined the development and evolution of two virtual research organizations in different disciplines through the identification and analysis of critical events and strategic choices. The findings provide several insights about the nature of critical events as a major element of organizational change.

First, the study shows that, virtual research organizations will face critical events that require strategic decision-making and these critical events will shape and punctuate the evolution of the VRE. Both the VAO and HTRC exhibited critical events that influenced the different working parts of the organization. Second, critical events can be a normal incident upon which the organization is well equipped to act due to the nature of the organization. On the other hand, some critical events arise due to a special situation or circumstance that is not necessarily expected. Third, in a similar vein, the situations that trigger a critical event can be internal or external to the organization. Fourth, the findings show that some critical events are discipline specific. Finally, the findings support strategic decision-making and strategic choice perspectives. The results of the study show that critical events require dominant coalitions to make strategic choices about how to proceed. In many cases, these choices resulted in additional critical events.

There were few limitations of the study. First, the study only dealt with two samples. Although both sample organizations were significant in the VRE world, the study does not capture the complexity of all ranges of VRE. Having more samples from hard and social sciences would strengthen the result. Second, the nature of documents from these organizations was different. VAO documents were specific and detail, whereas there were lack of details in HTRC annual reports. Third, VAO and HTRC were at different evolutionary trajectory during the study period. This situation was at the same time enabling and constraining for the study. During the course of study, VAO was going through a close-out period while HTRC was thriving as an organization. Although, it created a difference in data and interview responses, it also enabled us to capture critical events from both the beginning and ending phase of VREs.

This study provides insights into organizational change in an area that is relatively unexplored by organizational communication scholars up to this point. Computational research is becoming more prevalent within not only the physical, biological, or natural sciences, but also in the humanities and social sciences. Thus, it is important for communication scholars to investigate how organizations developing these VREs make strategic choices when faced with environmental disturbances. Future research should examine virtual research organizations within more polarized disciplines in order to develop a deeper understanding of the influence of disciplinary differences in organizational evolution. Future studies would also benefit from examining organizations that are in a relatively similar place within the organizational life cycle. Doing so would help to standardize findings in order to move towards a theoretical model of organizational evolution.

Table 1: Factors contributing to scientific collaboration

Factors contributing to scientific collaboration	**Definitions**	**References**
Proximity	The degree of physical separation between two individuals	Finholt & Olsen, 1997; Kraut, Fussell, Brennan, & Siegel, 2012
Productivity	The quantity and/or quality of the output of work	Lee & Bozeman, 2005; Pao, 1982; Pelz & Andrews, 1976; Price & Beaver, 1966; Zuckerman, 1967
Communication	The act of verbally or nonverbally sharing information between two or more people	Cummings & Keisler, 2005; Maglaughlin & Sonnenwald, 2005; Olson & Olson, 2000
Work ethic	Intrinsic motivation translated into effort to achieve some goal or reward	Pelz & Andrews, 1976
Conflict and Negotiation	Disagreement between individuals in which the parties involved must deliberate in order to repair the functionality of the relationship	Shrum, Genuth, & Chompalov, 2007
Competition	The motivation to reach a higher level of achievement and reward than a peer	Figg et al., 2006; McCain, 2010
Community	An in-group or a collection of people with a common trait who interact and collaborate with one another	Price & Beaver, 1966
Technology	Innovative and practical tools that assist an individual in accomplishing some task	Sahu, 2013
Organization	A structured collection of people and groups who utilize resources and maintain shared goals in order to work towards some outcome	IMNRC, 2002

Table 2: VAO Annual Report Content

Document	**Usual Content**
Executive Summary	Provides a snapshot of the overall progress made by the organization throughout the year
Management	Provides information about reports, meetings, and plans
Operations	Provides information about the maintenance and validation of technological tools
Standards and Infrastructure	Provides developmental information about international standards and an infrastructure that supports scientific tools and users
User Support	Provides information about quality assurance and testing of VAO tools as well as training and education for users
Science Applications	Provides information about the innovative arm of the VAO responsible for the development of new technologies
Participants	Lists the active participants within the VAO throughout the year
Presentations and Publications	Highlights the VAO-related publications and presentations throughout the year

Table 3: HathiTrust Annual Report Content

Document	**Usual Content**
News and Announcements	Highlights major happenings relevant to the organization that year
New Partners	Reports the addition of new universities to the HathiTrust
New Content	Reports the addition of new texts to the HathiTrust repository
Technology Updates	Reports major technological activities of the HathiTrust repository
Group Updates	Provides updates from each work group, committee, and governing body that makes up the organization.
Special Projects and Initiatives	Reports the activities of various special projects and initiatives including the HTRC
Papers and Presentations	Highlights the HathiTrust-related publications and presentations throughout the year

Table 4: VAO Technology Failure Critical Events

Critical Event Group	Critical Event Sub-Group	Critical Event	Critical Event Description
Technology Failure	Internal service failure	Missing records	Missing records in the VAO data registry/directory service
		Legacy footprint problem	Problem with the legacy footprint services that preventing simple queries from working
		Services failed to pass tests	Approximately 75% of VO services tested by validators did not pass without errors or warnings
		Spam issue	Spam problem on VAO Forum
	External service failure	HEASARC problem	Problem with HEASARC services noticed
		Decline in compliance	Decline in compliance from one of the VO data providers
		JIRA failure	Massive JIRA failure that depressed uptime metrics for the first half of the year
		Pass rate failure	Pass-rate for VizieR plunged towards 0
	Delay	Delay in release	SCCT delay in release due to documentation issues
		Plateau in science requests	A dramatic surge in Registry service queries caused a high plateau in science requests
	Circumstantial interruption	Federal shutdown interruptions	Federal shutdown caused major interruption to the VAO's science services, with the average availability of science services dropping below 90%

Table 5: VAO Personnel Changes Critical Events

Critical Event Group	**Critical Event Sub-Group**	**Critical Event**	**Critical Event Description**
Personnel Changes	Science/Technology	Task lead for standards	Departure of the lead personnel for Standards and Protocols work area
		Replacement of task lead for Standards	Personnel replacement of task lead for Standards and Protocol work area
		Lead of User Support	Departure of the lead personnel for User Support work area
		Project scientist	Resignation of project scientist
		Replacement of project scientist	Personnel replacement of project scientist
		Replacement of lead for User Support	Personnel replacement of lead for User Support work area
	Management/Administration	Business manager	Resignation of business manager
		Replacement of business manager	Personnel replacement of business manager
		Director	Resignation of director of the VAO project
		Personnel assumes role of director	Personnel replaces director as a formal representative of the VAO project

Table 6: VAO Funding Change Critical Events

Critical Event Group	**Critical Event Sub-Group**	**Critical Event**	**Critical Event Description**
Funding Change	Re-Scope	VAO re-scoped	VAO downsized and reorganized in response to redirection from NSF and NASA
	Close-Out	VAO close-out	Project terminated by NSF and NASA

Table 7: VAO Organizational Technology Critical Events

Critical Event Group	Critical Event Sub-Group	Critical Event	Critical Event Description
Organizational Technology	Adoption	JIRA release	JIRA service released for issue tracking
		VAO blog created	VAO blog created to allow for discussion of issues where the wiki format was not appropriate
		Google code	Decision to use Google code as the primary repository
	Technology change	JIRA upgrade	Substantial update made to JIRA
		Wiki moved	Wiki transferred to be co-hosted with the main VAO website

Table 8: VAO Re-scoping Activities Critical Events

Critical Event Group	**Critical Event Sub-Group**	**Critical Event**	**Critical Event Description**
Re-scoping Activities	Re-organization	Work areas eliminated	Data Curation and Preservation; Technology Assessment; and Education and Public Outreach tasks were discontinued when the project was re-scoped due to budget cuts

Table 9: VAO Close-out Activities Critical Events

Critical Event Group	Critical Event Sub-Group	Critical Event	Critical Event Description
Close-out Activities	Management	Data Cube decision	Board of directors agrees to take on development of data cubes as final significant development
		Work area reorganization	Standards and Infrastructure reorganized team according to revised task areas (Standards and Protocols/Product development)
		Work area changed	User Support work area changed to Documentation Services to support close out
		Support proposal submitted	Proposal submitted to NASA for plan to support the essential components of the VAO-developed infrastructure following close-out
	Technology Transfer	Transfer of VAO communication assets	Scheduled transfer of VAO website and mailing lists following close-out
		Transfer of IVOA communication assets	Scheduled transfer of IVOA website, wiki, and mailing lists following close-out
		Transfer of responsibility for publishing tools	Scheduled transfer of support maintenance role for publishing registries and resource publishing tool following close-out
		Transfer of responsibility for data sharing tool	Scheduled transfer of maintenance role for the data sharing tool following close-out

Table 10: HTRC Settlement Critical Events

Critical Event Group	Critical Event Sub-Group	Critical Event	Critical Event Description
Settlement	Suit	Authors Guild vs. Google Books	Authors Guild sued Google Books for copyright infringement
		Authors Guild vs. HathiTrust	Authors Guild sued HathiTrust for copyright infringement for the use of Google Books scanned materials.
	Court ruling	Settlement fallout	After back and forth over settlements, court dismissed the case, ruling in favor of Google Books.
		Court ruled in favor of HathiTrust	The court ruled that HathiTrust use of Google Books materials constituted fair use.

Table 11: HTRC Proposal Critical Events

Critical Event Group	**Critical Event Sub-Group**	**Critical Event**	**Critical Event Description**
Proposals	Decision to write proposal	Initial proposal	Decision to submit a proposal for the research center
	Accepted proposal	RFP accepted	Request for proposal for research center accepted
		MOU accepted	Memorandum of understanding accepted

Table 12: HTRC Funding Critical Events

Critical Event Group	**Critical Event Sub-Group**	**Critical Event**	**Critical Event Description**
Funding	Internal funding	HathiTrust funding	HathiTrust Board of Governors allocated nearly $1 million for four years to support HTRC
	External funding	Sloan Foundation funding	Initial grant of $300k for three years to the HTRC for the development of a data capsule
		National Endowment for the Humanities Funding	$324k grant for the project "Exploring the Billions and Billions of Words in the HathiTrust Corpus: HathiTrust+Bookworm"
		Andrew W. Mellon Foundation award	$1.7 million award for two years for the "Workset Creation for Scholarly Analysis + Data Capsules: Laying the Foundation for Secure Copyrighted Data in the HathiTrust Research Center, Phase I (WCSA+DC)" project

Table 13: HTRC Personnel Change Critical Events

Critical Event Group	**Critical Event Sub-Group**	**Critical Event**	**Critical Event Description**
Personnel Change	Management	Co-Director leaves project	Co-director leaves the HTRC when he was no longer needed to help organizationally
		Replacement of co-director	Co-director position filled
		Member of executive committee changes jobs	Key member of executive committee changes jobs and location
	Technical	Senior architect resignation	Senior architect leaves project during copyright conversations, leaving gap.

Table 14: HTRC Management Critical Events

Critical Event Group	**Critical Event Sub-Group**	**Critical Event**	**Critical Event Description**
Management	Decision	Decision to obtain partners for research center	Initial decision to find partners to help with the proposal and execution of the research center
	Structural change	Change in governance model	Establishment of HathiTrust Board of Governors, to whom the HTRC executive committee reports
		Full time executive director hired	Executive director established as liason between HTRC and HathiTrust Board of Governors
		Change from idea to operational	Management shift of co-directors when program required more operational leadership

Table 15: HTRC Technical Critical Events

Critical Event Group	**Critical Event Sub-Group**	**Critical Event**	**Critical Event Description**
Technical	Copyright data	Shifting data	Shifting of in-copyright data to the HTRC. Required technical and security review.

APPENDIX

INTERVIEW QUESTIONS

Can you tell me a little about yourself and your role in the development of the VAO/HTRC?

Has your role changed over time?

What were the overall goals of the organization? Have they changed?

Can you tell me a little bit about the communication and coordination practices in the VAO/HTRC?

What are the major tools available through the VAO/HTRC? How were they developed? How did they change?

Have there been any significant technical difficulties during the development of these tools?

How do you evaluate needs of the user community in terms of determining what kinds of tools to develop?

What kinds of products are in the pipeline for the future?

Do you think that the VAO/HTRC has its own unique organizational identity? If so, when did this identity emerge?

What is the nature of your organizational structures? Did they change throughout the years?

Can you think of any critical events in the history of the VAO/HTRC?

Have there been any major personnel departures during your time with the VAO/HTRC?

How do you all manage working with others from a distance?

Have there been any challenges in working with people with difference disciplinary backgrounds?

How do you envision the future of the HTRC? (not applicable to the VAO)

REFERENCES

Ahmed, I, Poole, M.S., Cooper, K., & Thomer, A. (2016). Virtual Research Environments, Communication, and Scientific Collaboration. Unpublished manuscript.

Ahmed, I., & Poole, M. S. (2012). A process method approach to study the development of virtual research environments: A theoretical framework.

Anandarajan, A., & Anandarajan, M. (2010). An overview of e-Research collaboration. In M. Anandarajan, & A. Anandarajan (Eds.), *e-Research collaboration: Theory, techniques and challenges* (pp. 3-13). New York: Springer.

Atkins Report. (2003). *Revolutionizing science and engineering: Report of the National Science Foundation Blue-Ribbon Advisory Panel on cyberinfrastructure.*

Australian Government, Department of Education and Training. (2015). *National Collaborative Research Infrastructure Strategy (NCRIS)*. Retrieved from Australian Government, Department of Education and Training: https://education.gov.au/national-collaborative-research-infrastructure-strategy-ncris

Babchuk, N., Keith, B., & Peters, G. (1999). Collaboration in sociology and other scientific disciplines: A comparative trend analysis of scholarship in the social, physical, and mathematical sciences. *The American Sociologist, 30*(3), 5-21. doi:10.1007/s12108-999-1007-5

Beaulieu, A., & Wouters, P. (2009). e-Research as intervention. In N. W. Jankowski (Ed.), *e-Research: Transformation in scholarly practice* (pp. 54-72). New York: Taylor & Francis.

Bødker, S., & Grønbæk, K. (1991). Cooperative prototyping: Users and designers in mutual activity. *International Journal of Man-Machine Studies, 34*(3), 453-478. doi:10.1016/0020-7373(91)90030-B

Borgman, C. L. (2007). *Scholarship in the digital age: Information, infrastructure and the Internet.* Cambridge, MA: MIT Press.

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

Braun, T., Glanzel, W., & Schubert, A. (2001). Publication and cooperation patterns of the authors of neuroscience journals. *Scientometrics*, *51*, 499-510. doi:10.1023/A:1019643002560

Brunvand, S., & Duran, M. (2010). A taxonomy of e-Research collaboration tools: Using Web 2.0 to connect, collaborate and create with research partners. In M. Anandarajan, & A. Anandarajan (Eds.), *e-Research collaboration: Theory, techniques and challenges* (pp. 105-116). New York: Springer.

Carusi, A., & Reimer, T. (2010). *Virtual research environment collaborative landscape study.* Bristol: JISC.

Charmaz, K. (2006). *Constructing grounded theory: A practical guide through qualitative analysis* . London: Sage Publications.

Child, J. (1972). Organizational structure, environment and performance: The role of strategic choice. *Sociology, 6*(1), 1-22. doi:10.1177/003803857200600101

Child, J. (1997). Strategic choice in the analysis of action, structure, organizations and environment: Retrospect and prospect. *Organization studies*, *18*(1), 43-76. doi:10.1177/017084069701800104

Cole, J. R., & Cole, S. (1973). *Social stratification in science.* Chicago: The University of Chicago Press.

Contractor, N. (2007). *From disasters to WoW: Enabling communities with cyberinfrastructure.* National Centre for e-Social Science.

Cummings, J. N., & Keisler, S. (2005). Collaborative research across disciplinary and organizational boundaries. *Social Studies of Science, 35*(5), 703-722. doi:10.1177/0306312705055535

Datta, A., Rzadca, K., Ang, S., & Hong, G. C. (2010). Serverless social software for nomadic collaboration. In M. Anandarajan, & A. Anandarajan (Eds.), *e-Research collaboration: Theory, techniques and challenges* (pp. 85-104). New York: Springer.

David, P. A. (2004). *Towards a cyberinfrastructure for enhanced scientific collaboration: Providing its 'soft' foundations may be the hardest part.* Oxford Internet Institute. The University of Oxford.

De Roure, D., Goble, C., & Stevens, R. (2008). The design and realisation of the myExperiment virtual research environment for social sharing workflows. *Future Generation Computer Systems, 25*, 1-15.

De Young, D., Hanisch, R. J., Berriman, G. B., & Fabbiano, G. (2010). A vision for the US Virtual Astronomical Observatory.

Dean, J. W., & Sharfman, M. P. (1996). Does decision process matter?: A study of strategic decision-making effectiveness. *The Academy of Management Journal, 39*(2), 368-396. doi:10.2307/256784

Duque, R. B., Ynalvez, M., Sooryamoorthy, R., Mbatia, P., Dzorgbo, D. S., & Shrum, W. (2005). Collaboration paradox: Scientific productivity, the Internet, and problems of research in developing areas. *Social Studies of Science, 35*(5), 755-785. doi:10.1177/0306312705053048

Dutton, W. H., & Jeffreys, P. W. (2010). World wide research: An introduction. In W. H. Dutton, & P. W. Jeffreys (Eds.), *World wide research: Reshaping the sciences and humanities* (pp. 1-17). Cambridge, MA: MIT Press.

Dutton, W. H., & Meyer, E. T. (2010). Enabling or mediating the social sciences? The opportunities and risks of bottom-up innovation. In W. H. Dutton, & P. W. Jeffreys (Eds.), *World wide research: Reshaping the sciences and humanities* (pp. 165-184). Cambridge, MA: MIT Press.

Evans, J. D., Plante, R. L., Bonaventura, N., Busko, I., Cresitello-Dittmar, M., D'Abrusco, R., . . . Thomas, B. (2012). Managing distributed software development in the Virtual Astronomical Observatory. *SPIE Astronomical Telescopes+Instrumentation*. doi:10.1117/12.927371

Figg, W. D., Dunn, L., Liewehr, D. J., Steinberg, S. M., Thurman, P. W., Barrett, C., & Birkinshaw, J. (2006). Scientific collaboration results in higher citation rates of published articles. *Pharmacotherapy, 26*(6), 759-767. doi:10.1592/phco.26.6.759

Finholt, T. A., & Olson, G. M. (1997). From laboratories to collaboratories: A new organizational form for scientific collaboration. *Psychological Science*, *8*(1), 28-36. doi:10.1111/j.1467-9280.1997.tb00540.x

Fraser, M. (2005). Virtual research environments: Overview and activity. *Ariadne*, n.p.

Fredrickson, J. W., & Mitchell, T. R. (1984). Strategic decision processes: Comprehensiveness and performance in an indistry with an unstable environment. *The Academy of Management Journal, 27*(2), 399-423.

Gaddis, J. L. (2002). *The landscape of history: How historians map the past.* New York: Oxford University Press.

Genoni, P., Merrick, M., & Wilson, M. (2009). e-Research and scholarly community in the humanities. In N. W. Jankowski (Ed.), *e-Research: Transformation in scholarly practice* (pp. 91-108). New York: Taylor & Francis.

Gersick, C. J. (1991). Revolutionary change theories: A multilevel exploration of the punctuated equilibrium paradigm. *Academy of Management Review, 16*(1), 10-36. doi:10.5465/AMR.1991.4278988

Hagstrom, W. O. (1965). *The scientific community.* New York: Basic Books, Inc.

Halfpenny, P., Procter, R., Lin, Y.-L., & Voss, A. (2009). Developing the UK-based e-Social science research program. In N. W. Jankowski (Ed.), *e-Research: Transformation in scholarly practice* (pp. 73-90). New York: Taylor & Francis.

Jankowski, N. W. (2009). The contours and challenges of e-Research. In N. W. Jankowski (Ed.), *e-Research: Transformation in scholarly practice* (pp. 3-31). New York: Taylor & Francis.

Jeffreys, P. W. (2010). The developing conception of e-Research. In W. H. Dutton, & P. W. Jeffreys (Eds.), *World wide research: Reshaping the sciences and humanities* (pp. 51-66). Cambridge, MA: The MIT Press.

Kanefsky, B., Barlow, N. G., & Gulick, V. C. (2001). Can distributed volunteers accomplish massive data analysis tasks? *Lunar and Planetary Science, 1*, 1-2.

Kraut, R. E., Fussell, S. R., Brennan, S. E., & Siegel, J. (2002). Understanding effects of proximity on collaboration: Implications for technologies to support remote collaborative work. In P. Hinds, & S. Kiesler (Eds.), *Distributed work* (pp. 137-162). Cambridge, MA: MIT Press.

Lariviere, V., Gingras, Y., & Archambault, E. (2006). Canadian collaboration networks: A comparative analysis of the natural sciences, social sciences and the humanities. *Scientometrics, 68*(3), 519-533. doi:10.1007/s11192-006-0127-8

Lee, S., & Bozeman, B. (2005). The impact of research collaboration on scientific productivity. *Social Studies of Science*, *35*(5), 673-702. doi:10.1177/0306312705052359

Lewis, J. (2000). Funding social science research in academia. *Social Policy & Administration, 34*(4), 365-376. doi:10.1111/1467-9515.00197

Lukkonen, T., Persson, O., & Sivertsen, G. (1992). Understanding patterns of international scientific collaboration. *Science, Technology, & Human Values*, *17*(1), 101-126. doi:10.1177/016224399201700106

Maglaughlin, K. L., & Sonnenwald, D. H. (2005). Factors that impact interdisciplinary natural science research collaboration in academia. *International Society for Scientometrics and Informatrics (ISSI) Conference,* (pp. 1-12).

McCain, K. W. (1991). Communication, competition, and secrecy: The production and dissemination of research-related information in genetics. *Science, Technology, & Human Values, 16*(4), 491-516. doi:10.1177/016224399101600404

Miles, R. E., Snow, C. C., Meyer, A. D., & Coleman, H. J. (1978). Organizational strategy, structure, and process. *Academy of Management Review, 3*(3), 546-562. doi:10.5465/AMR.1978.4305755

Mintzberg, H. (1978). Patterns in strategy formation. *Management Science, 24*(9), 934-948. doi:10.1287/mnsc.24.9.934

Mintzberg, H., Raisinghani, D., & Théorêt, A. (1976). The structure of "unstructured" decision processes. *Administrative Science Quarterly, 21*(2), 246-275. doi:10.2307/2392045

Myers, J. D. (2008). A national user facility that fits on your desk: The evolution of collaboratories at the Pacific Northwest National Laboratory. In G. M. Olson, A. Zimmerman, & N. Bos (Eds.), *Scientific collaboration on the internet* (pp. 121-134). Cambridge, Massachusetts: The MIT Press.

National Collaborative Research Infrastructure Strategy. (n.d.). *Evaluation report: Executive Summary.* Retrieved from Australian Government, Department of Education and Training: http://docs.education.gov.au/system/files/doc/other/national_collaborative_research_infrastructure_strategy_executive_summary.pdf

National Institute of Medicine National Research Council. (2002). *Integrity in scientific research.* Washington, DC: National Academy of Sciences.

National Science Foundation. (2005). *Cyberinfrastructure*. Retrieved from NSF-Wide investment: http://www.nsf.gov/news/priority_areas/cyberinfrastructure/index.jsp

Newell, S., & Swan, J. (2000). Trust and inter-organizational networking. *Human Relations, 53*(10), 1287-1328.

Newman, M. E. (2001, January 16). The structure of scientific collaboration networks. *Proceedings of the National Academy of Science of the United States of America, 98*(2), 404-409. doi:10.1073/pnas.98.2.404

Olson, G. M., & Olson, J. S. (2000). Distance matters. *Human-Computer Interaction, 15*, 139-178. doi:10.1207/S15327051HCI1523_4

Pao, M. L. (1982). Collaboration in computational musicology. *Journal of the American Society for Information Science, 33*(1), 38-43. doi:10.1002/asi.4630330107

Paterson, M., Lindsay, D., Monotti, A., & Chin, A. (2007). DART: A new missle in Australia's e-research strategy. *Online Information Review*, *31*(2), 116-134. doi:10.1108/14684520710747185

Pelz, D. C., & Andrews, F. M. (1976). *Scientists in organizations: Productive climates for research and development* (Rev ed.). Ann Arbor, MI: Institute for Social Research.

Perianes-Rodriguez, A., Olmeda-Gomez, C., & Moya-Anegon, F. (2010). Detecting, identifying and visualizing research groups in co-authorship networks. *Scientometrics*, *82*, 307-319. doi:10.1007/s11192-009-0040-z

Plale, B. (2014, May 20). HathiTrust and HTRC: The changing digital library. *El Colegio de Mexico*. doi:http://www.slideshare.net/BethPlale/hathitrust-plalemay2014htrcpptx

Poole, M. S. (2004). Central issues in the study of change and innovation. In M. S. Poole, & A. H. Van de Ven (Eds.), *Handbook of organizational change and innovation* (pp. 3-31). New York: Oxford University Press.

Poole, M. S., Van de Ven, A., Dooley, K., & Holmes, M. (2000). *Studying processes of organizational change and development: Theory and methods.* New York: Oxford University Press.

Price, D. J., & Beaver, D. D. (1966). Collaboration in an invisible college. *American Psychologist*, 1011-1018. doi:10.1037/h0024051

Price, D., & Gürsey, S. (1976). Studies in scientometrics. Part 1. Transience and continuance in scientific authorship. *International Forum on Information and Documentation*, 17-24.

Qln, J., Lancaster, F. W., & Allen, B. (1997). Types and levels of collaboration in interdisciplinary research in the sciences. *Journal of the American Society for Information Science, 48*(10), 893-916. doi:10.1002/(SICI)1097-4571(199710)48:10<893::AID-ASI5>3.0.CO;2-X

Reilly, E. D. (2003). *Milestones in computer science and information technology.* London: Greenwood Press.

Research Councils UK. (2014). *Previously funded programmes*. Retrieved from Research Councils UK: http://www.rcuk.ac.uk/research/xrcprogrammes/prevprogs/)

Sahu, C. (2013). History of technology in research. In N. Sappleton (Ed.), *Advancing research methods with new technologies* (pp. 1-11). Hershey, PA: Information Science Reference.

Shin, J. C., & Cummings, W. K. (2010). Multilevel analysis of academic publishing across disciplines: Research preference, collaboration, and time on research. *Scientometrics, 85*, 581-594. doi:10.1007/s11192-010-0236-2

Shrum, W., Genuth, J., & Chompalov, I. (2007). *Structures of scientific collaboration.* Cambridge, MA: MIT Press.

Sonnenwald, D. H. (2007). Scientific collaboration. *Annual Review of Information Science and Technology, 41*(1), 643-680. doi:10.1002/aris.2007.1440410121

Stohl, C., & Walker, K. (2002). A bona fide perspective for the future of groups. In L. R. Frey (Ed.), *New directions in group communication* (pp. 237-252). Thousand Oaks, CA: Sage Publications, Inc.

Storer, N. W. (1973). Introduction. In R. K. Merton, *The sociology of sciencce: Theoretical and empirical investigations* (pp. xi-xxxi). Chicago: The University of Chicago Press.

Wagner, C. S., & Leydesdorff, L. (2005). Network structure, self-organization, and the growth of international collaboration in science. *Research Policy, 34*(10), 1608-1618. doi:10.1016/j.respol.2005.08.002

Weedman, J. (1998). The structure of incentive: Design and client roles in application-oriented research. *Science, Technology, & Human Values, 23*(3), 315-345. doi:10.1177/016224399802300303

Zuckerman, H. (1967). Nobel laureates in science: Patterns of productivity, collaboration, and authorship. *American Sociological Review, 32*(3), 391-403. doi:10.2307/2091086

Zulueta, M. A., & Bordons, M. (1999). Multidisciplinary research: A global approach to the study of teams in multidisciplinary research areas through bibliometric indicators. *Research Evaluation, 8*(2), 111-118. doi:10.3152/147154499781777612