Supporting Multi-institutional Interdisciplinary Biomedical Collaboration (MIBC): A Biomedical Informatics Approach

Eunjung Sally Lee

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2008

Program Authorized to Offer Degree:
Medical Education and Biomedical Informatics

University of Washington
Graduate School

This is to certify that I have examined this copy of a doctoral dissertation by

Eunjung Sally Lee

and have found that it is complete and satisfactory in all respects,
and that any and all revisions required by the final
examination committee have been made.

Chair of the Supervisory Committee:

Peter Tarczy-Hornoch

Reading Committee:

Peter Tarczy-Hornoch

David W. McDonald

Fredric Wolf

Date: 11/2/08

University of Washington

Abstract

Supporting Multi-institutional Interdisciplinary
Biomedical Collaboration (MIBC):
A Biomedical Informatics Approach

Eunjung Sally Lee

Chair of the Supervisory Committee:
Professor Peter Tarczy-Hornoch
Department of Medical Education and Biomedical Health Informatics

The modern biomedical research community is facing ever more challenging research questions. Out of necessity, biomedical research has become increasingly interdisciplinary and large-scale in nature. Yet large-scale interdisciplinary biomedical collaborations are not easily established or maintained. Many funding agencies identify biomedical informatics as an important foundation to support biomedical collaboration to alleviate some of the challenges large-scale interdisciplinary collaborations face. However, biomedical informatics has yet to understand in detail how large-scale interdisciplinary biomedical collaborations operate and deal with day-to-day challenges associated with collaboration.

This research used contextual field study to describe the characteristics of large-scale interdisciplinary biomedical collaboration in-depth and to identify barriers, existing facilitators, and needs associated with various collaborative processes. The

study result was synthesized to develop a context-specific informatics framework to support large-scale interdisciplinary biomedical collaboration that extends prior research of collaboration in other fields. In the future, the framework can be used as a guide for design and evaluation of collaborative infrastructure.

Table of Contents

List of Figures

List of Tables

Table Number **Page**

Acknowledgements

I am grateful for support of my family, friends and peers. I also want to thank my chair, Peter Tarczy-Hornoch and my committee David McDonald, Fred Wolf and Barbara McGrath at the University of Washington for aiding me through the difficult dissertation process. Finally, I thank my fellow BHI students, in particular Anna Stolyar and Jim Tufano who assisted me data analysis process.

Chapter 1: Multi-institutional Interdisciplinary Biomedical Research

1.1 Need for Multi-institutional Interdisciplinary Biomedical Research

The modern biomedical research community is facing ever more challenging research questions. The complexity of today's research problems has led to a realization that a single lab or discipline often can no longer provide all the necessary expertise and resources to solve these questions (Hara 2003). Out of necessity, biomedical research has become increasingly interdisciplinary and collaborative in nature (Schur 1998; Horwitz 2002; Hara 2003; Sonnenwald 2004; Cummings 2005; NIH). The importance of the collaborative and interdisciplinary nature of biomedical research has already been recognized by many funding agencies (NCRR; NIH). For example, the Data and Collaboratories in the Biomedical Research Community meeting in 2002 supported by the NCR discussed a new round of investment specifically into biomedical research collaboration (NCRR). In its roadmap for modern biomedical research in the 21st century, the National Institute of Health (NIH) also stressed the need for interdisciplinary collaboration. NIH states that "the study of human biology and behavior is a wonderfully dynamic process, and the traditional divisions within health research may in some instances impede the pace of scientific discovery" (NIH interdisc). To lower the barrier of division within health research, NIH initiatives established a series of awards to enable interdisciplinary research, from training scientists for interdisciplinary research to creating centers to forge new interdisciplinary collaborations (NIH interdisc). NIH defines interdisciplinary research as one that "integrates the analytical strengths of two or more often disparate scientific disciplines" and one that "broadens the scope of investigation into biomedical problems, yields fresh and possibly unexpected insights" (NIH interdisc). According to the NIH, collaboration among disciplines will "open up new avenues of scientific inquiry and in the process, be able to tackle increasingly complex questions" (NIH roadmap).

As necessary and valuable as it is, research shows that interdisciplinary scientific collaborations, which are most often multi-institutional, are not easily established or maintained for several reasons. First, the researchers involved have a wide range of expertise and come from multiple disciplines with different cultures and social norms; therefore, these researchers have very little common ground to draw upon to effectively interact and communicate (Hara 2003). Second, biomedical research in particular is highly competitive (Hagstrom 1974). Many researchers involved are unwilling to share with and trust each other in fear of having their research stolen (Campbell 2002; Olson 2002; Birnholtz 2005). Third, many scientists are reluctant to trust tools they are not familiar with, which often results in lack of adoption of core technologies (O'Day 2001). Fourth, interdisciplinary biomedical researchers often lack proximity with their collaborators. Physical proximity is known to be important to scientific collaboration due to the necessity of informal communications in fostering a collaborative environment (Kraut 1998; Herbsleb 2000; O'Day 2001). Fifth, due to disparate locations, disciplines, and lack of adoption, there is a general lack of a common infrastructure connecting all disparate systems and workflows (CSIEDG; NHII; Lee 2008 (in Press)).

1.2 Motivation

Biomedical informatics has been increasingly recognized as an important part of foundation to support biomedical research to alleviate some of the difficulties of multi-institutional interdisciplinary research. In its roadmap and RFA (CTSA RFA; NIH roadmap), NIH proposes biomedical informatics as the key to enabling and facilitating new interdisciplinary multi-institutional biomedical collaboration. The National Meeting on Enhancing the Discipline of Clinical and Translational Sciences (NMEDCTS) identified biomedical informatics as "an essential part of the new clinical and translational effort, which places a strong emphasis on multidisciplinary collaborations." In the summary, biomedical informatics is also noted for "its role in bridging disciplines" and for being "integrative, facilitating communication across disciplines and analysis of data from disparate sources." The RFA (CTSA RFA) that subsequently followed this meeting, the Institutional Clinical and Translational Science Award (CTSA), calls for the establishment of the institutes of clinical and translational research (ICTRs). An ICTR is "a central unit whose exclusive purpose is to address all the issues and identify all the resources needed to conduct clinical and translational research". The CTSA RFA recognizes the importance of biomedical informatics by stating that it is "the cornerstone" of an ICTR (CTSA RFA; Lee 2008 (in Press)).

Although biomedical informatics must serve as an important support structure for multi-institutional interdisciplinary biomedical collaboration (MIBC), lack of prior research makes such endeavor difficult to accomplish. This research seeks to take the first step toward addressing the gap in the research of MIBC in the field of biomedical informatics. It does so by asking the following broad research question: "How can biomedical informatics best support current multi-institutional interdisciplinary biomedical collaboration?" and by focusing on following three gaps.

Gap 1:

Lack of cross fertilization between biomedical informatics and collaboration research in other fields

Not much literature of research on collaboration exists within the field of biomedical informatics. There is however, a large body of literature in fields such as computer supported cooperative work (CSCW) and information science (IS) on fostering collaboration (Olson 1996; Finholt 1997; Katz 1997; Casper 1998; Melin 2000; Schleyer 2001; Chin 2002; Finholt 2002; Addis 2003; Foster 2003; Jeffrey 2003; Fisher 2004; Halkola 2004; Myers 2004; Baker 2004a; Baker 2005; Lee 2006). Many social and technical factors that facilitate collaboration emerged from these prior studies (Finholt 2003). These studies can contribute to biomedical informatics valuable general concepts surrounding collaboration that are useful in supporting MIBC. Nevertheless, there has been very little cross fertilization of research between biomedical informatics and other fields that have traditionally studied collaboration.

Gap 2:

Lack of context-specific studies to characterize multi-institutional interdisciplinary biomedical collaboration

Gap 3:

Lack of context-specific theoretical framework to support Multi-institutional Interdisciplinary Biomedical Collaborations

Even if much can be learned from existing research, the prior studies of collaboration in other fields are not context specific to biomedical research collaborations. All the resulting theories and methodologies from these studies can support MIBCs in a general sense; however, there are factors specific to MIBCs that are beyond the scope of prior collaborative research. Only an in-depth study MIBC will be able to systematically identify the context-specific social and technical aspects of MIBC. Yet to date, only few small studies of MIBCs exist; as a consequence, no theoretical framework to support MIBCs exists.

1.3 Research Aims

This study will pursue following four aims to address three gaps mentioned in Section 1.2:

- Aim 1: Identify general collaborative concepts in related fields that biomedical informatics can utilize (to address Gap 1)
- Aim 2: Describe characteristics of multi-institutional interdisciplinary biomedical collaboration (to address Gap 2)
- Aim 3: Identify which characteristics (identified in Aim 2) biomedical informatics can support and describe associated barriers, existing facilitators, and needs (to address Gap 2)

- Aim 4: Develop a context-specific framework to support multi-institutional interdisciplinary biomedical collaboration (to address Gap3)

This research takes a first step at studying and identifying unique characteristics of Multi-institutional Interdisciplinary Biomedical Collaboration and developing a context-specific framework of MIBC that builds on and extends previous collaborative research in related fields.

1.4 Research Approach

Two approaches were used to conduct this study to address the four research aims above: a) the construction of a preliminary framework based on synthesizing findings in existing literature and b) a qualitative field study. To address Aim 1, a preliminary framework of biomedical collaboration was constructed based on a systematic analysis of existing literature. Chapter 3 describes the preliminary framework of MIBC that was constructed from the collaborative concepts researched in other fields that are applicable to biomedical collaboration. To address Aim 2 and Aim 3, a qualitative field study of two MIBCs was conducted with the preliminary framework of MIBC as its theoretical basis. Chapter 4 describes the theories and methods (interviews, observations, and document examinations) used in the qualitative contextual field study. The field study revealed the general characteristics of MIBCs, which are described in Chapter 5. Further analysis of the characteristics revealed associated barriers, facilitators, and needs that biomedical informatics research needs to focus on (see Chapter 6). The results of field study aided the development of the context-specific framework to support MIBC (see Chapter 7).

1.5 Contributions

The modern biomedical field, faced with complex research questions, is facing the challenges of multi-institutional interdisciplinary collaboration. Biomedical informatics is recognized as a foundation to address these challenges, and the biomedical informatics community faces the difficult question of how best to support such collaboration. This study makes several contributions to address the question of how biomedical informatics can support multi-institutional interdisciplinary biomedical collaboration.

- *Use of qualitative methods in studying collaborative biomedical settings:* Although qualitative methods (described in Chapter 4) have been used in biomedical settings in general, they have never been used to study and characterize complex biomedical collaboration. This study indicates that the use of qualitative methods is a practical approach to characterize (see Chapter 5 and Chapter 6) and develop a context-specific framework of a complex setting (see Chapter 7, Section 1).
- *Extension and verification of the existing Theory of Remote Collaboration (TORC):*
  The preliminary framework developed to address Aim 1 described in Chapter 3 extends the Theory of Remote Collaboration in the context of biomedical setting based on a synthesis of published literature. The preliminary framework was refined and verified through the contextual field study (see Section 7.1.1, Figure 7.1, and Table 7.1), and in turn, some of the concepts in TORC were also verified (see Table 7.2, Table 7.3). The contextual study also revealed some concepts that were not part of TORC. These are the extensions to the TORC that are context-specific to the biomedical field.
- *Characterization of multi-institutional interdisciplinary biomedical collaboration (MIBC):*
  To date, very little is known about multi-institutional interdisciplinary biomedical collaboration. To better support MIBC, this research identifies characteristics of MIBCs through a contextual field study (see Chapter 4 for methods). Chapter 5 describes these characteristics in detail and Chapter 6

delves further into a subset of characteristics with associated facilitators barriers, and needs.

- *The Framework for Multi-institutional Interdisciplinary Biomedical Collaboration (fMIBC):*

  Although general collaborative frameworks such as TORC exist, no biomedical context-specific framework exists to support MIBCs. Through a qualitative contextual study, this research was used to develop the framework for multi-institutional interdisciplinary biomedical collaboration (fMIBC). The study is described in Chapter 4 and its results are described in Chapters 5 and 6. The fMIBC points out factors essential to collaborations. They are composed of social and socio-technical issues, and core collaborative activities and processes. Chapter 7 shows a sample evaluation checklists for MIBC developed using fMIBC.

1.6 Dissertation Overview

Chapter 2: Biomedical Collaborations and Collaboratories

This chapter introduces research conducted in other fields related to collaboratories and collaboration. Although not context specific, existing research on collaboration can inform biomedical informatics on some of the general issues associated with large scale remote collaboration.

Chapter 3: Development of a Preliminary Framework (Aim1)

This chapter discusses the preliminary framework developed through a systematic synthesis of the existing body of literature. Unlike existing generic theories, the preliminary framework is a context-specific framework. The field study described in Chapter 4 has its theoretical base in the preliminary framework.

Chapter 4: Theory and Method to Study Biomedical Collaboration

This chapter describes the theories and methods used to conduct this research. After introducing the theories and methods, actual steps taken to carry out the contextual field study are described.

Chapter 5: Characteristics of Biomedical Research Collaborations (Aim 2)

This chapter describes the fundamental characteristics of MIBCs that emerged from the contextual field study. Two MIBCs under study and their members, as well as the similarities and differences between the two sites are introduced. MIBCs are portrayed by introducing the core activities and processes that are associated with collaboration as well as social and technical issues that influence collaboration.

Chapter 6: Opportunities for Biomedical Informatics in Collaborative Barriers, Facilitators, and Needs (Aim 3)

This chapter delves deeper into the core activities and processes described in Chapter 5. Each of the core activities and processes are broken into facilitators, barriers, and needs. The facilitators are factors and processes that current participants consider to work well and aid in collaborative efforts. The barriers are factors that hinder collaboration from working smoothly or from forming. The needs are factors that those involved in collaboration thought could help with the barriers they face during collaborative research.

Chapter 7: Contribution, Future Work, and Conclusion (Aim 4)

This chapter describes the contributions of the study including fMIBC, as well as limitations and future work.

Chapter 2: Biomedical Collaborations and Collaboratories

Biomedical research has become more interdisciplinary and collaborative in nature due to increasingly complex research questions that are difficult to address by one researcher or even a single team of researchers. Large interdisciplinary collaborations spanning multiple institutions are also becoming a norm in the biomedical field due to the recent push for such collaboration from funding agencies such as the NIH (CTSA RFA; NIH roadmap). According to the NIH roadmap as outlined in the Clinical and Translational Science Award call for proposals, the biomedical informatics community is to play a key role in enabling and facilitating such collaborations. Yet virtually no prior studies of multi-site biomedical collaborations exist within the biomedical informatics field. Although biomedical informatics community have almost no prior knowledge of what biomedical collaboratories are and how they function, large scale collaborations have existed in other fields and have been studied by researchers in fields such as Computer Supported Cooperative Work (CSCW) and Information Science (IS) (Finholt 2003).

In the CSCW and IS the term "collaboratory" is used to describe large scale remote collaborations. Although biomedical collaboration faces unique challenges and existing studies of collaboratories are not specifically applicable to the biomedical field, they can provide an insight into how biomedical informatics can support current multi-institutional interdisciplinary biomedical collaboration (MIBC). In this chapter, the history of collaboratory development is introduced in Section 2.1. Then, Sections 2.2 and 2.3 review collaboratory theories and design methodologies. The chapter then concludes with Section 2.4 that describes some of the gaps in collaboratory research.

2.1 Collaboratory Development

The collaboratory concept started with the concept of remote collaboration defined by William Wulf in 1989. Essentially, a collaboratory refers to a "center without walls, in which the nation's researchers can perform their research without regard to physical location, interacting with colleagues, accessing instrumentation, sharing data and computational resources, and accessing information in digital libraries" (Kouzes 1996; Finholt 2003). Collaboratories were formed in various fields such as physics, engineering and natural sciences and have been heavily researched in the fields of CSCW and IS (Bly 1997; Finholt 1997; Katz 1997; Casper 1998; Baker 2000; Melin 2000; Schleyer 2001; Sonnenwald 2001; Teasley 2001; Arnstein 2002; Arnstein 2002; Bafoutsou 2002; Chin 2002; Finholt 2002; Olson 2002; Addis 2003; Ash 2003; Foster 2003; Jeffrey 2003; Sonnenwald 2003a; Baker 2004; Fisher 2004; Halkola 2004; Myers 2004; Baker 2004a; Wagner 2005). Initially, collaboratory research spent most of its efforts on building tools support remote collaboration and remote access to common instrumentation. Then as collaborator development matured, the focus of collaboratory research moved from tool development to supporting both social and technical processes of scientific research. Sections 2.1.1 to 2.1.3 describe various collaboratories and their development to date.

2.1.1 Earlier Collaboratories: Focusing on Technology Development

When the concept of remote collaboration was in 1989, the technical infrastructure to support such collaboration did not yet exist. Therefore, the early collaboratories, were tool-centric, they focused on developing specific functional parts of collaborator system such as tools for communication and shared access to instrumentation systems to support remote collaboration. One of the earliest examples of a collaboratory is a biological research collaboratory – the Worm Community System (WCS) (Figure 2.1). The WCS started in 1990 as a customized piece of software for few biologists studying *c. elegans* (a tiny sea worm) at a single location, but soon geographically dispersed researchers began to use it to collaborate together. The WCS developed into a distributed web system with the capability to enable informal and formal communication and data access across geographically dispersed sites. The tools available in the software include graphics of the organism's physical structure, a genetic map, formal and informal research annotations, directories, a thesaurus, and a database of research findings on *c. elegans*. (WCS; Star 1994). In 1994, Star and Ruhleder conducted observations and interviews to describe the use of WCS (Star 1994; Finholt 2003).

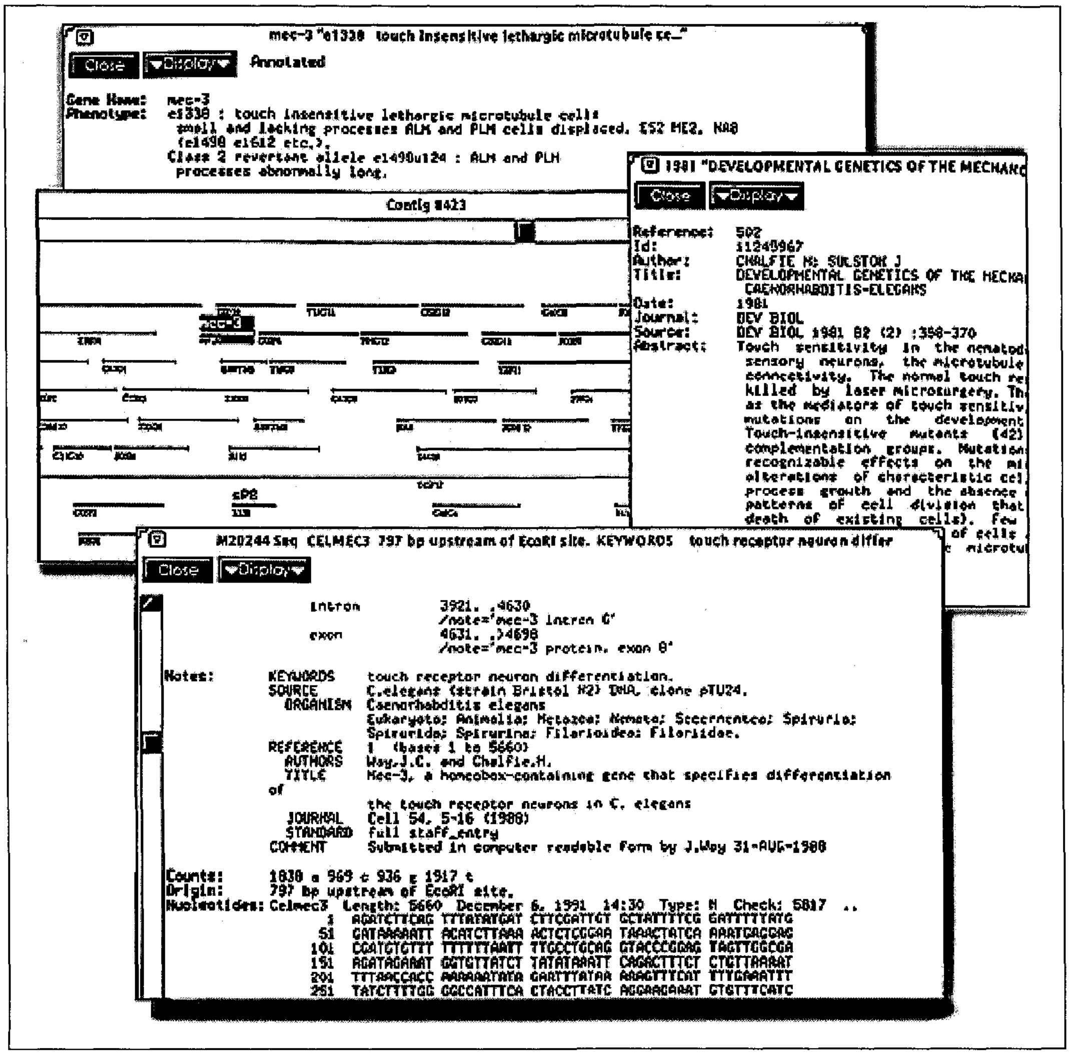


Figure 2.1
The Worm Community System interface
(http://www.canis.uiuc.edu/projects/wcs/national_collaboratories.html)

Another tool-centric collaboratory in the biological research domain was the Biological Collaborative Research Environment (BioCoRE) (BioCoRE; Bhandarkar 1999) (Figure 2.2). The BioCoRE was developed in 1998 by the University of Illinois at Urbana-Champaign as a web-based collection of collaborative tools for structural biologists. The BioCoRE provides a suite of collaborative visualization programs, web-based forms and interfaces for submitting and monitoring networked supercomputing jobs, electronic notebooks for collecting data and notes, asynchronous and real-time communication tools for discussions, and collaborative writing tools for generating reports and publications (Bhandarkar 1999; Chin 2004). BioCoRE conducted an ongoing simple multiple choice user satisfaction survey from 2002-2005 (BioCoRE).

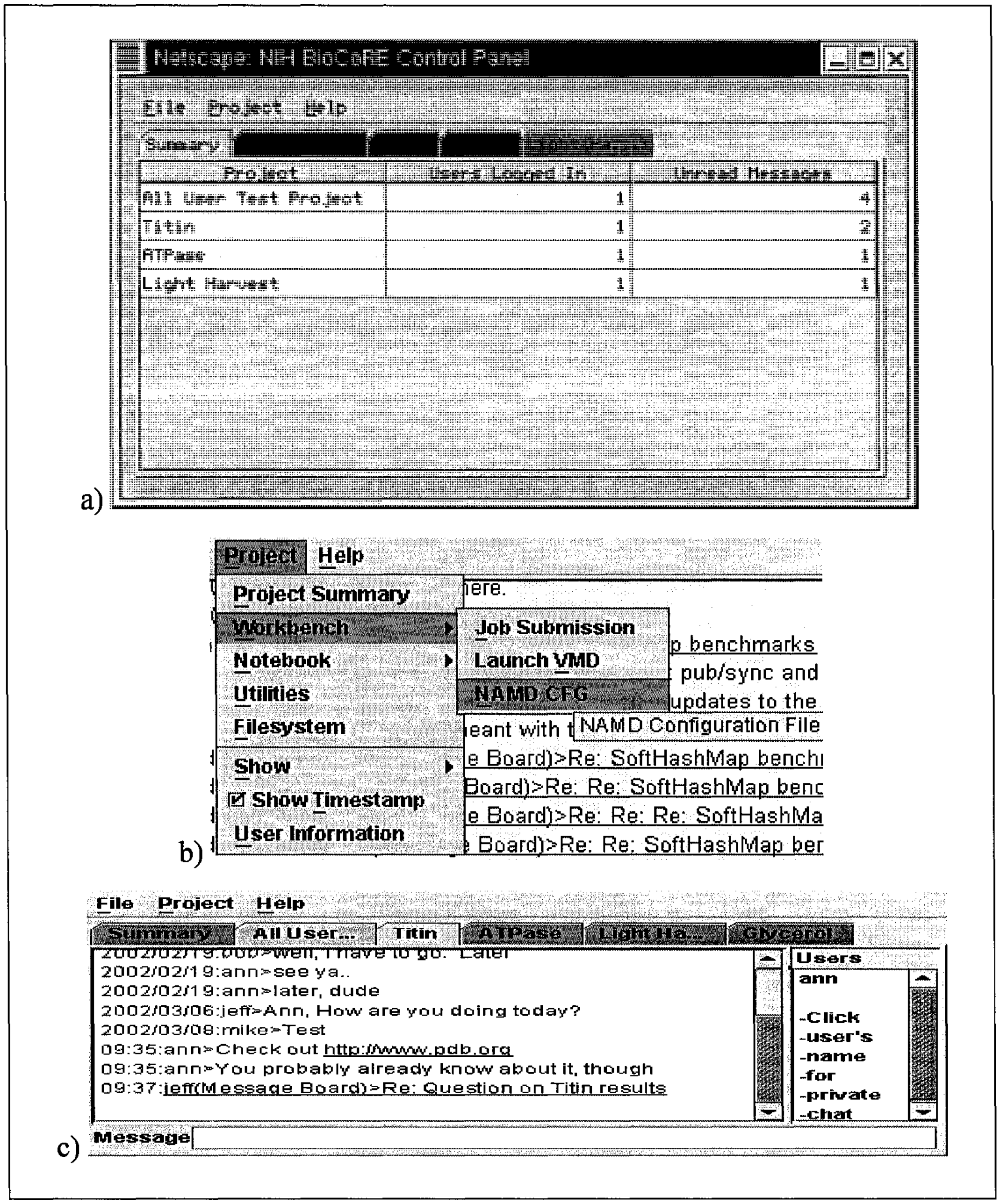


Figure 2.2
The BioCoRE interface (http://biocore.ks.uiuc.edu/biocore/docs/controlpanel.html).
a) Control Panel: Informs user of the latest occurrences in BioCoRE projects and gives an easy way to communicate and to navigate. b) Project Menu: Contains a list of shortcuts to the various BioCoRE applications. c) Message Line: At the bottom of the Control Panel is a message line. The text typed in this box will be sent to others in the same BioCoRE project. They will be notified of the message and can then read and respond to it.

The collaboratories that focused on access to a common instrumentation system developed various tools to enable remote manipulation of a scientific instrument. For example, the goal of the Upper Atmospheric Research Collaboratory (UARC) (Figure 2.3) was to give access to instruments at a NSF funded observatory located above the Arctic Circle on the west coast of Greenland. The UARC was formed in 1993 for a distributed community of physicists and provided real-time control of remote instruments used to study space physics and other upper atmospheric events. It also allowed communication among geographically distributed colleagues, shared access to real-time data, and access to archived data (CTSA RFA; Olson 1998; Finholt 2003). Olson et al describe in detail user experience with UARC system (Olson 1998).

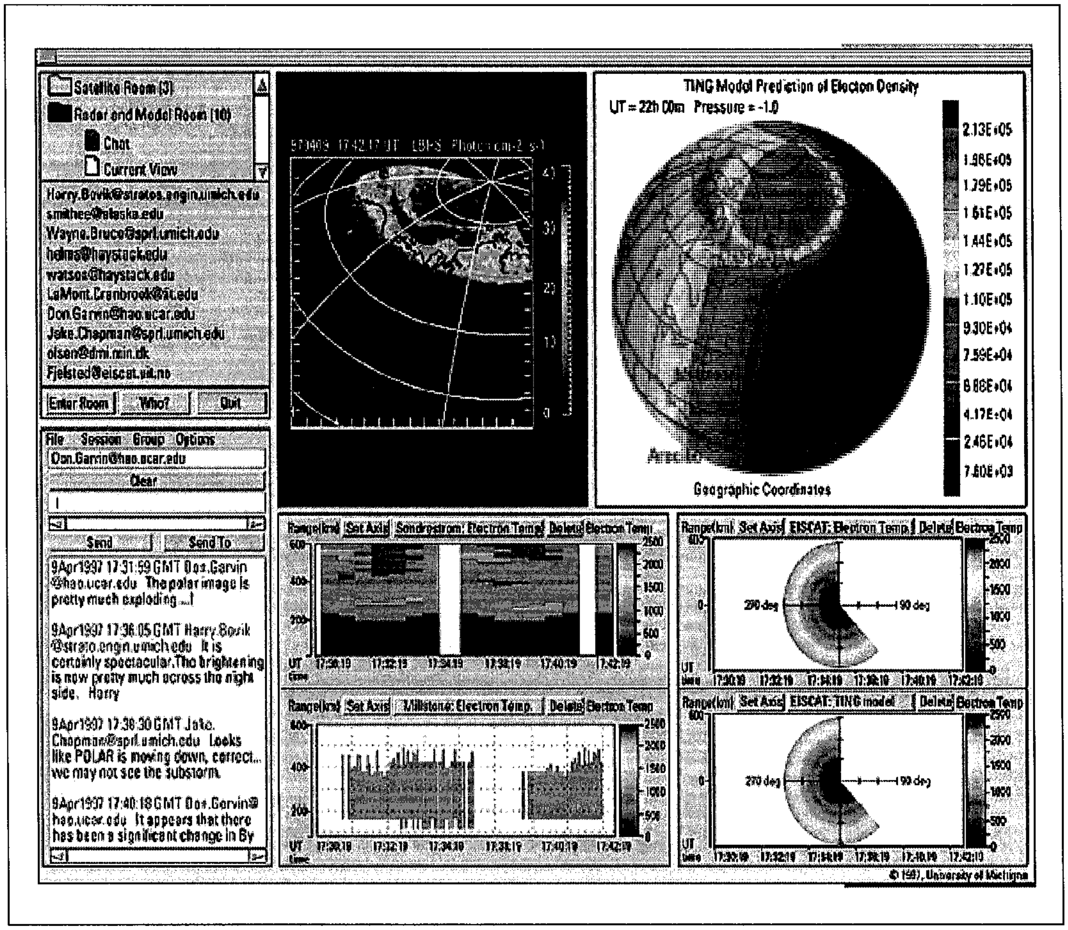


Figure 2.3
The Upper Atmospheric Research Collaboratory interface
(http://portal.acm.org/citation.cfm?id=275276&coll=portal&dl=ACM&CFID=18180356&CFTOKEN=14805760)

Another example of a collaboratory created to provide an access to instruments is the Environmental Molecular Sciences Laboratory Collaboratory (EMSL) at the Pacific Northwest National Laboratory (Figure 2.4). The EMSL was developed in 1995 to provide access to a collection of instruments and expertise for environmental molecular science researchers. The EMSL facility consists of data resources, magnetic resonance instruments, mass spectrometers, and applications to support remote operation of this hardware, as well as an electronic laboratory notebook. In addition, EMSL users have access to a set of generic collaboration tools such as whiteboards, chat rooms, audio and video conferencing, and application sharing (EMSL; Kouzes 1996; Schur 1998; Finholt 2003). Schur et al. conducted interviews with scientists prior to development of EMSL and observed its use afterwards (Schur 1998).

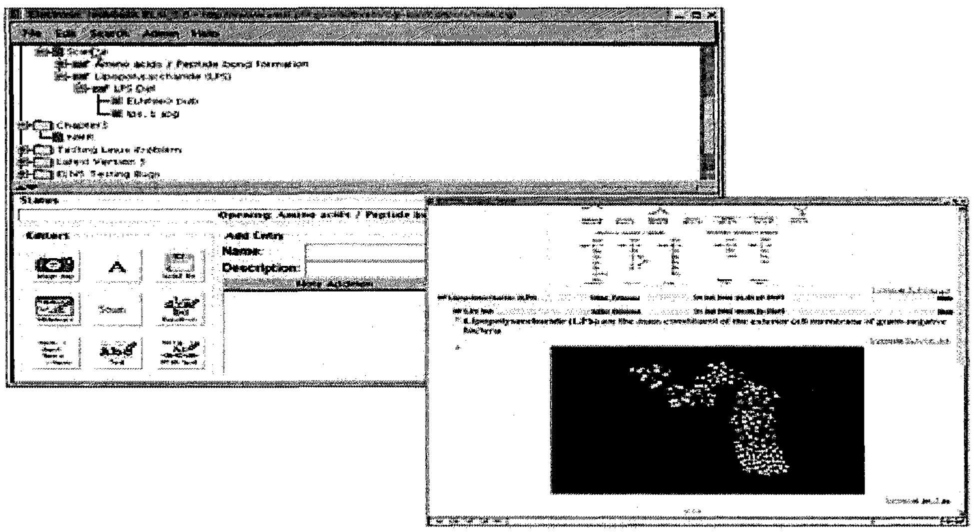


Figure 2.4
The Environmental Molecular Sciences Laboratory interface (http://collaboratory.emsl.pnl.gov/software/eln/)

2.1.2 Later Collaboratories: Focusing on More Complex Research

As the technology matured, the focus of collaboratories gradually began to shift from developing technologies to support remote collaboration within a single field to supporting interdisciplinary research. For example, the nanoManipulator collaboratory developed in 1995 is one of the first collaboratories to facilitate interdisciplinary collaborative research of physicists, chemists, and gene therapy researchers (Figure 2.5). The nanoManipulator collaboratory is a virtual laboratory for geographically-distributed. In this virtual laboratory, scientists can access nanoManipulator, an instrument that allows them to remotely investigate physical samples through an atomic force microscope. The NanoManipulator collaboratory also incorporates visualization and haptic technology, which allows scientists to see, feel, and modify biological samples being studied (Sonnenwald 2001).

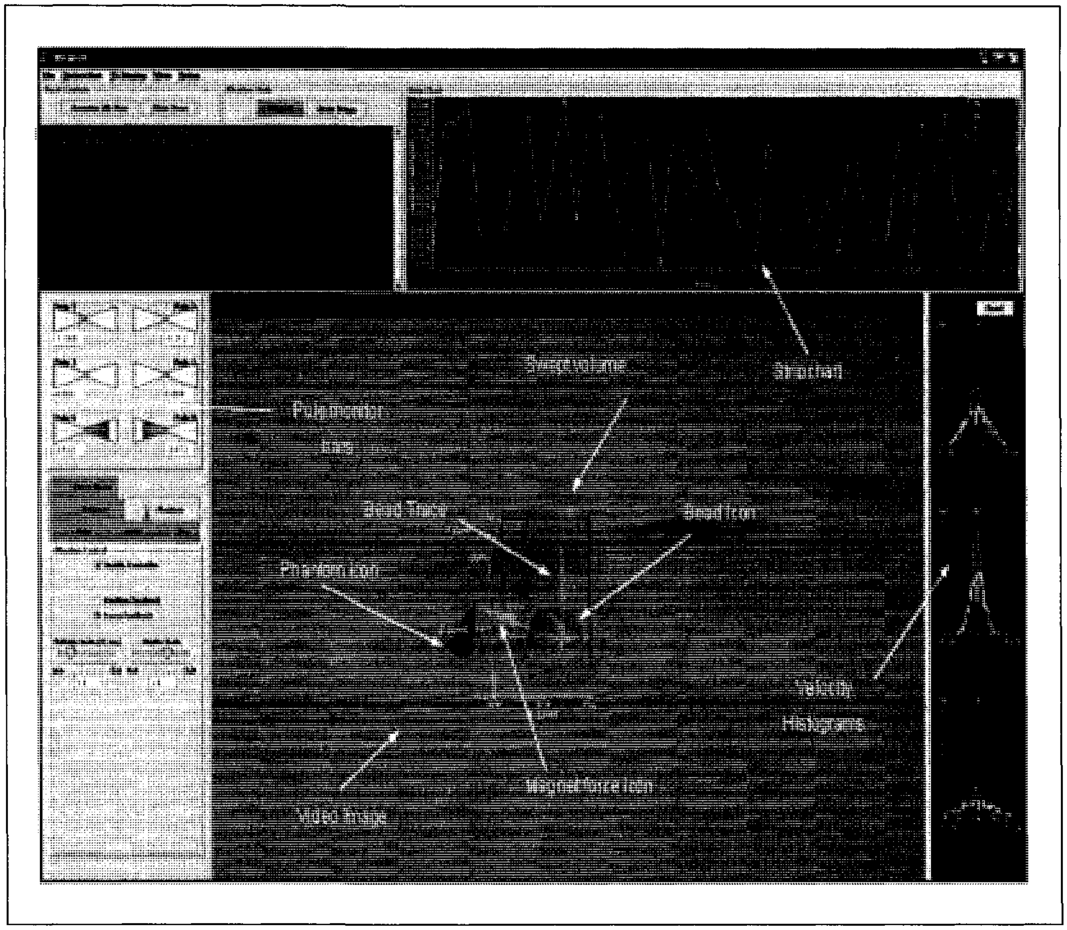


Figure 2.5
The nanoManipulator interface
(http://www.cs.unc.edu/Research/nano/cismm/ui/index.html)

The Biological Sciences Collaboratory (BSC) is another biological research collaboratory that differ from earlier collaboration efforts in that it focuses much more importance on sharing data (Figure 2.6). The BSC, created in 2004, recognized that scientific collaboration focuses heavily on the sharing and joint analysis of scientific data (BSC; Chin 2004). For that reason, the BSC enables the sharing of biological data and analyses through metadata, electronic notebooks, data organization views, data provenance tracking, analysis notes, task management, and scientific workflow management (Chin 2004). The recognition of the importance of metadata and data provenance tracking is especially important to the data-centric approaches and had not been formally recognized by previous tool-centric collaboratories.

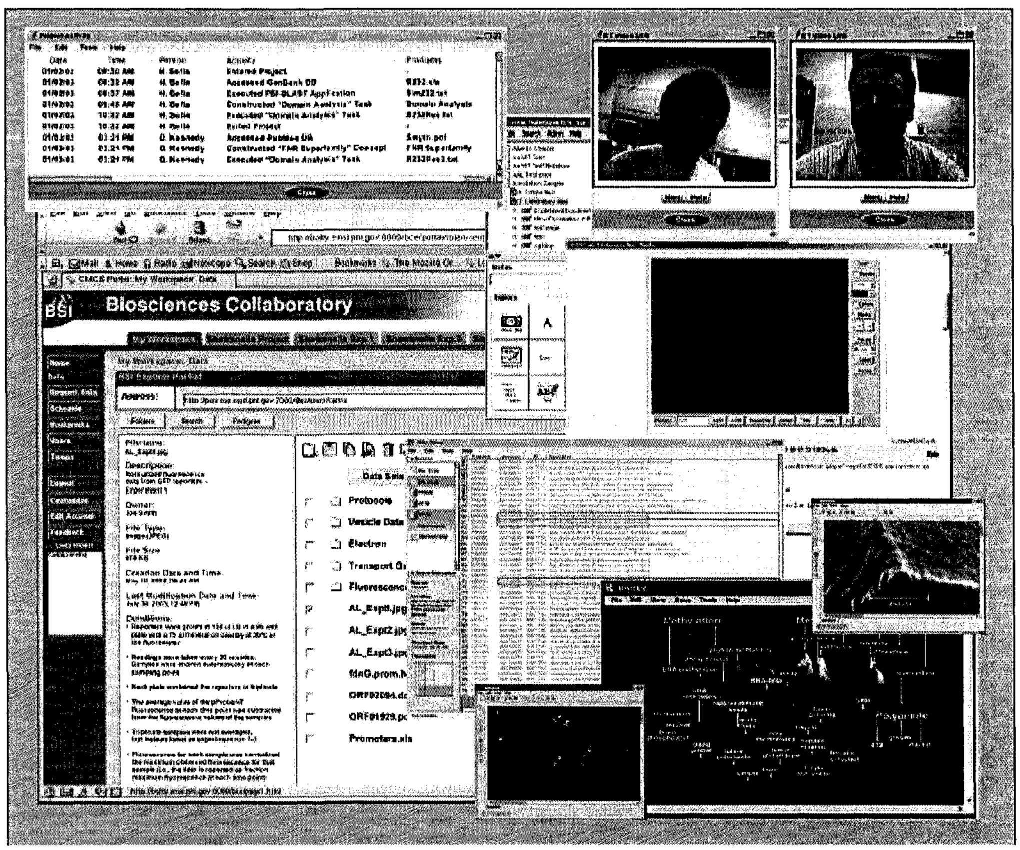


Figure 2.6
The Biological Sciences Collaboratory interface showing integrated set of tools: data, applications, and communications
(http://portal.acm.org/citation.cfm?id=1031677)

2.1.3 Modern Collaboratories: Focusing on Multi-institutional Collaborations

Following earlier smaller scale remote collaborations came several large multi-institutional collaborations such as the Biomedical Informatics Research Network (BIRN) and the cancer Biomedical Informatics Grid (caBIG). Both of these remote collaborations are based on a flexible network or grid-like technology (Foster 2002; Cao 2003; Foster 2003). The BIRN is a virtual community that provides shared resources for the diagnosis and treatment of disease and enables communication and collaboration among investigators (BIRN) (Figure 2.7). The caBIG is a grid connecting individuals and institutions to enable the sharing of data and tools (caBIG) (Figure 2.8). Recently, Lee et al. conducted an ethnographic study of BIRN to understand "human infrastructure", human and organizational properties of collaboration (Lee 2006).

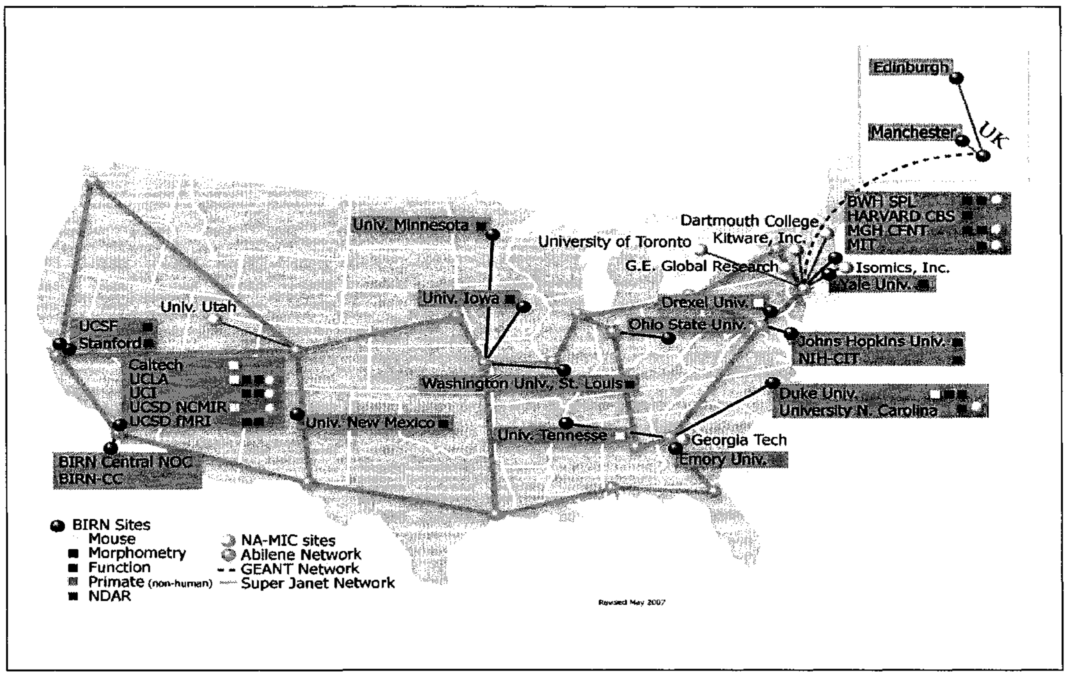


Figure 2.7
The Biomedical Informatics Research Network, a multi-institutional interdisciplinary collaboration (http://www.nbirn.net/tools/index.shtm)

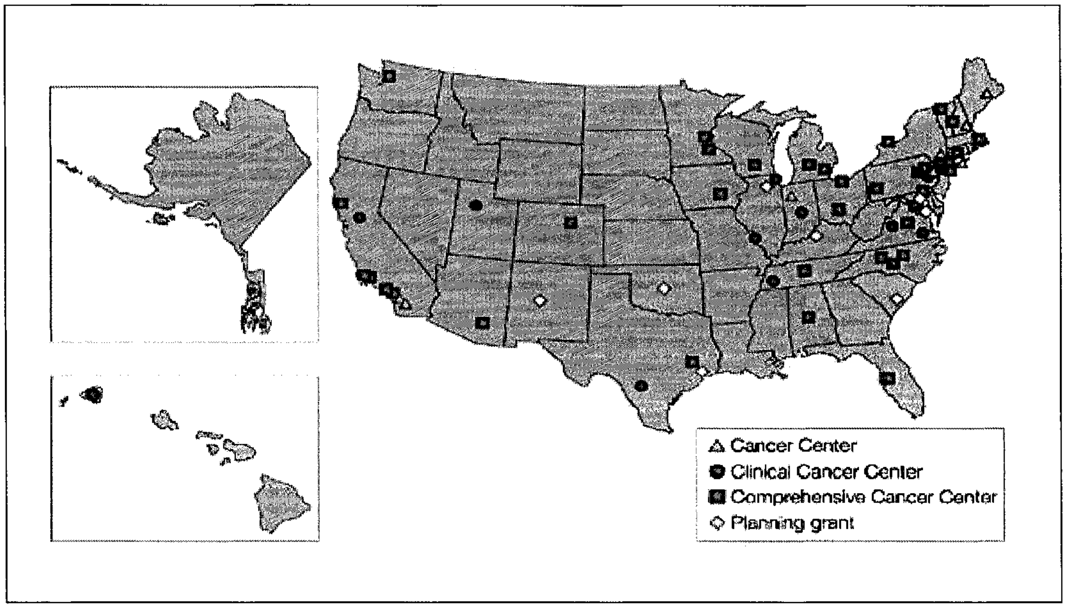


Figure 2.8
The caBIG, cancer Biomedical Informatics Grid
http://www.nature.com/nrc/journal/v4/n10/fig_tab/nrc1458_F2.html

As initial multi-institutional biomedical collaborations began to mature, the future vision of collaborative biomedical research became multi-institutional centers that facilitate large-scale interdisciplinary research. More recently, with the NIH roadmap stressing the importance of interdisciplinary research, collaborative biomedical research became a nationally driven endeavor (NIH roadmap). A prime example of such an interdisciplinary collaboration initiative is the Institutional Clinical and Translational Science Award (CTSA) (CTSA RFA). The CTSA calls for establishment of the Institutes of Clinical and Translational Research (ICTRs), "a central unit whose exclusive purpose is to address all the issues and identify all the resources needed to conduct clinical and translational research". According to CTSA RFA, Biomedical Informatics is seen as "the cornerstone of communication within centers/departments/institutions and with all collaborating organizations (CTSA RFA). Other collaborative centers such as ICTRs are expected to be seen more in the future.

The term "collaboratory" formerly described simple collaborations within a similar field of study that focused on the use of technology to access remote instruments or data. This term is no longer widely used in a same sense. Instead, the concept of collaboratories has evolved and expanded to describe large collaborations and consortiums spanning multiple institutions, covering multiple domains of science, and leveraging various collaboratory technologies. The collaboratories are now called by other names such as consortium, eScience, Grid, center, core and network.

2.2 Theories Related to Collaboratory Research

Not only have collaboratories and the research they support become more complex, but the nature of research on collaboratories has also changed. Initial research on collaboratory focused on developing and improving technology to enable simple remote-collaboration; however, as collaboratory technologies matured, the focus of research on collaboratories shifted toward trying to understand to better support the process of remote, technology-supported scientific research that happens in collaboratories. Many of the later studies of collaboratories were guided by theoretical principles, which will be introduced in this section.

2.2.1 Diversified Approach to Collaboratory Theories

Currently there is no all-encompassing theoretical framework that describes collaboratories to help informatics to better support collaboratories. This lack of theoretical framework is due to the complex socio-technical nature of collaboratories which does not align well with one general theory or a one solution fits all approach. That is why most collaboratories have been one-off affairs, custom designed and implemented to meet specific needs (Winograd 1994; Finholt 2003). Although customized developments for specific needs have better chance of being adopted, the costs of one-offs are high. That is why collaboratory research has sought to find general principles behind collaboratory infrastructure, but not by trying to find an overarching framework, but by breaking down the research into different disciplinary pieces. The reason for this multi-disciplinary piece approach to a common collaboratory framework is simple. There is no single discipline that could address all the socio-technical questions collaboratory research has to face.

Thus far, collaboratory research has addressed several different pieces of collaboratory framework. Some of these pieces include:

- Sharing: What are the social issues associated with sharing in a competitive environment? (Birnholtz 2003)
- Communication: What technologies are needed for remote communication? How do those in collaboration interact with each other? (Kraut 1998; Chin 2004)
- Coordination: How do those involved in collaboration coordinate their activities? (Hoffelner; Cummings 2005)
- Infrastructure: What are some of the technologies needed for remote collaboration? (Olson 2005)
- Organizational: What are some of the organizational and institutional effects collaboration faces? (Olson 2005)

2.2.2 The Theory of Remote Collaboration (TORC)

Although there is no all encompassing theory in the domain of collaboratory research, there is a theory that attempts to summarize the collaboratory knowledge collected to date. The Theory of Remote Collaboration (TORC) is a descriptive theory that is based on both a set of evaluation studies and a survey focused on identifying the critical aspects of remote collaboration (Olson 2005). The TORC defines "success of collaboratories" and codifies the major factors that lead to collaboratory success (Olson 2005).

Definition of Collaboratory Success

Olson et al. defined the success of collaboratories in terms of short and long term results. Some indications of short-term collaborative success include:

- More co-authored publications
- Actively aggregating and sharing of data
- Wide access to expertise and instruments
- Active use of tools developed for collaborative purposes are being used

Some indications of long-term collaborative success include:

- Change in science careers (e.g. those who participate in collaborative research are more likely to get tenure)
- Greater diversity among scientists
- Greater quality of life (e.g. due to reduced amount of time spent in travel)
- More students entering science where collaboratories are used, which inspires further development of collaboratories resulting in more visibility and funding

The ultimate success of collaborative research would be that the collaboration process results in new breakthroughs or more rapid progress in science (Olson 2005).

Five Factors that Lead to Successful Collaboration

The TORC outlines five factors that lead to successful collaboration, they are: 1) Nature of the work, 2) Common ground, 3) Collaboration readiness, 4) Management, planning and decision making, and 5) Technology readiness (Olson 2005).

"Nature of the work" refers to answers to the questions "what to do?" and "how to do it?" in science. TORC states that "loosely coupled work", work that is easily divided, can be done successfully at a distance, while "tightly coupled work" necessarily requires collocation (Olson 2005). Tightly coupled work is "work that strongly depends on the talents of collections of workers, and is non-routine, even ambiguous. In tightly coupled work, components of the work are highly interdependent" (Olson 2005). In contrast, loosely coupled work "has fewer dependencies or is more routine" (Olson 2005).

"Common ground" in TORC refers to the mutual knowledge, beliefs, and assumptions people have in the collaboratory. The successful collaboratory requires common ground. One example of this would be a need for common vocabulary. Unless there is an explicitly defined common vocabulary, those involved in a collaboratory may not be able to communicate and the structure may fail (Olson 2005). For example, MouseBIRN is a collaboratory of scientists studying mouse brain from molecular structure to morphometry. Although they were all studying the mouse brain, they did not map and label the mouse brain in the same way due to the variability in their backgrounds. To solve this problem, therefore, they built an "atlas" (a common vocabulary) that shows the relationship between various terms they use (Olson 2005).

"Collaboration readiness" refers to an individual scientist's willingness to collaborate. The willingness could be due to many reasons such as monetary benefit, recognition that people have reciprocally needed skills, or a need to share equipment. In order for collaboration to succeed, there has to be some benefit for all participants. Willingness also means that the people involved have to like working together and trust each other. They must trust that one will not take advantage of the other's vulnerability, that others will keep their promises, and that they will produce high quality work. Trust is why collaborators who have worked successfully together in the past have a better chance of success in collaboration. They have already found common ground and established trust during their previous collaboration. Another aspect of collaboration readiness is that participants should have common goals and each group participating should believe that they can overcome obstacles to achieve that goal (Olson 2005).

"Management, planning and decision making" refers to the quality of the leadership and management involved. Those participating in collaboration must have time and resources to commit to the collaborative project. This is especially important at the beginning of a project where everyone involved needs to spend a significant amount of time developing a common ground and aligning goals. Before a successful collaboration can occur, a communication and management plan and an oversight committee or advisory boards must be in place. The communication and management plan will ensure that the decision making within the collaboration is free of favoritism and happens fairly and openly so that each participant feels they can influence or challenge decisions. The collaboratories must also have a good knowledge management plan so that they do not loose data or records (Olson 2005).

"Technology readiness" refers to the fit of the technologies involved to the collaboratory work. Due to the geographic distribution of participants all collaboratories must connect people via technology for both communication and core work. In order for this to happen, participants must adopt and be able to use the tools provided. They must feel comfortable using the technology. It is essential that the technology benefits those individuals who are expected use it. The TORC framework also suggests that interoperability, support, and standards are important. The chances of success in collaboration are greater if the participants agree on a single platform. The collaboration also benefits if there is technical support at each location (Olson 2005).

2.3 Collaboratory Design Methodologies

Initial collaboratory research began without the notion of theoretical models of collaboration guiding the designs of collaboratories. This was partially due to the failure of prior development attempts involving theoretical design. For example, the Coordinator or Action Workflow, two systems that implemented models of human collaborative activity based on Speech Act Theory, ended in user resistance and system rejection (Winograd 1994). However, while there has been no theoretical framework that has guided the design, there do exist several common methodological approaches that have been used to design collaboratories. Many of the early collaboratory studies involved researchers who understood the complex nature of scientific collaboration. For this reason, these researchers formally engaged the design of collaboratory infrastructure using Participatory Design (PD) (Bødker 1993; Muller 1993; Sumner 1997; Weng 2006) and User Centered Design (UCD) principles (Beyer 1998; Consolvo 2002). Adopting PD and UCD principles allowed collaboratories to address the specific needs of each collaboratory and avoid some of the user resistance faced by the theoretical designs.

These ad hoc approaches to design have been used in the majority of existing collaboratory research. Many collaboratories started with field research where the designer observed and interacted with the users during their normal activities. Most field studies were then followed with semi-structured interviews to get users' ideas about their work and background information. This information was then used to develop the collaborator application and related technologies. Once the applications were deployed, some collaboratories conducted usability testing to find errors or improperly designed features (Finholt 2003). Other collaboratory studies conducted initial short and long term evaluations of collaboratories (Sonnenwald 2003; Olson 2005).

The UARC and The Space Physics and Astronomy Research Collaboratory (SPARC) are two examples of collaboratories that involved the user-centered design techniques. The UARC was one of the first collaboratories to involve user-centered iterative rapid prototyping design. The collaboratory was developed by a multidisciplinary team composed of domain specialists, computer scientists, and behavioral scientists. The designers, composed of both computer and behavioral scientists, conducted many laboratory and instrument site visits and the collaboratory design emerged through a slow iterative evolution (Olson 1998). Designers of the SPARC also spent time to understanding users' work practices in order to guide development of system specifications. Through user testing and other design evaluation methods, the designers insured that the interface they created had useful functionality and was easy to use (Olson 2005). Many other collaboratories that followed also applied similar design methods (Bly 1997; Schur 1998; Sonnenwald 2001).

Yet user-centered design techniques were not panacea for all adoption barriers. Most collaborator applications and projects still had adoption problems. Many of the adoption issue arose from technical problems and lack of technical support. However, the adoption issues also included many complex social and socio-technical issues related to collaboration and the use of technologies.

2.4 Socio-Technical Issues in Collaboratories

Despite the successes in design and development of many collaboratories, not all collaboratories have succeeded. Many were either never adopted or had problems with continuous use after the initial adoption. One of the biggest barriers to collaboratory adoption have been faults in the collaboratory technologies themselves. For example, the WCS was not adopted because participating biologists, the users of the system, had to master relatively complex system installations within an alien computing environments (e.g. the UNIX operating system) (Olson 2005). The UARC faced the challenge of collaboratory adoption when their migration toward the java platform resulted in system problems (Olson 2005). Yet another project, SPARC also faced adoption challenges because individual scientists went directly to system developers to request new features, which finally resulted in the system becoming unstable because of frequent new versioning (Olson 2005).

Another major problem that has contributed to the failure of collaboratories is that technical support is seldom available to collaborator users. For example, some of the reasons why the WCS was not adopted included users not knowing where to download the system and the difficulty of maintaining the operating system maintenance. Users of the UARC users were challenged by frequent browser downloads, this extra effort for users severely undermined confidence in the UARC system (Finholt 2003). Technical support would have eased such difficulties and promoted adoption of these collaboratories.

Political and social issues have also hindered adoption of collaboratories. For example, one of the social issues WCS faced was poor system adoption partly due to scientists' reluctance to share data publicly (Star 1994; Finholt 2003). The difficulty of sharing data in science is a well known social phenomenon associated with collaboration (Campbell 2002; Birnholtz 2003).

2.5 Summary

Large scale multi-institutional, interdisciplinary biomedical collaborations are rapidly becoming a norm in science and biomedical informatics has been identified as a field that can help facilitate such research activity (NIH roadmap). Yet large scale collaborations have not been studied within the biomedical informatics community. Other fields such as CSCW and IS however, have studied many different aspects of collaboration and have accumulated a large body of literature on the topic. However, further research is needed to understand biomedical collaborations and how to support them, as earlier collaboratory research mostly focused only on supporting a generic collaboration issues such as building the collaboratory tools themselves and studying their use and measuring user satisfaction.

In this chapter, some existing studies of collaboration, specifically those pertaining to collaboratory research were introduced. Although collaboratory researches were not specifically looking at collaborations in biomedical domain, we can learn from the research because some of the characteristics of collaboration are similar across domains. Therefore, biomedical informatics can get some insight into the overall collaborative process from some of the existing work. Nevertheless, some gaps still exist in our understanding of collaborative research (i.e. how to deal with some of the technical and social barriers?). There is also no available comprehensive theoretical framework based on research data or a comprehensive review of all collaboratory work in the field of biomedical informatics. TORC (section 2.2) is broad enough that it is difficult to apply to MIBC and biomedical informatics. To fill this gap in the literature, a preliminary framework of biomedical collaboration was developed based on existing collaboratory studies with a focus on biomedical research (see Chapter 3).

Chapter3: A Preliminary Framework of Biomedical Collaboration

As discussed in Chapter 1 and Chapter 2, it is difficult to find an all encompassing framework that can be used to guide the study of biomedical collaboration. The Theory of Remote Collaboration (TORC) described in Chapter 2, Section 2.2 does fulfill some of the roles of a theoretical framework in that it summarizes some general properties that lead to successful collaboration, but although such a descriptive theory can be used to inform general collaboration research; the broad nature of TORC makes it difficult to apply to a specific domain. Currently, there is no existing framework that addresses issues of collaboration specifically from a biomedical point of view. This research tries to address the problem of lack in biomedical context-specific framework.

To address the lack of needed framework (see Gap 1 and Aim 1 in Chapter 1), a comprehensive preliminary framework of biomedical collaboration was developed by systemic analysis of existing literature. Qualitative literature analysis methods were used to combine concepts from TORC and various other collaboratory research in the context of the biomedical setting. This preliminary framework will be used to guide the study of multi-institutional interdisciplinary biomedical collaboration (MIBC) as its foundational framework (described in chapter 4 sections 3.4-3.6 and 3.8). Section 3.1 explains how existing literature was synthesized and how the preliminary framework was developed. Section 3.2 describes the details of preliminary framework itself. This concluding Section 3.3 which illustrates how the preliminary framework described in Section 3.2 can be used to advance biomedical informatics research and collaboration.

3.1 Literature Review

The concept of collaboration has been studied by researchers in many different fields and from various different angles and points of view. For that reason, the literature that applies to collaboration is distributed over many different disciplines and is heterogeneous in nature. This synthesis of literature therefore, includes a wide range of literature from a variety of relevant disciplines that study and support collaborations. The review is composed of a survey of literature in Computer Science, Information Science, Biomedicine and Social Sciences. The contents of this literature include formative studies designed to collect requirements for new collaborative infrastructure, participatory design studies, systems design and deployment studies, as well as qualitative summative evaluations of collaborative infrastructure. The focus of the search was not only in technical infrastructure associated with collaboration, but also naturalistic studies that highlighted social issues. The studies vary from descriptions of the systems, evaluations of systems, analyses of the systems without a formal methodology, to qualitative analyses of themes across collections of systems.

The primary search began in Google Scholar (Google scholar) and PubMed (PubMed), but subsequent searches included other relevant literature indices and databases such as: Library Literature & Information Science, Library and Information Science Abstracts (LISA), ISI web of Knowledge (Social Sciences Citation Index), Association for Computing Machines (ACM) and the Institute for Electrical and Electronics Engineers (IEEE). Published articles, books, conference proceedings, and grey literature such as dissertations and websites were used. The initial keywords that began the search were: collaboratory, collaboratories, and biomedical collaboration. Additional literature was then found through citation and reverse citation searches and through discussions with colleagues in the area of collaborative research. The review was restricted only to science and engineering research collaborations. Although collaborations exist in many fields, the challenges within these science disciplines were felt to be more germane to the biomedical setting. Over 150 documents were screened and close to 100 relevant documents were reviewed.

A qualitative approach was used to characterize and identify themes in each article rather than a traditional literature review approach of serially reviewing the existing frameworks from a certain perspective (i.e. biomedical informatics). It was felt that qualitatively categorizing the emergent and recurring themes would present a more synthetic overall view of the existing literature. Due to the heterogeneous nature of literature under study, the qualitative approach was better suited to identify critical results that cross methodological approaches and disciplinary boundaries. The concepts in the review were iteratively developed through careful reading of each selected piece of literature. After the initial list of concepts was compiled, the individual concepts were sorted and categorized into larger concepts.

Although The Theory of Remote Collaboration (TORC) described in Chapter 2 exists to guide a general study of collaboration. However, it is not context-specific enough to guide the study of biomedical collaboration. For example, data security is one of the most important factors to consider during biomedical collaboration due to the nature of data involved. Yet TORC does not mention this concept at all. The Preliminary Framework within the context of biomedical research was developed for to better guide the field study.

3.2 Preliminary Framework of Biomedical Collaboration

This novel preliminary framework was developed through qualitative analysis of published literature of studies associated with complex collaborations in the biomedical context. The preliminary framework includes the following concepts: a) general concepts that are necessary in any collaboration, b) specific concepts that are desirable in biomedical collaboration, c) environmental factors that support collaboration, and d) factors that support long-term collaboration. The framework describes an ideal set of attributes that modern multi-institutional biomedical collaboration should have. The general attributes that are necessary in any collaboration include: communication, common workspace/coordination, and data sharing and management. The attributes that are particularly desirable in biomedical collaborations include: data integration/analysis, security, metadata/data provenance, and interoperability/data standards. The environmental factors that support collaboration includes: administration/management, technical support, and funding. The factors that aid long-term collaboration include: training and iterative evaluation. Each of these groups of concepts can be used as a guide by the biomedical informatics community to construct needed informatics infrastructure in support of biomedical collaboration. A summary of these concepts and definitions are provided in Table 3.1. The applicability of each concept in the framework to biomedical collaboration is described in section 3.3 using a scenario.

Table 3.1
A summary of the preliminary framework of biomedical collaboration.

		Collaboratory Features	**Definition**
Concepts General to Many Domains (Section 3.2.1)	**Communication**	Communication-asynchronous	A method of communication that enables researchers to communicate non-simultaneously (e.g. e-mail, discussion list, newsgroup)
		Communication-synchronous	A method of communication that enables researchers to communicate in real time (i.e. by phone, videoconferencing, whiteboard, chat)
		Awareness	Awareness is an understanding of the activities of others, which provides a context for your own activity
	Common workspace, Coordination	Common workspace	A physical or a virtual space where researchers can interact and work together (i.e. web portals, wiki)
		Coordination tools	Tools that help manage scheduling and coordination of various tasks involved in collaboration (i.e. scheduler, calendar)
	Data sharing and management	Data management	Technologies that effectively help handle (i.e. retrieve, search, access) data created during research
		Data sharing	Any data produced during research being shared as well as technologies that are involved in sharing
		Archive/Repository	A place where researchers can put their data to store, retrieve, and access data
		Trust	Trust is a belief in the integrity and abilities of others involved in the collaboration, which helps collaborators to work together
		Willingness to share	Willingness is an inclination or readiness to be a part of collaboration and to work toward a common research goal

Table 3.1 (continued)

Concepts Particularly Relevant to Biomedical Research (Section 3.2.2)	**Data Integration and Analysis**	Data Integration	A mechanism to integrate data that are in multiple formats and located in different places
		Data analysis tools	Tools that help analyze data in any way (i.e. visualization, display, statistics)
	Security	Security/Variable Access	A measure to secure access to a system or data, limiting access to resources only to those authorized to use them
	Metadata, Data Provenance	Metadata (Annotations)	Contextual information, descriptions of datasets
		Data tracking (Data provenance)	Tools to help keep track of history of a dataset (i.e. what changes are made and who makes them)
	Interoperability Data Standards	Common vocabulary/Standards	A common set of defined terms or defined formats for a data set
		Interoperability	All the technologies in a collaboration interconnected and integrated to work together and interconnected
Environmental Factors that Support Collaboration **(Section 3.2.3)**	**Administration, Management**	Management structure	A body of people responsible for managing the overall structure of a collaboration and solving conflicts
	Technical Support	Technology support	Personnel that helps with technology (i.e. set up new hardware, software support)
	Funding	Funding	Funding for collaboration, incentives to collaborate
Factors that Aid Long-term Collaboration (Section 3.2.4)	**Training**	Education/Training	Any tools and activities related to educating, mentoring, and training
	Iterative Evaluation	Iterative evaluation	Continuous evaluation of needs and barriers throughout the collaborative process (both technological and social)
		User-centered design	Designing tools with users in mind (i.e. participatory design, ethnographic study of workflow)
		Understanding workflow	Understanding the overall structure of work involved in collaboration

3.2.1 General Concepts that are Necessary in any Collaboration

During the development of the preliminary framework, three concepts that are associated with any working collaboration effort were identified. These three concepts are essential to any collaboration if it is to function successfully as a collaborative effort, they: a) nature of remote communication and communication technology, b) common workspace and coordination tools to bridge distances, and c) defined approaches to data sharing and management. These categories of factors emerged from literature across multiple fields and in a variety of collaborations. They are cross cutting concepts without which collaborations cannot function.

The Nature of Remote Communication and Communication Technology

The importance of communication in collaboration is widely known. Collaboration, by nature, involves people working together toward a certain goal and working together is not possible without some form of communication. Researchers involved in collaboration work together, discuss their research, jointly analyze data, and share ideas. Throughout these processes, they communicate formally and informally using various mediums such as email or phone. This formal and informal communication fosters collaborative relationships by improving working relationships and maintaining shared knowledge (Kraut 1998). The importance of informal communication in collaboration, such as chance meetings in the hallways, is well documented (Hollan 1992; Kraut 1998; Herbsleb 2000; LaCoursier 2004). It is important to note that one reason remote collaborations often fail is partly due to the lack of these informal face-to-face communications (Herbsleb 2000).

Because researchers involved in modern collaborations often work in different institutions and are located geographically apart from one another, their communication in modern collaborations is often aided by technology. Technology can support both synchronous or asynchronous communication. Through synchronous tools such as phones, videoconferencing, and chat systems, scientists can connect to each other in real time. Real time communication is important because it allows for activities such as brainstorming that are otherwise not easily achieved. On the other hand, asynchronous tools such as email or discussion lists allow researchers to communicate outside of a set time; thus, these asynchronous communication technologies enable researchers to work flexibly around their busy schedules. Through a number of mechanisms, communication technologies facilitate both planned and unplanned interactions that promote collaboration and increase awareness of research (Dourish 1992; Hollan 1992; Kraut 1998; Hara 2003).

Minimal communication support for multi-institutional collaboration should include at least one asynchronous communication tool, such as email and a synchronous communication tool, such as communication by phone. Even this minimal set of tools will enable basic communication necessary for collaboration. However, more advanced tools such as videoconferencing will additionally benefit the communication process in collaboration.

Common Workspace and Coordination Tools to Bridge Distance

Collaborations must have a common workspace, a "place" where researchers can work together (Bannon 1997). It is needed for coordination, communication, and information and knowledge transfer (Ackerman 2000). Historically, scientific research was typically done in a single lab and within a single lab where having a common work space was not an issue. However, with the advent of large complex collaborations brought on by more complex scientific questions and related funding initiatives, having common workspace for everyone involved in a research initiative is often no longer possible (Cummings 2005). Large scientific collaboration frequently involves scientists from multiple institutions that are geographically dispersed. Because collaborative research cannot succeed without some type of a common place where researchers can work together, information tools can be used to create a virtual common space.

In order to alleviate the lack of shared physical space due to distance; technologies have been developed to give researchers a simulated co-located space. Such a virtual workspace constitutes a place where research information (e.g., data, analysis results) can be collected and accessed by the researchers involved in collaboration (Appelt 1999). Through these technologies, researchers can come together, share data, analyze findings, and conduct research discussions. Some examples of collaborative technologies designed to address this particular issue include Basic Support for Cooperative Work (BSCW) (Hoffelner; Appelt 1999) (Figure 3.1), Wikis (Wagner 2004) (Figure 3.2), and Microsoft Sharepoint (MS Sharepoint; Barga 2007) (Figure 3.3). BSCW is "a 'shared workspace' system which supports document upload, event notification, group management and much more" (BSCW). A Wiki is "a collaboratively created and iteratively improved set of web pages, together with the software that manages the web pages" (Wagner 2004). Microsoft Sharepoint provides a central platform to connect and share knowledge. It supports "intranet, extranet, and web applications across an enterprise within one integrated platform" (MS Sharepoint). These virtual workspace technologies are known to reduce negative impacts of remote collaboration (Cummings 2005).

The minimal common workspace/coordination support for multi-institutional collaboration should include a common repository of information such as a shared server and a way to coordinate schedules. However, more advanced tools such as Sharepoint or customized workspaces such as BSCW can provide better support for collaboration. Furthermore, shared calendaring such as the Microsoft Exchange system (MS Exchange) can also help coordinate schedules among collaborators.

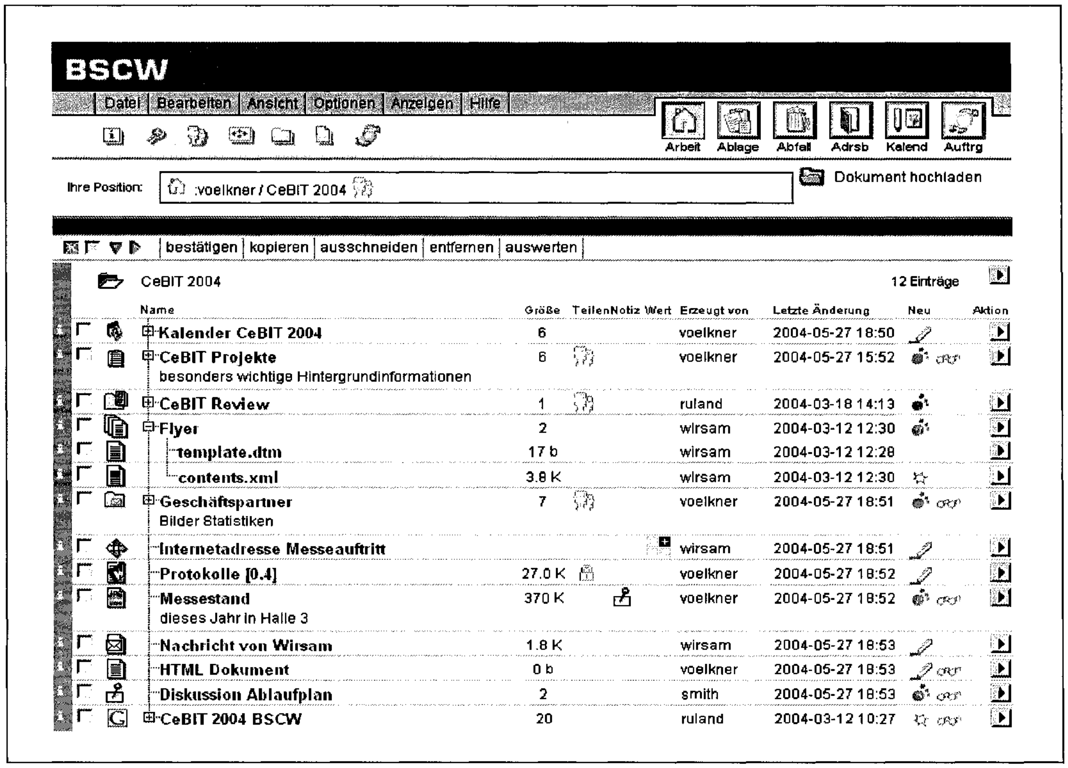


Figure 3.1

Basic Support for Cooperative Work (BSCW) Interface
(http://www.fit.fraunhofer.de/products/bscw_en.html)

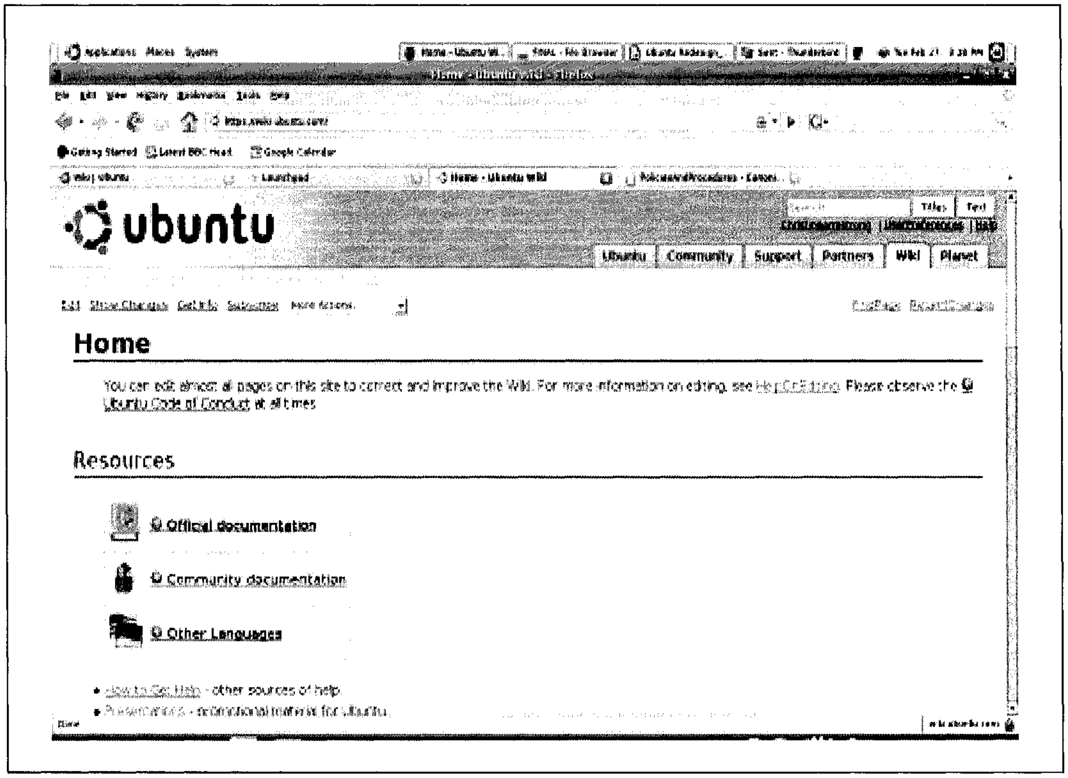


Figure 3.2
Wiki Interface (http://www.fit.fraunhofer.de/products/bscw_en.html)

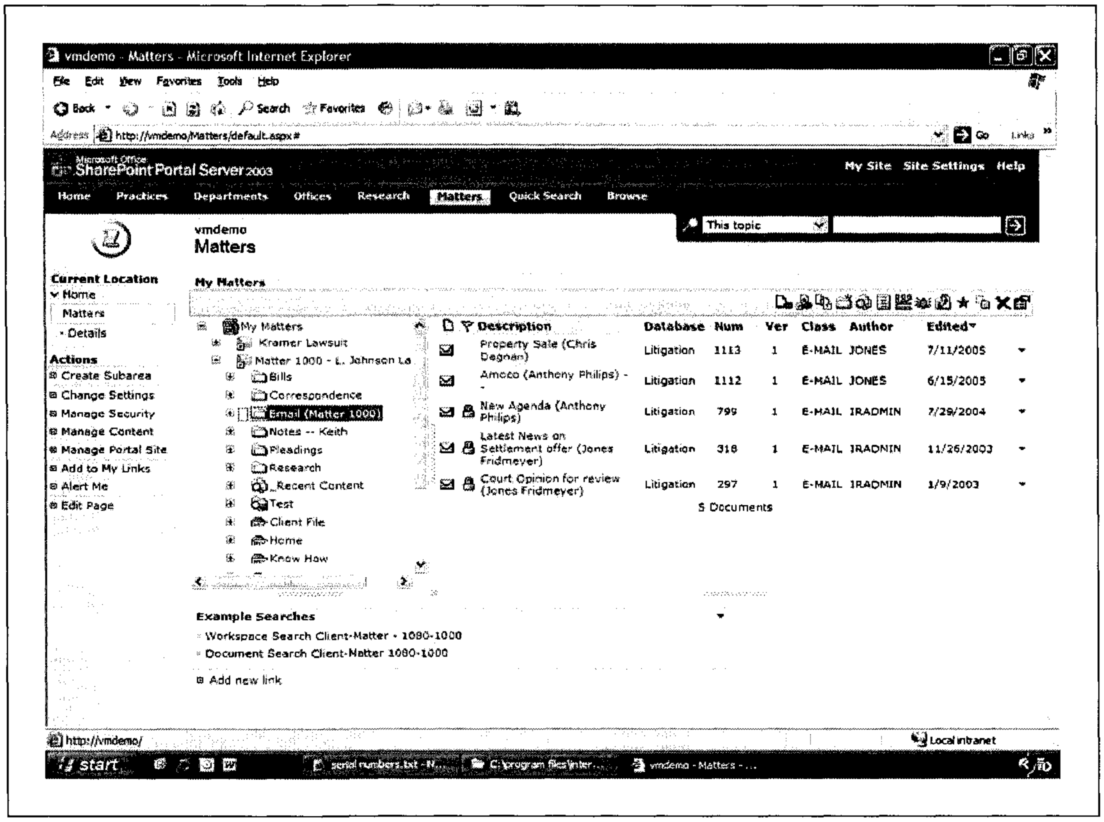


Figure 3.3 Sharepoint Interface
(http://www.officeinfo.com.au/Products/Interwoven/WorkSite/WorkSiteforSharepoint.asp)

Approaches to Data Sharing and Data Management

Data sharing and management have to occur as a necessary part of any scientific collaboration. Large amounts of data are generated throughout collaborative research. Data can be the result of an experiment performed by a single lab or multiple labs or the result of aggregated analyses of existing forms of data. Regardless of the type of data generated, they are often shared and analyzed by all individuals involved in collaboration. Hence, data sharing and data management go hand in hand. Data sharing in science is important for several reasons: 1) a researcher's findings can be validated by peers, 2) new analyses can be performed on existing data, 3) data can serve as the basis for new research, and 4) it can prevent unnecessary duplication of effort (Birnholtz 2003; Ball 2004). Furthermore, new knowledge and insight can emerge from aggregated data that were not visible or obvious in individual sets of data. For these reasons, funding agencies such as NIH have been increasingly requiring data sharing as part of their funding initiatives (NIH data; NMEDCTS).

Data sharing in large collaboration often involves technologies that enable data exchange. Data sharing tools can be as simple as spreadsheets sent over email or they can be more complex systems like a common database or a repository to which researchers can submit their data. However, even if there are mechanisms to share data, sharing cannot occur without trust and willingness to share from participants. Although data sharing is generally perceived as beneficial, researchers often find sharing difficult due to the competitive nature of scientific research and the cultural tendency toward not sharing (Campbell 2002). Researchers often do not trust others involved in collaboration, thinking that a collaborator might steal their findings (Birnholtz 2003; Olson 2005). Such mistrust often stems from the highly competitive nature of biomedical research (Campbell 2002; Olson 2002; Birnholtz 2003). When designing and implementing data sharing technology, such social barriers should be taken into account.

Support for multi-institutional collaboration should include some tools that enable management and sharing of data across institutions. Data management tools could be as simple as a spreadsheet and as complicated as a large scale database. Data transfer can be a spreadsheet sent as an email attachment or be as complex as a web-based standardized transfer system that every institution in collaboration uses. Although even simple tools can enable data sharing, more advanced tools are needed as collaborations get larger and more complex.

3.2.2 Concepts Specifically Relevant to Biomedical Collaboration

During the development of the preliminary framework, four concepts were identified as particularly relevant to collaborations in the biomedical field. These are concepts that might not be applicable to collaborations in other fields, but are essential to biomedical collaboration. These concepts include: 1) approaches to data integration and analysis, 2) security protocols, 3) tracking data through metadata and data provenance, and 4) interoperability and data standards.

Approaches to Data Integration and Analysis

Even if data are managed and shared well, they cannot be fully taken advantage of in collaborative research unless different types of data and analysis are integrated. For example, an image generated by a microarray experiment is useless without relevant statistical analysis to indicate what part of that image is significant (Allison 2000). In addition to accessing and reviewing individually analyzed data, such as the microarray experiments and analyses described above, in collaborative research, researchers often need to collectively analyze data. The collaborative analysis has been described as "an interactive process of brainstorming where researchers share their individual interpretations, understanding, and insights, which build upon one another to form cogent findings and conclusions" (Chin 2004). Because in collaborative research input is often needed from a number of researchers located remotely, the collaborative analysis must occur through a technical medium.

Collaborative data analysis must often involve sophisticated statistical or analytic tools as well as a mechanism for the researchers to tie the analysis processes together (Chin 2004). Integrated analysis technologies enable scientists to view and analyze data in one place, share tools during analysis, and capture notes, working ideas, and interpretations. These technologies can enable the process of analysis to occur in real time or asynchronously, depending on the type of analysis needed and researcher preferences. Probably the most important function of integrated analysis technology in biomedical research is data integration. Collaborative analysis of the data generated in collaboration cannot happen without integrating data generated by different individuals and institutions involved in the collaboration. The concept of data integration is widely recognized as an important issue and is actively being researched in biomedical research due to the advent of data explosion challenges (Altman 2006; Louie 2007).

The minimal common data integration and analysis support for multi-institutional collaboration should include tools that enable merging of various datasets in one place and tools to access and analyze several datasets at once. In their review, Louie et al. (Louie 2007) suggest various existing data integration approaches that complex collaborations can take advantage of. Of the suggested approaches, database federation and peer data management approaches best fit the multi-institutional collaboration since these approaches deal with data at disparate locations. Data analysis tools can also be as simple as a single spreadsheet, or a complex software package such as Rosetta Bio-software (Rosetta). Although most labs are currently utilizing spreadsheets (Anderson in Press), more advanced tools are needed for larger and more complex collaborations.

Security Protocols

Modern biomedical research cannot begin without a well-developed security protocol. Biomedical research often involves highly sensitive data such as private medical records or individual genetic information. Personal health data are legally protected by the Health Insurance Portability and Accountability Act (HIPAA). HIPAA mandates personal health information to be constantly under tight security protection (HIPAA). Funding agencies also require that any biomedical research project have a carefully planned security structure (CTSA RFA). In addition to satisfying HIPAA and funding mandates, it is possible that tight security constraints can alleviate the mistrust among researchers not wanting to lose control of their data (Birnholtz 2005).

Although a general security structure is the first step toward fulfilling the security requisites constraints of biomedical research, it alone is not enough. For example, certain data in biomedical research are more sensitive than others. It is also possible that some more sensitive data needs to be only available to a subset of researchers within a collaboration. To support all levels of constraints, a flexible security structure needs to provide differential access to a common collaborative system (Chin 2004). A minimal collaborative security structure should involve authentication (minimum login/password) and authorization (variable access to different parts of research data and workspace) as well as encryption of communication and audit trails to allow review of record accesses. The security structure of a collaborative research project must also be compliant with HIPAA and funding agency mandates.

Tracing Data through Metadata and Data Provenance

Metadata and data provenance provide important information about data used in the process of collaborative research. Before the era of large scale collaboration, the data sets created in biomedical research were manageable in size and narrow in scope. Modern collaborative biomedical research, however, generates large quantities of highly diverse data (Ball 2004). Furthermore, collaborative biomedical research often involves researchers from various fields with different set of knowledge and often from multiple institutions. For example, a research collaboration might involve a bench scientist sequencing genes at one end, while at another end involve a clinical researcher performing clinical trials of a drug for cancer (Mankoff 2004). To facilitate sharing and collaborative analysis, collaborative technologies must bridge the gap of diversity in data and variability in researchers' field of knowledge. Metadata and data provenance can help bridge that gap (Simmhan 2005).

Metadata is a description of data, data about data (Buetow 2005). Since collaborative research involves researchers with different backgrounds, some of them can find it difficult to understand data without the context of how research was conducted and how the data were analyzed (Allison 2000). For example, a biochemist discovers a possible treatment for cancer and gives resulting data to a clinical researcher. Without some explanation or description of the data, the clinical researcher would find the data impossible to decipher. Data provenance is related to metadata in a sense that it is essentially concerned with the history of data (Simmhan 2005), the importance being that it functions like a versioning mechanism. Without a detailed data history, researchers have no way of determining whether data is outdated or still valid. Not having data provenance in collaborative research could result in confusion where some researchers access outdated or invalid data while others use the most updated data. Data provenance enables scientists to keep track of large data sets generated by collaborative research.

A minimal level of metadata and provenance is maintained by keeping careful track of files and comments within the files since many biomedical researchers are still using spreadsheets for data analysis and sharing (Anderson in Press). However, multi-institutional interdisciplinary biomedical collaboration can benefit greatly from more complex tools for managing metadata and provenance. Simmhan et al. (Simmhan 2005) reviews five tools that enable metadata and data provenance. Tools such as Taverna (myGrid) (Zhao 2007) are gaining popularity in the biomedical field.

Interoperability and Data Standards

Interoperability is a concept that is assumed to exist in collaboration, but rarely explicitly mentioned. Most literature on collaboration implies interoperability by stating that technology involved in collaboration must be able to work together (Olson 2005)or saying common infrastructure support exist (Finholt 2003). Very few, however explicitly discuss the importance of interoperability. Interoperability is a term used to describe a common infrastructure that seamlessly integrates technologies from all levels of research within a given collaboration. It includes not only technological infrastructure, but also interoperability of data structures. Data standards enable interoperability by creating a common frame of reference for data sharing. Heterogeneous sets of data are able to converge through a common frame of reference so that researchers can easily share and analyze data (Neches 1991).

Although a necessity, interoperability is difficult to achieve in collaborative biomedical research due to the heterogeneity of technologies and data formats used by different collaborators (Maojo 2003). For example, often institutions involved in collaboration use different operating systems. Some use Unix, some, Windows, and others can use a completely different platform; hence, even reaching a basic level of interoperability is often difficult. The difficulty escalates when the collaboration needs to standardize heterogeneous data for sharing and analysis. For example, a genetic sequence of a disease gene is significantly different from an x-ray image of a tumor; yet both must be integrated into a common system to support collaboration. Interoperability should provide a way to enable research processes such as data analysis regardless of differences in system platforms or heterogeneity in data. A minimal approach to interoperability can be a simple standard that connects various pieces of data together. At its most complex, interoperability can be a complex central system with various tools that span multiple institutions and enable sharing of complex data (i.e. caBIG, BIRN) (BIRN; caBIG; Buetow 2005).

3.2.3 Environmental Factors that Support Collaboration

Large scale collaborations can be situated among several academic and non-academic institutions. Since collaboration cannot exist in a vacuum, the policies and politics of the institutions involved in collaboration can have positive or negative effects on the collaboration. Although the surrounding environment is important to address in collaboration research, very few studies have examined the environment surrounding collaboration. The majority of previous studies have focused only on the technical or the social issues within the collaborations, leaving out the factors that might influence collaboration from the outside. Three critical environmental factors that are important to collaboration are outlineded in this preliminary framework. They are: 1) overseeing collaboration with administrative and management structures, 2) institutional support to relieve technical difficulties, and 3) available funding enabling the formation of collaboration.

Overseeing Collaboration with Administrative and Management Structure

Having a good administrative and management structure in a multi-institutional collaboration enables better collaboration. According to Olson et al. (Olson 2005), an administrative and management structure is composed of individuals that oversee, coordinate, resolve conflicts, and make decisions within a collaboration. Ideally, this administrative body supports the overall collaborative process and functions as a central reference point for those involved so that no one in a collaboration feels isolated (Olson 2005). Furthermore, an administrative and management body also manages all the legal issues that might arise related to collaboration (David 2003). In small collaborations such as one that involves two small labs, one of the principal investigators can essentially act as an administrative body that manages collaboration. In large collaborations, however, the amount of work involved in managing collaboration is too great for any one investigator; therefore, a designated administrative body is often in charge of managing collaboration as whole.

Support to Relieve Technical Difficulties

Large collaborations often involve a wide range of technical tools. For example, rather than walking a few doors down to talk to a collaborator, researchers often have to rely on remote communication technologies such as email or phone. They also have to use a variety of complex tools to remotely share and analyze data. Yet researchers are not inclined to dedicate time and energy to learning these complex tools, they feel that the primary objective of collaboration is to advance scientific knowledge, not to learn how to use a set of tools. Such extra work that is required to learn technology often deters researchers from continuing with the collaboration. For example, the Worm Community System (WCS) was faced difficulties of acoption because researchers had to learn to install complex systems and use unfamiliar UNIX operating system (Finholt 2003). The Upper Atmospheric Research Collaboratory (UARC) researchers found frequent system downloads a huge challenge (Finholt 2003). Having technical support personnel can alleviate such technical difficulties and benefit collaboration by removing potential causes for failure (Olson 2005). At a minimum, a MIBC should have a available technical support to help with difficulties of collaborative infrastructure.

Available Funding to Enable the Formation of Collaborations

Large scale collaboration typically has a large financial overhead due to the technical infrastructure necessary to support the processes of scientific research and collaborative work among a large group of people. To alleviate financial difficulties of large scale collaboration that can cuase both initial and continued adoption problems, adequate funding structure should be available as part of the environment fostering collaboration. Olson et al. have found that collaborations based on funding initiatives are more likely to fail since a lasting collaboration cannot be based purely on financial incentives (Olson 2005); however, without adequate funding, it would be difficult for large collaborations to form. Funding agencies should recognize that MIBCs are difficult to achieve and often require more resources to bridge the physical distance and differences among its members. Providing extra resources might help facilitate more MIBCs to form.

3.2.4 Factors that Support Long-term Collaboration

A number of research studies and technological developments have made collaboratories possible. Developments in communication, data sharing and analysis technologies have helped to alleviate the difficulties specific to collaboration at distance. The social issues in collaboration such as willingness to share and trust have been identified and researched in depth. Yet despite growing research in the area of collaboration, very little is known about how to sustain collaborations over a long period of time. Many earlier large scale collaborations simply ended when their initial funding period ran out. Others dissolved due to the overwhelming challenges they faced during their first few years of existence. During the development of the preliminary framework, two factors emerged as those that might help foster long-term collaboration: a) collaborative process and technology training, and b) long-term planning.

Collaborative Process and Technology and Training

Technology plays an important part in modern collaboration. It enables everything from simple every-day communication between researchers, to sharing of complex research data. However, since technology is often perceived as difficult to use, such dependence upon it in collaboration can negatively affect the collaborative process. The knowledge of the research process particular to each collaboration is also important for the success of collaboration. Due to the variability in backgrounds, the researchers involved in collaboration often focus only on their local research objectives and overlook the larger goals of the collaborative effort. It is often overlooked fact that it is important for all collaborators to have an understanding of the overall goal and the high-level scientific processes of the entire collaboration. Technology training and research process training can help foster a more successful collaborative environment through a better understanding of the overall collaboration process and more effective and easier use of technology (Olson 2005). Minimal training can include basic courses on tools being used and holding open conferences to introduce various aspects of collaboration.

Long-term Planning

Continued long-term planning of the infrastructure development, process, and on-going objectives of the multi-institutional collaborative project, is key to fostering long-term collaboration. Collaborations often form out of necessity, either due to a need for additional expertise or in response to funding requirements. Therefore due to this "on-the-spot" formation, collaborations seldom begin with long-term plans. When plans are made, they often do not extend beyond the initial funding period due to uncertainty in funding, even when the complex problems the collaborations seek to solve cannot be expected to be resolved within the first funding period. Yet, if collaboration is to last, long-term planning of goals, infrastructure and the overall collaborative process is needed from the beginning (Olson 2005). Lack of such planning has contributed to the dissolution of collaborations before their aims were achieved (Olson 2005). For long-term planning of collaboration, it should be considered that collaboration is a dynamic process that changes constantly. Supporting collaborative research through a continuous evaluation of needs and the identification of gaps in current processes is imperative for the survival and success of a long-term collaboration (Sonnenwald 2003; Kaplan 2004; LaCoursier 2004). MIBCs should consider, at minimum, an annual evaluation to identify needs and gaps that arise within the collaboration.

3.3 Applying the Framework to Biomedical Informatics

Section 3.2 described in detail the preliminary framework that combines concepts from various collaboratory studies in the context of the biomedical setting. This section illustrates how each dimension of this framework can be applied to biomedical collaborations. A motivating scenario is used to demonstrate how each dimension of the framework is important to make collaboration effective.

3.3.1 Scenario

Consider a hypothetical case: Several investigators at multiple institutions across the United States are all interested in how a certain antibody might be used for treatment of human breast cancer. They wish to create a computer model of the structure of the antibody and the receptor to which it binds as to permit simulation of a variety of modifications to the antibody that might increase its binding and effectiveness against the cancer. The investigators involved in this collaboration have different resources to contribute and different areas of expertise. For example: a) clinical investigators have expertise in the use of antibodies in the treatment of cancer through clinical trials, b) computer scientists have expertise in simulation algorithms and modeling, and c) translational researchers have developed a model of breast cancer and preliminary data that in this model the antibody they are studying helps treat the breast cancer. In order for these investigators to successfully conduct research together, they will need to:

- Share data (i.e. clinical outcome measures, cancer model outcome measures, simulation constants, and human body models)
- Communicate with all the investigators involved in the collaboration at both local and affiliated institutions
- Integrate all the various pieces of heterogeneous data, information and knowledge spread across different groups
- Work with local administration and institutional structures (i.e. IRBs) to approve the study procedures

3.3.2 Applying the Framework to the Biomedical Research Scenario

General Concepts that are Necessary in any Collaboration:

The Nature of Remote Communication and Communication Technology, Common Workspace and Coordination Tools to Bridge Distance, Approaches to Data Sharing and Data Management

In the scenario above, the investigators are geographically distributed across multiple institutions. Because most researchers are not co-located, frequent face-to-face meetings to discuss research and share ideas are difficult to achieve. To facilitate remote communication, a technical medium needs to be employed since collaboration is not likely to happen otherwise. As described in Section 3.2.1, systems to facilitate communication among investigators may include tools as simple as phone or email or more complex tools such as videoconferencing systems.

However, providing a means of communication is not enough to facilitate collaborative research. When meetings occur in a co-located physical space, investigators bring their research data and notes and discuss the research and how to coordinate efforts. Such sharing and coordination are integral to collaborative research. A virtual workspace for investigators in the scenario can enable such work. Using a virtual workspace, the investigators can access data from a common location as well as communicate through discussion forums and chat and coordinate their efforts through a scheduler and common calendaring tools.

The investigators from the scenario describe above will need to share and manage data during every phase of the study. At the design phase, plans will need to be developed for data capture (e.g. who will capture data about the behavior of the antibody in a human body), data management (e.g. where and how all the data collected will be managed), and data sharing (e.g. what infrastructure will be used to allow the computer scientists to access and integrate clinical and animal models). During the execution phase, systems will need to be in place to enter, store, and monitor the highly diverse and heterogeneous information from each subgroups (i.e. multiple modelers, algorithms, simulations, and animal models).

Concepts Specifically Relevant to Biomedical Collaboration:

Approaches to Data Integration and Analysis, Security Protocols, Tracing Data through Metadata and Data Provenance, Interoperability and Data Standards

The investigators in the scenario above need mechanisms to integrate heterogeneous data (i.e. data from clinical research, computer models, animal models, algorithms for simulation) from groups scattered across several institutions. The data will need to be well integrated so that a common set of tools can be used across the data to uncover and distill relevant information. As mentioned in Section 3.2.2 data integration is especially important for interdisciplinary collaborations because they tend to work with highly heterogeneous data.

The issues of data privacy and security have become more prominent in recent years in part due to regulatory pressures (HIPAA) and especially so in biomedical research due to the recognition that it is very difficult to de-identify genomic data (Altman 2002; Malin 2005). When supporting collaborative biomedical research, such as the research described in the scenario, the biomedical informatics community should consider regulatory issues not only at the individual institutional or laboratory level, but also at a collaborative, cross- institutional level. Having a security structure as described above in Section 3.2.2, is important to ensure that sensitive patient information obtained from clinical investigations are safe and are protected within the bounds of HIPAA and institutional regulations. Appropriate security measures will also ensure that investigators can more comfortably share data with their collaborators, knowing that the novel data and findings they share within the collaborative system will not be accessed by unauthorized individuals, such as competing researchers.

Even if a mechanism for data integration and protecting data security exists, the use of such data in the described scenario might be limited if no information exists about the data. For example, in our biomedical collaboration scenario, a clinical investigator who receives data from a computer scientist may not be able to interpret it or even understand it without associated metadata explaining what the data represents. A mechanism for versioning and keeping a historical record of data also needs to exist. A translational researcher would not be able to trust data he receives from a clinical investigator unless he can be sure he has the latest version and knows who has accessed and modified the data.

Data flowing from computer scientists to clinical investigators to translational researchers in the above mentioned collaborative scenario cannot be effectively used without an existing common infrastructure. A collection of interoperable systems using a common interface will present to the investigators using the system a seemingly common infrastructure where they can gather all the information they need in one place instead of searching for information at various different sources across institutions. Adoption of standards by the community is key to enabling different systems to interoperate.

Environmental Factors that Support Collaboration:

Overseeing Collaboration with Administrative and Management Structure, Support to Relieve Technical Difficulties, Available Funding to Enable the Formation of Collaborations

Biomedical informatics support for biomedical collaboration is often thought to end with the technology; however, the biomedical informatics community should consider issues that extend beyond just the technical infrastructure. The organizational and environmental issues surrounding collaboration can significantly affect the technological and organizational implementation of a collaborative infrastructure. Managerial structure, technical support, and adequate funding are factors crucial to the survival of collaboration that are often overlooked. Biomedical informatics should take part in the development of the collaboration to ensure a financially viable infrastructure is set up and resources to provide needed technical support are included. For example, the investigators in the above scenario will need a central resource to help set up a low-cost collaborative infrastructure. During the research process, they will also need technical support to learn to navigate the various data management and communication systems.

Factors that Support Long-term Collaboration:

Collaborative Process and Technology and Training, Long-term Planning

Most biomedical researchers are not familiar with the process or the technologies involved in collaborative research. The investigators in the scenario described above would benefit greatly from general training to inform them about the common technical infrastructure that exist for the collaboration, tools such as virtual workspace and discussion boards. Biomedical informatics community can also contribute to long-term planning by performing continuous evaluations throughout the collaborative process and making appropriate adjustments to the infrastructure to adapt to the changing needs of collaboration.

3.4 Summary

In Chapter 3, a preliminary framework of biomedical collaboration (Table 3.1) was qualitatively developed using existing literature on collaboration. *The Nature of Remote Communication and Communication Technology, Common Workspace and Coordination Tools to Bridge Distance, Approaches to Data Sharing and Data Management* are general concepts in collaboration that are essential to all collaborations. *Approaches to Data Integration and Analysis, Security Protocols, Tracing Data through Metadata and Data Provenance, Interoperability and Data Standards* are concepts specifically relevant to complex, multi-institutional biomedical research. These are also key areas for biomedical informatics in terms of research and development of applications to support collaboration. Also identified are critical environmental factors that support collaboration (*Overseeing Collaboration with Administrative and Management Structure, Support to Relieve Technical Difficulties, Available Funding to Enable the Formation of Collaborations*) and factors that support long-term collaboration (*Collaborative Process and Technology and Training, Long-term Planning*). These factors are important to all aspects of biomedical research collaboration.

The preliminary framework of biomedical collaboration (Table 3.1) provides insight into how biomedical collaboration can be supported by the biomedical informatics community. It can guide planning and design of collaborative infrastructure to better support interdisciplinary biomedical collaboration. In particular it has implications for biomedical informatics researchers and developers who are working on ways to support biomedical collaborations. However, the preliminary framework is not yet complete in that it has never been verified in the field. This preliminary framework will be used to guide the field study of MIBCs described in Chapter 4.

Chapter 4: Theory and Method to Study Biomedical Collaboration

A wide range of theories and methods to conduct in-depth contextual study exist in fields such as Library and Information Science, Computer Science, Communications and Organizational Studies. This is due to a recognition that a thorough understanding of a complex setting resulting from contextual studies has been an important step in better supporting users in that environment. Some of the contextual studies have led to better system design and adoption while others have provided different types of support for processes occurring within that environment (Bødker 1993; Muller 1993; Sumner 1997; Weng 2006). This study attempts to gain a better understanding of Multi-institutional Interdisciplinary Biomedical Collaboration (MIBC) to better support the overall collaborative process. In order to do so, a contextual study of two MIBCs was conducted with the concepts in the preliminary framework of biomedical collaboration developed in Chapter 3 as a guide. This chapter briefly introduces in Section 4.1 and 4.2 the theories and methods that were used to conduct the contextual study. Then in Section 4.3, the actual study steps are described in detail.

4.1 Theory

Grounded Theory (Glaser 1967) is a systematic methodology that often uses qualitative methods to generate theory from data through iterative categorization of themes. Since its inception, Grounded Theory has diverged into two different schools of thought due to the differences in opinion of the original creators, Glaser and Strauss. Glaser maintains that grounded theory focuses on deriving theory based purely on data gained from the study environment (Glaser 2001). In this school of thought, the researcher goes into the field with no prior assumptions and without any bias from a particular theoretical framework. On the other hand, Strauss interpreted Grounded

Theory as one that uses theoretical frameworks to guide the fieldwork (Strauss 1998). In this model, the research can enter a field with some prior sensitivity to existing theory and thus is able to compare the theory in context.

Although the researchers' mindset going into the field might be different, both schools of Grounded Theory have remarkably similar approaches to research. They both approach research with no formal hypothesis and try to generate explanatory theories of certain human behavior through systematic data collection and analysis. Data is often collected through the use of qualitative methods such as interviewing and observation. Furthermore, data collection and analysis often occur simultaneously and through constant comparison. The analysis process occurs through reading and re-reading and reflecting on data where patterns in data and relationships between them are identified. Grounded Theory is often used in the study of complex settings where there is little prior knowledge of the setting (Bernard 1994; Morse 1995).

As stated in Chapter 2.2.1, there is a general lack of an all encompassing theoretical framework that can be used to guide the study of collaboration. Although the Theory of Remote Collaboration (TORC) partly fulfills the role of the all encompassing collaboratory theory, the broad nature of the theory makes it difficult to apply to a specific domain "as-is". That is why, as described in Chapter 3, a preliminary framework of collaboration that is more specific to the biomedical field was developed through a literature review. Yet neither of these theories is complete by itself and neither has been verified in the field. That is why this study uses the Strauss interpreted Grounded Theory approach in an attempt to derive a framework based on both TORC and the preliminary framework. The use of Grounded theory firmly bases the emerging framework in real data, while extending existing theoretical knowledge.

4.2 Qualitative Methods

Studies guided by Grounded Theory often use qualitative methods since they enable the researcher to gain in-depth understanding of individuals, groups and settings.

Thus far, qualitative methods have been used in various fields. They help to understand participants and their experiences, the particular context in which participants act, identify unanticipated phenomenon, and the process by which events and actions take place (Maxwell 1996). The representative common forms of qualitative methods are in-depth interviewing, observational field research, and analysis of existing documentation in the setting.

4.2.1 Common Data Collection Methods

Interviewing

According to Denzin and Lincoln, interviewing is one of the most "common and powerful way to understand our fellow human beings" (Denzin 1994). This is due in part to whatever thought process that is going on in individuals' heads is not otherwise accessible to others. There are three basic forms of qualitative interviewing: a) unstructured, b) semi-structured, and c) structured (Crabtree 1999; Patton 2002). The structured interviewing approach/form starts with a set of formalized questions and pre-determined choices of answers; therefore, it is survey-like and has very little variability in answers. In semi-structured interviewing, the researcher is guided by a set of questions and topics, but has no pre-formed answers he seeks. The unstructured interviewing is by far the loosest form of interviewing. During unstructured interviewing, the researcher takes the interview to wherever the interviewee leads within the confines of a general topic. Each of these three forms of interviewing have strengths and weaknesses depending on the research and the question being answered.

The interviews are typically recorded, contents de-identified, and transcribed for analysis. Although interviews are a powerful method of collecting otherwise difficult data, there are several weaknesses to interviewing method which might result in erroneous results. Sources of errors include: a) respondent giving socially desirable response instead of truth, b) confusing wording of questions resulting in answers to wrong questions, and/or c) interviewers with flawed questioning techniques leading the interviewee to answer questions in a certain way (Denzin 1994). The specific methods used in this study will be described in section 4.3.4.

Observational Fieldwork

Direct observations in the field are a rich source of data. Observational field notes often contain some or all of the following: a) space: information about a physical place, b) actors: people involved, c) activity: set of related acts people do, d) object: physical objects present, e) act: actions that people do, f) event: set of related activities that people carry out, g) time: sequencing that takes place over time, h) goal: what people are trying to accomplish, and i) feeling: emotions felt and expressed (Crabtree 1999). There are several different ranges of observational roles, from a role of full participation in the setting to a complete spectator, and everything else in between. The observational notes are often typed up and their contents are used as data in the analysis (Patton 2002).

Observational fieldwork is a powerful tool in that it gives a more objective account of what is occurring in the field. Unlike interview data, observational data are not clouded by participants' points of view. Nevertheless, observational field data also have their weaknesses. One common weakness might be that those being observed might behave differently than under normal circumstances. It is also possible the observer might lose objectivity due to prolonged exposure to the setting (Patton 2002). The specific methods used in this study will be described in section 4.3.5.

Artifact Examination

Complementing interviews and observations, documents and other artifacts associated with the setting can shed a different light into the complex environment. Artifacts can sometimes give information about the setting that is otherwise not possible to obtain. For example, an overall history of the organizational structure or processes unclouded by any one participant can be glimpsed at through document examination. The artifacts also provide snap shots of some of the processes that occur in the setting (Patton 2002). However, due to the static nature of artifacts, they are used more as supporting data than as central data. They may contain data that is out of date or incomplete and might be inconsistent with the current setting. Furthermore, the artifacts can be difficult obtain and once obtained, it can be hard to determine what is of interest and what is not. The specific methods used in this study will be described in section 4.3.6.

4.2.2 Triangulation

Triangulation means employing a combination of methods, investigators, and/or theories that complement one another to study same phenomenon or process (Denzin 1994). The word triangulation has roots in radio triangulation where the use of directional antennas set up at two ends of a known baseline determines the point of origin of a radio broadcast. In qualitative research, triangulation helps researchers to guard against research bias by drawing on a variety of data, and to gain deeper and clearer understanding of the setting and people being studied. There are three common types of triangulation: a) data triangulation: using multiple sources of data to study a single setting, b) investigator triangulation: having more than one investigator involved in the study, and c) methodological triangulation: using multiple methods to study a single setting. The specific methods used in this study will be described in section 4.3.7.

4.2.3 Analysis

There are four general steps to analysis in qualitative research guided by grounded theory (Bernard 1994; Strauss 1998; Patton 2002):

1. Produce data (i.e. transcripts from interviews, observations, and/or artifact examination).
2. Identify, name, categorize and describe themes (codes) in data

3. As themes emerge, relate themes (categories and properties) to each other according to how they fit into the subject under query
4. Develop a central framework that ties all categories together by identifying a core category and relating all other categories to that core

Step two, commonly called "open coding" is done at the level of each line of transcription. Open coding is a step where the researcher carefully reads produced data and identifies any themes (codes) associated with the study question. Step three, also called "axial coding" is an iterative process that categorizes, re-categorizes, condenses, and identifies relationships among the generated codes. Step four, called "selective coding" where the identified themes and relationships are formalized toward a theory (Strauss 1998). The specific methods used in this study will be described in section 4.3.8.

4.2.4 Trustworthiness

Internal and external validity, reliability and objectivity that are often used to measure the dependability of quantitative research are inapplicable to qualitative studies. Instead, qualitative research uses a slightly different, but equivalent set of criteria to show the legitimacy of study findings. Validity in qualitative research is trustworthiness and has its own equivalents to quantitative validity (see Table 4.1) (Lincoln 1985). The specific methods used in this study to maintain trustworthiness will be described in section 4.3.10.

Table 4.1
Trustworthiness in qualitative research vs. validity in quantitative research. Equivalent terms.

Qualitative Term	Equivalent Quantitative Term
Credibility	Internal Validity
Transferability	External Validity
Dependability	Reliability
Confirmability	Objectivity

Credibility

There are five common activities that increase the credibility of findings: a) prolonged engagement, b) persistent observation, c) triangulation, d) member checking, and e) peer debriefing. Prolonged engagement involves engaging in a setting for a sufficient amount of time to learn the culture, minimize distortions, and build trust. It provides a more in-depth view of a setting. Persistent observation occurs when a researcher identifies and assesses relevant factors and atypical findings. Persistent observation provides depth to the study. Triangulation, as explained above refers to using a variety of data, methods, and/or investigators in a study. In member checking, data, analysis and interpretations, and conclusion are presented and tested by those from whom the data were originally collected. In peer debriefing, the researcher shares and discusses his research process with a disinterested peer in order to gain further objectivity of his research (Lincoln 1985; Patton 2002).

Transferability

Transferability is the qualitative equivalent to external validity and can be achieved through "thick description". "Thick description" provides an in-depth description of setting and methodology involved in the study so that it can be meaningful to outsiders. It enables other researchers to understand whether the conclusion of the research in question can be transferred to another study. Transferability is very different from the conventional external validity in that it does not claim that data can be generalized to apply to multiple settings (Lincoln 1985; Patton 2002).

Dependability

Dependability informs about the consistency and reliability of the study. It is able to inform whether the study, if replicated with same or similar participants in a similar context, can give repeatable findings. In order to have dependability, the study must take into account any factors that might lead to change in study process or findings. Dependability can be achieved through an "audit trail" which consists of keeping meticulous field notes and constantly comparing a variety of data (Lincoln 1985; Patton 2002).

Confirmability

Confirmability informs about the objectivity of the study. Because by nature qualitative studies cannot be completely objective, the researcher using qualitative methods recognizes that objectivity as an illusion. That is why qualitative researchers do not attempt to ensure their data are free from subjectivity of the researcher. Rather, they trust in that data can be tracked to their sources and that the logic used for analysis is coherent. Confirmability can be best achieved through an audit trail and reflexive journaling, a diary in which the investigator, as needed, records a variety of information about the study and methods (Lincoln 1985; Patton 2002).

4.2.5 Qualitative Methods in Collaboratory and Biomedical Settings

Applications of Qualitative Methods in Collaboratory Settings

As described in Chapter 2.3, many of the collaboratory studies have employed Participatory Design (PD) (Bødker 1993; Muller 1993; Sumner 1997; Weng 2006) and User Centered Design (UCD) principles (Beyer 1998; Consolvo 2002). Collaboratory designers used these design principles because they understood that the adoption of a system is highly impacted by the complex environment that it is to be used in. Qualitative methods are often employed during PD and UCD because these studies occur in complex settings. For example, UARC development was based on a number of field visits (Olson 1998). SPARC was developed after much time spent learning users' work practices through field studies (Olson 2005). The use of qualitative methods has helped designers better identify specific needs of each complex setting and has resulted in better adoption of the system developed. Section 4.3 will describe how this study used qualitative methods to study two complex settings.

Applications of Qualitative Methods in Biomedical Settings

The first well known case of using qualitative methods in a biomedical setting is described in Latour's 1979 study of researchers in a neuroendocrinology lab at the Salk institute (Latour 1979). Latour's work highlighted the complex scientific setting and shed some light on the life of scientists at work. However, Diana Forsythe was the one who conducted pioneering work on exploring socio-technical issues in the biomedical field (Forsythe 1992; Forsythe 1996; Forsythe 1998) . Forsythe paved the way for social scientists to apply qualitative methods to study the complex biomedical setting. Following in the footsteps of Forsythe, other researchers have used qualitative methods to study system implementations and the effects of systems in the medical settings. For example, Ash et al studied socio-technical factors surrounding Computerized Physician Order Entry (CPOE) systems using observational fieldwork and interviews (Ash 2003; Ash 2005). Through the use of qualitative methods they were able to gather factors that influence successful implementation of CPOE systems. The use of qualitative methods in other biomedical settings was also able to inform better design and implementation of systems as well as evaluations of systems (Bartlett 2002; Bartlett 2005; Shachak 2007 prepring; Anderson in Press). Section 4.3 will describe how this study used qualitative methods to study two complex biomedical settings. The results described in Chapter 5 and 6 will show how the use of qualitative methods helped to develop a framework to better support multi-institutional biomedical collaboration.

4.3 Study Steps

The section 4.1 and 4.2 introduced the theories and methods that were used in this study. In this section 4.3, the steps used to conduct the actual study described in Chapters 5 and 6 are described in detail.

4.3.1 Human Subjects

All steps of this research, recruitment, observation, interviews, and artifact examination were approved by the Human Subjects Committee of the University of Washington Institutional Review Board (IRB) to safeguard the rights and welfare of human subjects.

4.3.2 Sites

Two sites (referred to as GH and TH) were chosen as sample multi-institutional interdisciplinary biomedical collaborations (MIBC) to study. The criteria for choosing sites were that they involve two or more institutes remotely located collaborating together in biomedical research and involve participants from at least three different fields.

4.3.3 Recruitment

Recruitment emails (Appendix A) were first sent to a key participant at each site who facilitated introductions. The rest of the participants were recruited through "snowball" sampling (Patton 2002) where each recruited participant recommended subsequent participants to participate in the study by forwarding the recruitment email. Initial informational interviews were scheduled with those participants who agreed to participate. Each interviewee was given a required consent form (Appendix B) and a brief explanation of the study prior to the actual interview process. Observation recruitment was also done through a key participant at each site who helped to obtain permission to sit in at various local and phone conference meetings. Before each observation, a brief explanation of the study was given, but the IRB did not require the consent of each individual present during the observation.

4.3.4 Interviews

At each site, two different sets of interviews were conducted: a) informational interviews, and b) in-depth interviews. The informational interviews were conducted to gather overall background information of the setting to focus and guide the in-depth interviews and to build rapport with the participants at each site. The informational interviews were on average 15-25 minutes in length and semi-structured in nature (see Appendix C for questions structuring the interviews). The questions were guided by the preliminary framework developed in Chapter 3 and were initially tested with the key participant at each site.

Ten informational interviews were done at site GH and eight informational interviews were done at site TH. A subset of informational interviewees was contacted to participate in the second in-depth interview. The in-depth interviews probed deeper into the nature of the collaboration and were informed by the results of the informational interviews. Each in-depth interview was on average 40-60 minutes in length and semi-structured in nature (see Appendix D for questions structuring the interviews). The questions were guided by the analysis of informational interviews (see section 4.3.8) and by the preliminary framework developed in Chapter 3. Ten in-depth interviews were carried out at site GH and six in-depth interviews were carried out at site TH. Both informational and in-depth interviews were recorded using a digital recorder (iPod with microphone extension) and transcribed to be analyzed later. The interviews reached a saturation point (i.e. majority of themes arising from each interview that is similar in nature) after this point.

4.3.5 Observations

Observations of activities relevant to MIBC were performed at each site. Five observations were carried out at GH, each averaging 50-60 minutes in length. Four of those observations were observations of multi-site conference call meeting and one was of a local chapter meeting. Two observations of average 60-80 minute length phone conference meetings were carried out at TH. At each observation, a set of questions guided field notes (Appendix E). Each set of observation notes were taken as described in section 4.2.1 and transcribed to be analyzed later. Note taking was guided by the preliminary framework developed in Chapter 3.

4.3.6 Artifact Examination

At each site, existing documents associated collaborative activities were examined for their content. A total of 46 different documents were studied at GH. These documents included: a) 5 journal articles describing the structure and infrastructure related to GH, b) 3 collaborative process guidelines, c) 8 sets of presentation slides on collaboration at GH, d) 26 meeting minutes and agendas associated with collaborative activities at GH, e) 2 internal gazettes associated with collaboration at GH, and f) 3 documents describing collaborative infrastructure. At TH, three sets of meeting minutes and agendas associated with collaborative activities were examined. Careful notes were taken during the examination of each document as described in section 4.2.1 to be used in content analysis later. The artifact note taking was guided by the preliminary framework developed in Chapter 3.

4.3.7 Triangulation

To guard against research bias, data and methodological triangulation was used in this study as described in section 4.2.2. Multiple sites (2 field sites, see section 4.3.2) were used as well as multiple sources of data (i.e. documents, interview transcript).

Multiple methods were also used to study each field sites (i.e. interviewing, observation, artifact examination). These methods were described in detail in section 4.2.

4.3.8 Analysis

The analysis of this research generally followed the common qualitative analysis steps described above in Section 4.2.3. The informational interviews from both GH and TH were transcribed after the first round of interviews. The potential themes (codes) related to MIBC were initially identified on paper transcripts with open coding. The initial open coding was guided by the preliminary framework developed in Chapter 3. Themes related to preliminary framework as well as new emerging themes were identified through the open coding process. The transcripts were then gathered in ATLAS.ti analysis software (Atlas Ti) (Figure 4.1) and initial codes were repeatedly refined through multiple readings of each transcript. The generated codes were iteratively grouped into categories which then were further grouped into higher categories. These initial codes and categories were then used to guide the in-depth interview questions.

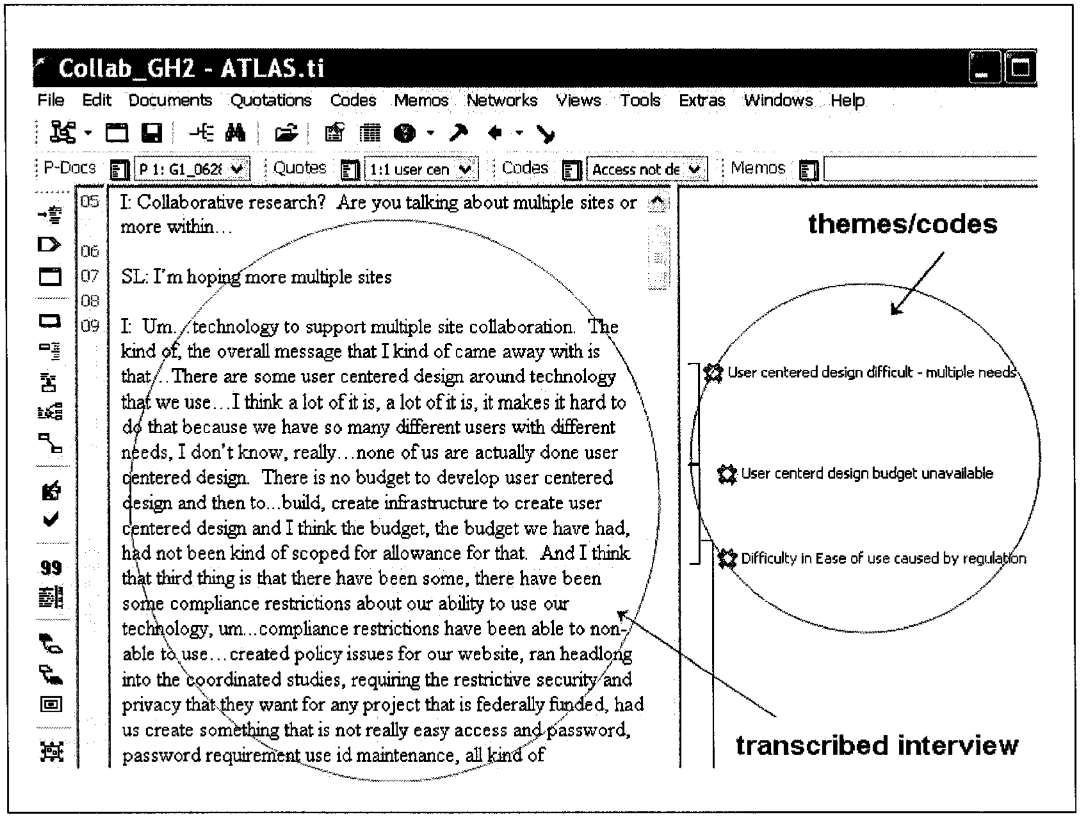


Figure 4.1
Screenshot of Atlas.TI interface. It shows the transcript of an interview to the left and codes/themes associated to the right.

The preliminary analysis of informational interviews to guide more in-depth interviews revealed both procedural and social aspects of collaborations (Appendix F). The procedural nature of collaboration was revealed by participants talking about different processes related to collaboration such as initiation of collaboration vs. the research phase of collaboration. Each process had associated factors that facilitated (facilitators) and impeded (barriers) the process. Aside from processes, there were also many social and socio-technical issues associated with the overall collaborative process. These issues crossed procedural boundaries and affected the collaboration as whole. Most of the factors that contributed positively or negatively to collaboration were social in nature (Appendix F). Social factors such as trust, existing relationships, and similar interest and vision helped participants overcome some of the difficulties of MIBCs and helped participants continue their collaborative research endeavor (Appendix F.1). Social network was used primarily to find collaborators. Incentives such as access to more resources and expertise led the participants to be involved in the MIBC (Appendix F.2). The barriers to collaboration were associated with distance and differences in context. The differences in local culture, institutional practices and funding mandates led to contextual differences. The distance created difficulties for participants to find other collaborators or resources with MIBCs. Most of the concepts that emerged as factors that facilitated the various processes associated MIBCs were technical in nature. MIBCs involved a large array of tools such as email, data repository and online resources that helped bridge the distance among collaborating institutions.

Guided by initial themes, the in-depth interview questions were formed to tease out the process boundaries (Appendix D) and get a more in-depth view of social issues. The process was roughly grouped as pre-, during, and post-collaboration. For each of the processes, the participants were asked to describe both technical and non-technical attributes that they found helpful and aspects of the process that hindered them. They were also asked if they had any needs associated with the collaboration. The questions about some of the social issues associated with collaboration such as trust were interspersed throughout the interviews.

After all the data were collected, the transcriptions of all the interviews, observational notes, and artifact notes were gathered and any themes relevant to MIBC were initially identified on paper transcripts. Any concept that was related to preliminary framework or to the processes of MIBCs was identified as a relevant theme. The data transcripts were then gathered in ATLAS.ti analysis software and the codes/themes arising from observational and artifact transcripts were iteratively refined. Three different sources of data were used to determine whether a certain theme should be included into a final code set. Themes that were mentioned in less than three separate sources of data were excluded from the final code set. Each interview participant was counted as a separate source of data. The resulting codes/themes and categories from all data sources were combined afterwards (Appendix G). These themes are described in detail in Chapter 5 and Chapter 6.

The final codes (see individual codes in Appendix G) were loosely categorized into different groups based on their naturally emerging similarities and preliminary framework. For example, themes face-to-face meetings, phone, and email were grouped into a communication process because most participants mentioned them together as methods of communication during the interviews and preliminary framework also contained communication as a collaborative process. Each group was refined by iterative examination of raw data and preliminary framework.

After few iterations, an initial version of framework (Appendix H.1) was built by classifying different groups of codes along the collaborative project timeline. The social and socio-technical issues were shown to influence the entire collaborative project cycle. In the initial framework, many processes were classified as pre- and during collaboration and no processes were really classified into post-collaboration. After several more iterations, the second version of the framework (Appendix H.2) removed the post-project timeline and grouped all the process categories into pre and during collaborative research with surrounding social and socio-technical issues. Although the second version of the framework better characterized MIBCs, many of the processes didn't fit wholly into the process timeline (i.e. just into pre- or during collaborative process). For example, the communication process happened most during collaboration, but also had to happen pre- and post- collaborative project. After many more iterations, the third version of the framework (Appendix H.3) discarded the idea of a timeline, but grouped the processes into core processes that needed to occur throughout the collaboration with surrounding social and socio-technical issues. The significance of the biomedical nature of MIBCs was also highlighted as along with external factors such as funding. The final framework was developed after more iterations and is described in detail in Chapter 5 Section 2 (general overview), Chapter 6 (in-depth description of core processes), and Chapter 7, Section 1.1 (framework).

4.3.9 Inter-coder Check

The resulting preliminary framework of MIBC (Appendix H.3), and associated categories and codes (Appendix G) along with a random subset of transcribed interviews were given to two graduate students with experience in qualitative analysis. Each student received three transcripts from site GH and two from site TH. Only one identical GH transcript was given to both students. During an initial meeting, the preliminary framework, categories and codes were explained. Then the graduate students were asked to code the transcripts using the given codes and categories and note the following: a) any new emerging themes related to domain of multi-institutional collaboration, b) unclear codes and categories, and c) any other issues of note.

In subsequent meetings, the three coders met to discuss the codes. The GH transcript coded by all three coders was first discussed in detail. Each code associated with transcript passages was compared and discussed. When conflicts arose, they were either resolved through group discussion or the code in question was refined. The rest of the GH and TH transcripts were discussed individually with the coders and the same method of resolving conflicts was applied. The overall analysis approach was refined according to the discussions and led to the final results described in detail in Chapters 5 and 6.

4.3.10 Trustworthiness

This research rigorously followed the steps recommended by Lincoln and Guba (Lincoln 1985) as described in Section 4.2.4 to ensure trustworthiness of the study. The steps implemented were:

a. To ensure credibility:

- All interview questions were pre-tested.
- Data were gathered through prolonged engagement in the field.
- The observer engaged in rigorous note taking.
- The research used triangulation of methods (observations, interviews, artifact examination)
- The interviewer used member checking during interview process.

b. To ensure transferability:
   - "Thick description" was provided of methods and theory used in the research so that findings can be compared with other research

c. To ensure dependability:
   - An "audit trail" was used. The "audit trail" consisted of the researcher keeping scrupulous observation field notes and constantly comparing them with other types of collected data for consistency. Furthermore, to ensure accuracy of interview data, interviews were audio-recorded and inter-coder checks of the analysis were conducted as described in Section 4.3.7.

d. To ensure confirmability:
   - The researcher reflected throughout the research period on the research process, method and data.

4.4. Summary

This chapter describes theories and methods used to approach the in-depth study of MIBCs as well as the analysis steps and validity methods used to check the obtained data. As described in Section 4.1 this study has theoretical base in grounded theory and the preliminary framework described in Chapter 3. Qualitative methods such as interviews, observations, and artifact examination were used to collect data. To increase trustworthiness of the study, well known methods such as triangulation and inter-coder analysis were used (for full list please see Section 4.3.8). This study is one of the first studies to broadly use qualitative methods to characterize large complex collaborative settings in the biomedical field. The field study described in this chapter validates the Preliminary Framework developed in Chapter in the context. This is different from TORC in that TORC is based on survey results while this study builds theory from the field data. The result of this study illustrates that the use of qualitative methods is a valid and practical approach to inform system design and development of theory to describe a complex environment. Chapters 5 and 6 will give a detailed account of the result of the study.

Chapter 5: Characteristics of Biomedical Research Collaborations

The first step to addressing the question posed in Chapter 1, "How can biomedical informatics best support current multi-institutional, interdisciplinary biomedical collaboration?" is to gain a thorough understanding of what Multi-institutional Interdisciplinary Biomedical Collaboration (MIBC) is (see Chapter 1, Section 3). It would be impossible to support something without knowing what is being supported. As described in Chapters 2 and 3, collaboration in general has been studied extensively. Yet little research exists that describes specifically biomedical collaboration and still less research exists that describes MIBC and none that look at it from a biomedical informatics perspective. The reason for this gap in research (mentioned in Chapter 1 Section 2) is partly due to the fact that the large effort invested in forming MIBCs has been a relatively recent event. The effort has been largely driven by funding agencies that have recognized MIBCs as a solution to solving complex research questions (NIH roadmap).

This study bridges the gap in knowledge of MIBCs by describing some of the fundamental characteristics of MIBC that emerged from the contextual field study described in Chapter 4. Chapter 5 starts with the descriptions of two MIBCs under study and their members, and the similarities and differences between two sites are described in Section 5.1. Then in Section 5.2.1, all the core activities and processes that are associated with MIBCs are introduced. Although not part of the core collaborative activities, social and technical issues that influence overall collaboration are discussed in Section 5.2.2. The detailed description of methods (observations, interviews, and artifact examination) that were used to conduct the contextual study of two MIBCs is in Chapter 4 as well as the stages of development of the framework (see also Appendix F, G, and H).

5.1 Context

5.1.1 MIBC 1: GH

GH Structure

GH is a non-academic research institution that is a member of a large loosely federated Research Network (Figure 1) of institutions where membership is voluntary. An institution is the smallest unit within the Research Network that contains its own separate IRB, a defined location, a single goal and vision, and a set of researchers belonging to that unit. The Research Network has a governing board that only makes membership decisions and has no formal funding mechanism or legal status tied to it; however, the member institutions have a common unwritten code of conduct. The Research Network was formed for the purpose of attaching a unified identity to existing collaborative efforts and to gain a better competitive edge for funding. Even before the formation of a formal Research Network, many multi-institutional collaborative projects already existed among the member institutions of the Research Network.

Various subsets of institutions within the Research Network are involved in a number of multi-institutional collaborative projects funded by various funding agencies; however, only one institution within each multi-institutional collaborative project is the lead institution for that project. Multi-institutional collaborative projects have goals and set agendas which are overseen by a committee composed of a subset of investigators from member institutions. Each of the multi-institutional collaborative projects also has multiple sub-collaborative projects associated with it. Some of the multi-institutional collaborative projects work together closely while others do not.

Currently, GH is part of several multi-institutional collaborative projects with other institutions within the Research Network and outside. These projects are all tied to different funding mechanisms. Each of these multi-institutional collaborative projects has a different research goals and infrastructure due to their separate funding. Because an institution is a single contained unit within the Research Network, researchers at GH must first get approval from GH before participating in any collaborative project even within the Research Network.

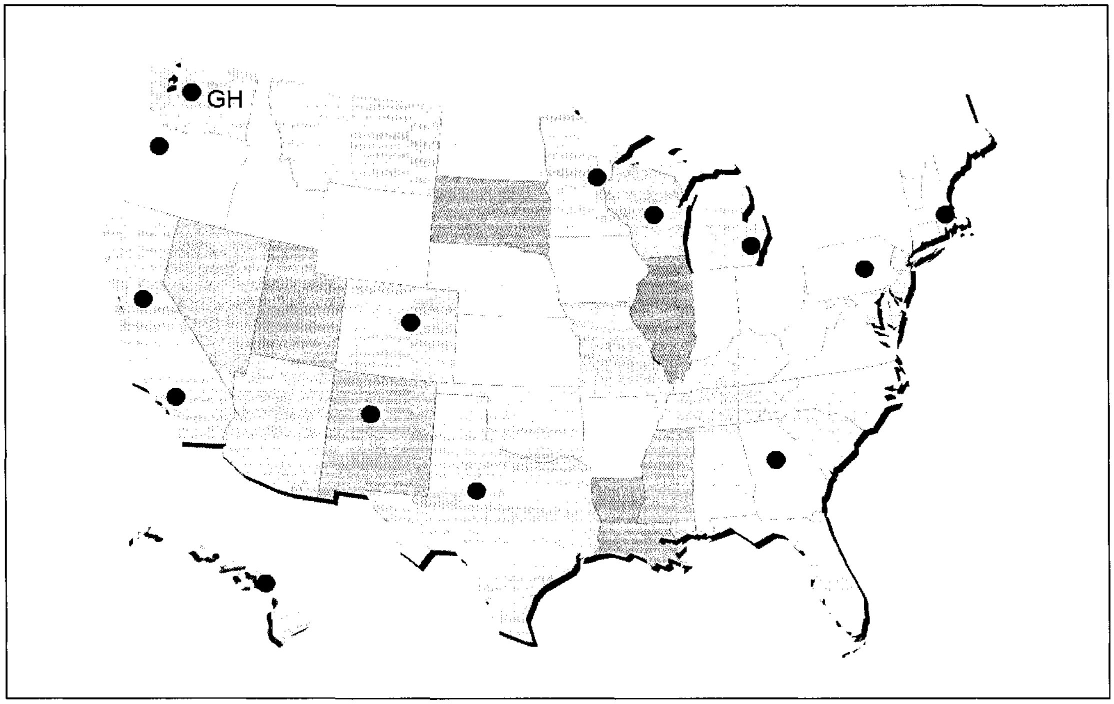


Figure 5.1
GH and the Research Network of institutions. Each of the dots represents an institution involved in the Research Network that is collaborating together.

GH Participants

The participants at GH were comprised of investigators, researchers, project managers, and IT specialists (see Table 5.1 for details). The majority, especially investigators, had backgrounds in health-related fields or in medicine and had post-graduate degrees. The majority of the participants have also worked at GH for many years although not necessarily within the same collaborative projects. Many of the participants identified themselves as health researchers and also identified with the goals of GH. Most participants were involved in multiple collaborative projects associated with the Research Network and had affiliated appointments outside GH.

Table 5.1
Descriptions of GH investigators whose interview data was used in this Chapter.

Investigator	Description
G1	G1 is an investigator who's been with GH for over 3 years. Because he is a medical doctor, not a PhD, he believes brings different perspective to GH research. He is involved in several collaborative projects at GH. One of his primary research interests is in facilitating collaboration within the greater GH network from the infrastructure stand point. He calls himself a "facilitator".
G2	G2 is an investigator who has been with GH for over 10 years. He became involved with GH due to a research assistant position he had during his PhD. He is involved in several collaborative projects at GH. His primary research interest is in health economics.
G3	G3 is an investigator who has been with GH for over 6 years. He is a medical doctor who has training in internal medicine and has a Master's degree in Public Health. He had been a faculty member and a researcher at a university and has worked with GH for about 20 years. He is currently involved in several collaborative projects at GH.
G4	G4 is an assistant scientific investigator who was a trained pharmacist for over 8 years. After getting her Master's and PhD, she came to work for GH. She is involved in several collaborative projects and has primary interest in pharmaco-based research at GH. She also teaches and mentors graduate students part time.
G5	G5 has worked as a project manager for over 4 years and recently changed her role to become a research scientist. She has a Master's in Public Health and works on several collaborative projects at GH. Due to her past role as a project manager, liaison and coordinator, her current research interest is in better facilitating multi-site research.
G6	G6 became involved with GH in the 90s and worked as a project manager from 1997 until 2001. After a 2- year break, she came back to work as a project manager for several multi-site collaborative projects.
G7	G7 is a project manager for several multi-site projects and has worked for GH for over a year. She has a Master's in Public Health and had managed clinical research project after her graduate work. She believes that a position of project manager for multi-site research needs a lot of experience in many areas such as having prior experience in multi-site research and in aspects of business management.
G8	G8 has woked at GH for over 11 years, first as an assistant investigator and then an investigator. She has her doctoral training in breast cancer and works in a similar field at GH. She plays various roles in several multi-site collaborative projects at GH.
G11	G11 is a programmer who has been working at GH for over 20 years. He is involved in many collaborative projects. He observes that over the years the circle of what GH considered "in-group", insider and therefore can be trusted, has grown significantly to include other institutions and universities.

5.1.2 MIBC 2: TH

TH Structure

TH is an academic collaboration that evolved from a single collaborative project at an institution in Missouri. TH gradually became a multi-institutional collaboration as the original members of the team moved to different locations. Now the participants of TH are located at 4 different states and at an international site. TH has several different collaborative projects that were initiated by different investigators within TH.

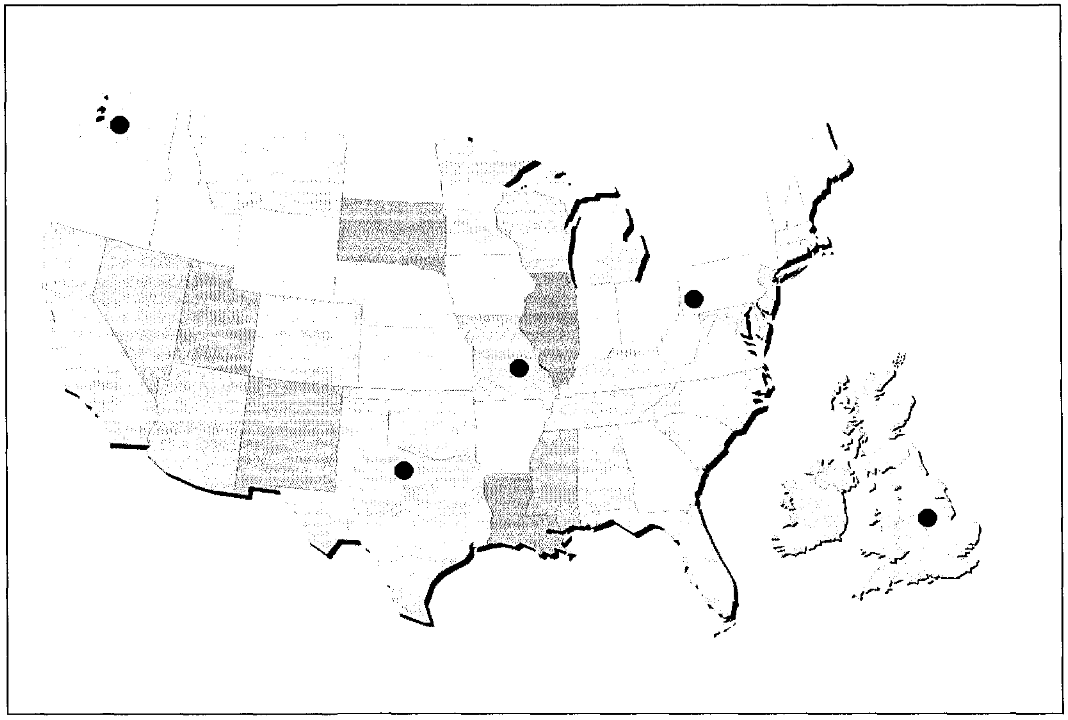


Figure 5.2
Each of the dots represents an institution involved in TH (4 institutions located in the United States and 1 institution in England).

TH Participants

TH has total of 11 members composed of investigators, researchers, and research assistants (See Table 5.2 for details). The participants at TH have various backgrounds: nursing, social work, informatics, and communications. Every participant at TH is working on multiple projects, but within the same area of research and funding. Each investigator has a different role depending on the project.

Table 5.2
Description of TH investigators whose interview data was directly quoted in this chapter.

Investigator	Description
T1	T1 is an investigator and one of the founding members of the TH collaboration which started in 2002. Since 2007, he has relocated to a university in WA and works remotely from the main TH site. His field of interest is in using information technology to support patient comfort. He is involved in various collaborative projects within TH.
T2	T2 is a principal investigator and formerly a graduate student associate of TH. She has a PhD in nursing and informatics and was brought into group by T1 through mentorship. T2 has recently moved to another university and works remotely from the main group. She is involved in various collaborative projects within TH.
T3	T3 is an investigator and has been involved with TH for over 3 years. Her primary field of interest is in communication. She was brought into the group by T4 due to her expertise. She has, from the beginning of the collaboration, worked remotely from the main group.
T4	T4 is an investigator and one of the founding members of the TH collaboration. She started collaborating with T1 because their work complemented and they had similar research goals. For the last five years, she and T1 have recruited many other members to the TH collaboration. She has a PhD in sociology and is involved in various collaborative projects within TH.
T6	T6 is a researcher who collaborates with TH outside the U.S. Initially she became locally involved with the group due to her nursing background and common research interest with T1 and T4. She has had to return to the UK, but has maintained her role as part of the TH group. She is currently involved in several collaborative projects at TH.
T7	T7 is a research assistant in TH. She has an undergraduate degree in biology and a Master's in Social Work. She claims to be a "social worker". She was in the social work field for 5 years before deciding to come back to school and by chance, through T4, got involved with TH.

5.1.3 GH vs. TH: Similarities and Differences

One of the major differences between GH and TH is that TH is an academic collaboration while GH is a non-academic collaboration. Being an academic collaboration that originally started in a single institution, TH is much smaller in size, effectively only one collaborative team functioning as an umbrella over several collaborative projects. On the other hand, GH is an institution involved in multiple large collaborations with sub-collaborative projects that occur among other institutions. Nevertheless, although GH and TH are different in size and in nature of collaboration, there are also many similarities. Both GH and TH are multi-institutional interdisciplinary collaborations, involving investigators from various fields. Even though one is non-academic and the other academic, both GH and TH largely depended on federal funding for their research. Furthermore, both conduct research in the biomedical field. Due to similarities in funding and the common research area, a lot of commonalities in terms of research practices were discovered.

5.2 Internal Characteristics of MIBC

Many activities, processes, and issues that are integral to multi-institutional interdisciplinary biomedical collaborations (MIBCs) emerged from the analysis of contextual field data. In this section, these characteristics are categorized into two large groups: a) core collaborative activities and processes: factors that are integral to daily operation and maintenance of collaboration and its projects, and b) social and socio-technical issues surrounding overall collaboration: factors that influence overall collaboration although are not daily processes required to keep collaborative projects going (Figure 5.3).

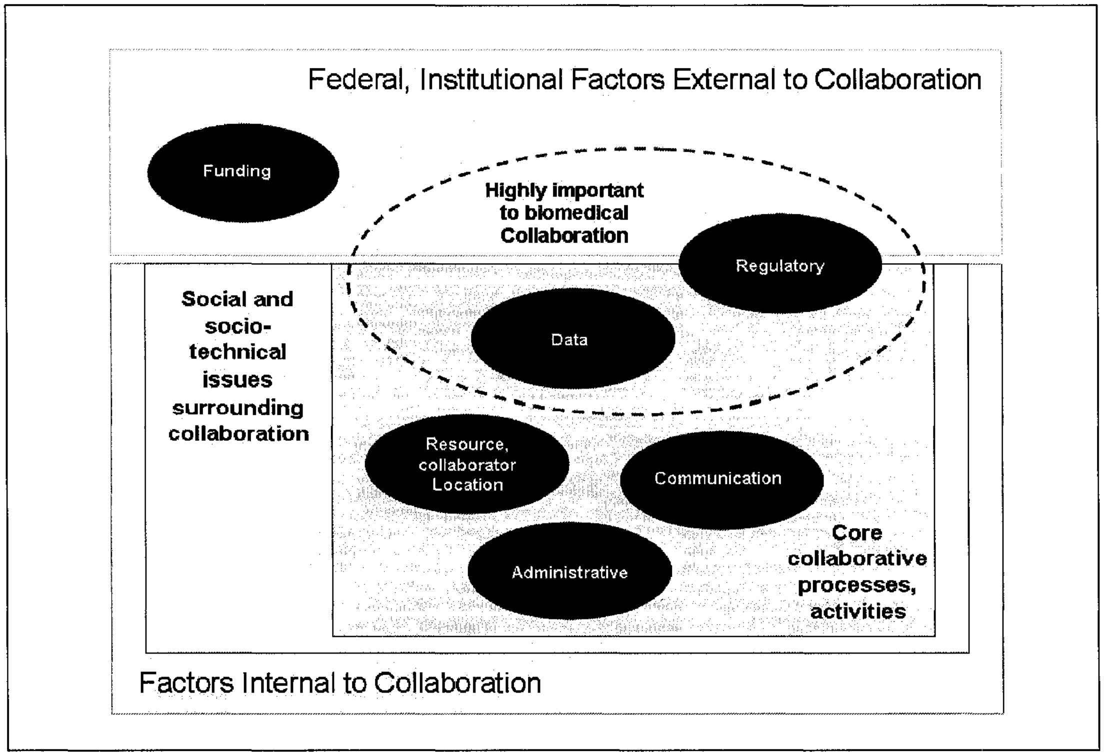


Figure 5.3
The characteristics of multi-institutional interdisciplinary biomedical collaboration. The yellow squares represent top most categories (internal and external factors). The blue square represents social and socio-technical factors that affect overall collaboration. The pink square represents the core collaborative processes and activities with each process represented by a red oval.

5.2.1 Core Collaborative Processes and Activities

Core collaborative processes and activities are those that are integral from the beginning to the end of the lifecycle of collaboration. These activities and processes can be broken down into 5 large groups: a) data, b) regulatory, c) communication, d) administrative, and e) resources/collaborator location (see Figure 3, inside pink rectangle). The factors related to data and regulatory processes are of particular importance to biomedical collaborations due to the unique nature of biomedical research having to deal with sensitive data. These processes have not previously been stressed as of great importance in collaborative studies done in other fields. Biomedical collaborations nevertheless have characteristics that are similar to other collaborations that have been studied extensively. Despite differences in focus and the nature of research, biomedical collaborations are still all collaborations among researchers in the pursuit of knowledge. Hence, many common characteristics of collaboration exist across research fields. The activities and processes introduced in this section will be further broken down and discussed in detail in Chapter 6.

5.2.1.1 Characteristics Particularly Important to Biomedical Collaboration

Among these categories of core collaborative processes and activities, data and regulatory-related activities are of particular importance to biomedical collaborations. These two activities are tightly coupled together since biomedical collaborations often deal with health-related data that is highly regulated. Due to their sensitive nature, the health-related data are highly regulated at every level. Both MIBCs were keenly aware of this and had processes in place to ensure data protection. Since data are highly important to biomedical collaboration and essential to being able to answer research questions, there is no collaborative project that does not have some need to share data. The importance of data processes to biomedical collaboration was noted in the preliminary framework described in Chapter 3, Section 2 and Table 3.1. Although not specifically mentioned, regulatory processes are also related to security structures (described in Chapter 3, Section 2 as a factor also particularly important to biomedical collaboration) since they both deal with the protection of sensitive data. These themes that emerged during the development of the preliminary framework emerged again during the field study described in Chapter 4, Section 3.

The data-related and regulatory-related processes presented in this section were not mentioned in TORC, the general theory of collaboration. Since data important to biomedical research, the context-specific Preliminary Framework developed in Chapter 3 included data-related processes as a part of the framework. However, field study not only discovered data-related processes, but regulatory-related processes as an important part of biomedical collaboration.

Data-related Processes

The data-related processes and activities occur before, during, and after a collaborative project. Before starting a new collaborative project, researchers often have to analyze preliminary data to better formulate research questions. Throughout the collaborative project, data are generated, analyzed, and shared among collaborators. Even after a project is completed, data are often revisited for further analysis or used for follow up project ideas. When asked to describe some of the daily collaborative activities, the participants at both MIBCs mentioned several data related processes.

One of the most notable data-related processes was related to security. The participants at both TH and GH were well aware of the importance of keeping data private and secure. Investigator T1 at TH mentioned that even during private phone conferences, his group has a policy never to discuss any research subject by name, thereby protecting the anonymity of their subjects:

> *"We have policy anyway not to discuss patient by their names. So you make hear things like subject number blah blah blah, but we don't use patient identifying info during these calls..."*

Investigator G2 readily described security protocols they have at GH to protect data during transfer. Even though data was already de-identified, GH took further care to transfer them securely:

> *"And then when the data is transferred, what we're transferring is the de-identified, anonymized data through secure web transfers so..."*

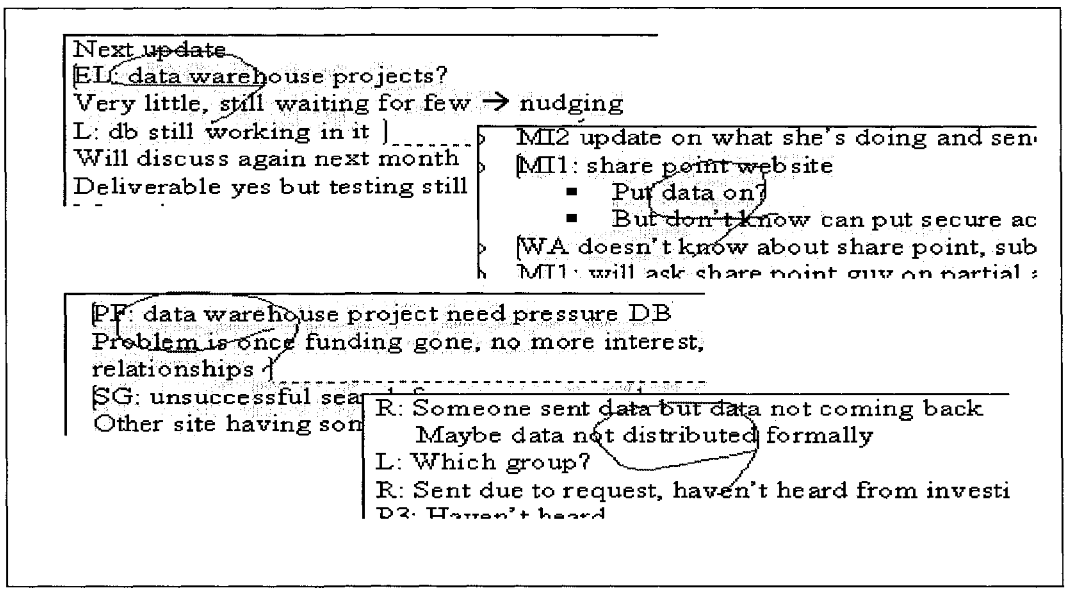
When asked to describe tools that are used during data-related process, investigator G1 described a custom-built data warehouse technology that exists at GH to facilitate data sharing:

> *"So if we get the grant, which is likely, will use combination of virtual data warehouse, develop some of our own variables and we will ship those through, back and forth, de-identified, through secure file transfer protocols and...that's it."*

Although investigators at TH did not have a sophisticated technology such as a custom-built data warehouse, they still had a tool to store data. Investigator T4 at TH mentioned an off-the-shelf technology, Sharepoint that they utilize as a data storage tool:

> *"Uh, we consistently use, a joint website where we store data, a sharepoint website, so we can all have access to that very different documents, data, etc."*

During observations of both local GH and national GH meetings, data-related processes such as data sharing and tools that aid the process such as warehouses, were often mentioned (see examples of observation notes in Figure 5.4). The artifacts examined also revealed many data processes that exist within GH such as data collection, data definition, and data aggregation (see examples of artifacts in Figure 5.5).



Next ~~update~~
EL: data warehouse projects?
Very little, still waiting for few → nudging
L: db still working in it
Will discuss again next month
Deliverable yes but testing still

MI2 update on what she's doing and sen
MI1: share ~~point~~ ~~web~~site
- Put data on?
- But ~~don't~~ ~~know~~ can put secure ac

WA doesn't know about share point, sub

PF: data warehouse project need pressure DB
Problem ~~is~~ ~~once~~ funding gone, no more interest, relationships
SG: unsuccessful sea
Other site having son

R: Someone sent ~~data~~ ~~but~~ ~~data~~ not coming back
Maybe data not distributed formally
L: Which group?
R: Sent due to request, haven't heard from investi

Figure 5.4
Sample observation notes showing data related processes.

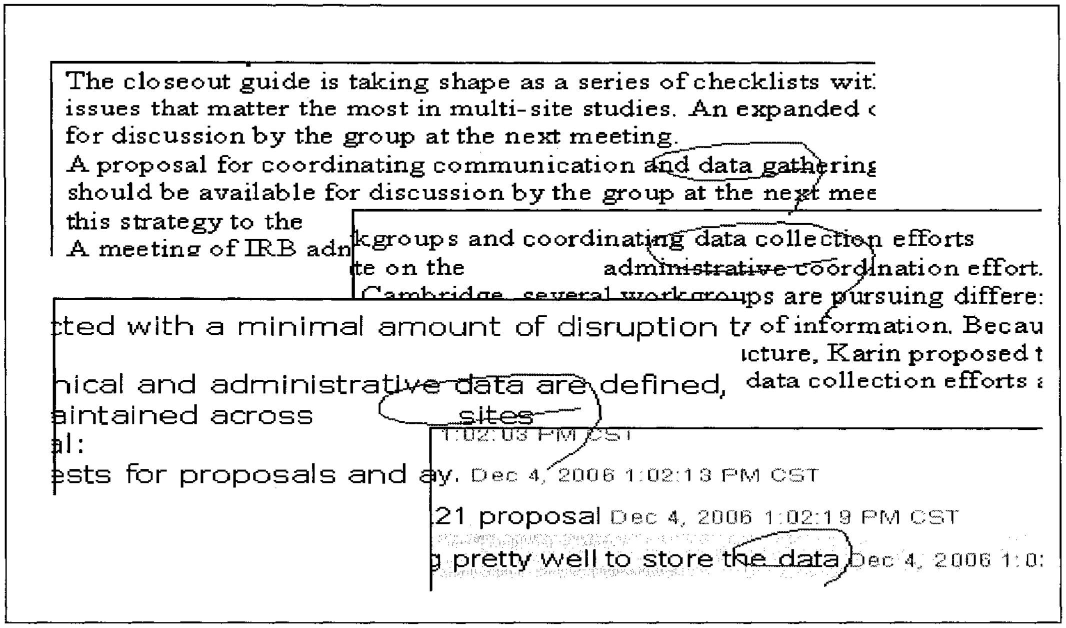


Figure 5.5
Sample artifacts displaying data related processes.

The field study revealed many data related processes such as data security protocols and data sharing. Also associated with the processes were tools that facilitate the processes such as data warehouses and data transfer tools. The data-related process is one of the five main processes that occur during MIBC and will be further analyzed and described in detail in Chapter 6 Section 1.

Regulatory-related Processes

The regulatory process must be in place at every level of collaborative research to meet federal and institutional standards. Before any project begins, an institutional regulatory board must ensure that the research about to be conducted is done properly. Through the research process, the process approved by the institutional review board (IRB) must be closely monitored and followed. Although not specifically asked about, the most frequently mentioned processes that exist within MIBC are regulatory-related processes, especially those that involve institutional review boards (IRBs).

In regards to regulatory process, investigator G3 at GH described the IRB process and the security consciousness that exist within GH:

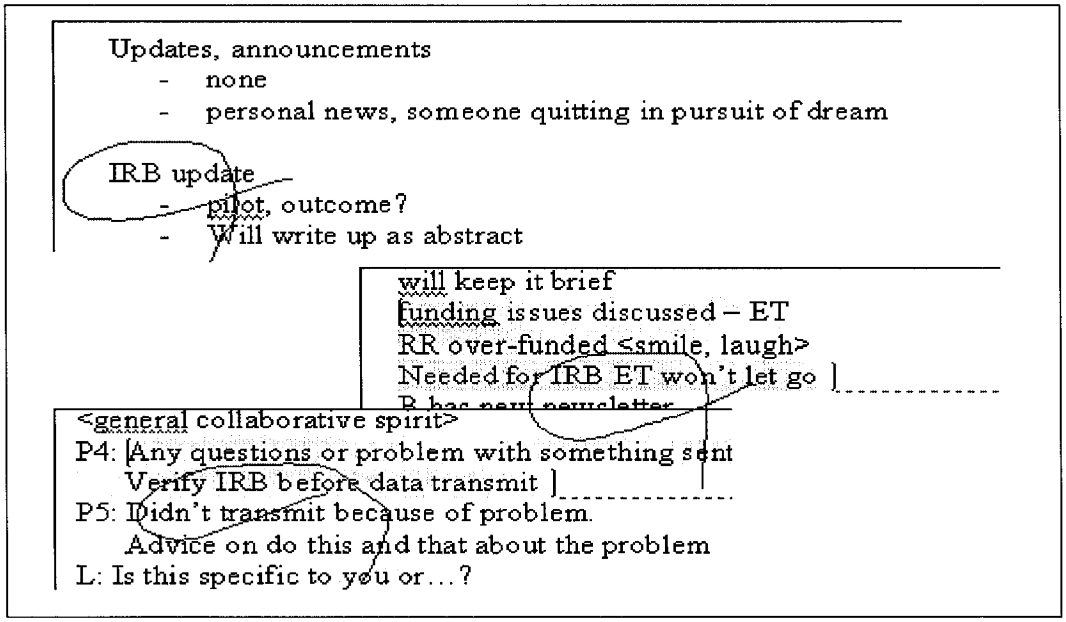
> *"We have IRBs and the...just very high level of security consciousness here."*

The investigator T1 at TH also similarly described the regulatory process at TH in association with IRB :

> *"...we never could control who come in and out of the room. IRB is signed, but we will not be using identifying information so..."*

Although they are members of two separate collaborations, the importance of the IRB process to each collaboration can easily be understood from their descriptions.

Regulatory-related processes, such as IRB, were also frequently mentioned during observation sessions (see sample observation notes in Figure 5.6). Since the institutional review board is involved in every step of the research process, it is not surprising that each observation session of a meeting had at least one mention of the IRB process. The artifacts examined also revealed a similar trend in regards to references to IRB processes (see sample artifacts in Figure 5.7). Examined artifacts of various kinds (i.e. meeting notes, agendas, guidelines) describe various aspects of the IRB process.

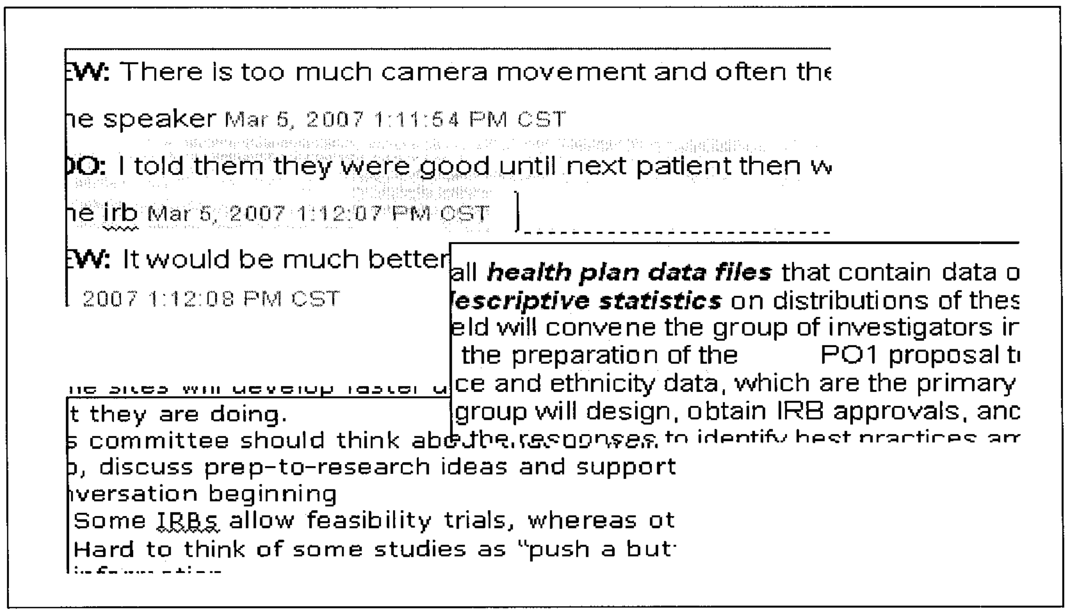

Updates, announcements
- none
- personal news, someone quitting in pursuit of dream

IRB update
- pilot, outcome?
- Will write up as abstract

will keep it brief
funding issues discussed – ET
RR over-funded <smile, laugh>
Needed for IRB ET won't let go]

<general collaborative spirit>
P4: [Any questions or problem with something sent
Verify IRB before data transmit]
P5: Didn't transmit because of problem.
Advice on do this and that about the problem
L: Is this specific to you or…?

Figure 5.6
Sample observation notes with references to regulatory processes.


W: There is too much camera movement and often the
e speaker Mar 5, 2007 1:11:54 PM CST
O: I told them they were good until next patient then w
e irb Mar 5, 2007 1:12:07 PM CST]
W: It would be much better
2007 1:12:08 PM CST

all **_health plan data files_** that contain data o
escriptive statistics on distributions of thes
eld will convene the group of investigators ir
the preparation of the PO1 proposal t
ce and ethnicity data, which are the primary
group will design, obtain IRB approvals, and
the responses to identify best practices an

t they are doing.
committee should think abo
, discuss prep-to-research ideas and support
versation beginning
Some IRBs allow feasibility trials, whereas ot
Hard to think of some studies as "push a but

Figure 5.7
Sample artifacts with references to regulatory processes.

The results of the field study indicate the importance of regulatory processes to the collaborative research process. As one of the five main processes occurring in MIBC, regulatory process must be studied and supported. The regulatory process will be further analyzed and described in detail in Chapter 6 Section 2.

5.2.1.2 Characteristics Important to Any Collaboration

Of all the core collaborative activities and processes that emerged from the field data, several categories can be seen as important to any collaborative effort. They are: a) finding collaborators/resources, b) communication, and c) administrative processes that support the collaborative project. These categories of processes have been studied extensively in fields ourside biomedicine (Finholt 2003; Olson 2005), but not specifically by the biomedical informatics community. The preliminary framework in Chapter 3, Section 2.1 described general concepts associated with any collaboration. Although organized and grouped differently, these themes were established and confirmed during the field study described in Chapter 4, Section 3. The field study revealed that communication (described as the nature of remote communication and communication technology in the preliminary framework) and administrative processes (described as common workspace and coordination tools to bridge distances) occur regularly in both MIBCs. The field study, however, revealed that data processes are tightly connected and are best grouped as one entity, instead of separate pieces as described in the preliminary framework (i.e. in section 5.2.1.1. described the following data processes: data sharing and management, data integration and analysis, data standards, and data provenance).

Communication

Communication occurs throughout the lifecycle of collaborative projects. Communication can include anything from discussing initial project ideas to regular discussions of research progress during the project period. Essentially, without communication, no collaborative project can start or proceed. Many mechanisms of communication exist throughout collaboration. The participants of both of MIBCs mentioned three general categories of communication mechanisms: a) asynchronous communication such as email, b) synchronous communication via technology such as phone, and c) synchronous communication without medium such as face-to-face meetings.

When asked what are some mechanisms used to communicate during collaboration, investigator G1 at GH said the two most frequently used methods were email and phone:

> *"We have telephone, cell phone, word documents and emails..."*

Investigator T2 at TH had a similar response to that same question:

> *"...we do really low tech things like telephone, um, for conference calls..."*

Although phone and email are used most frequently, communication without tools such as face-to-face meetings is also often mentioned. Investigator G4 at GH described frequent face-to-face meetings that occur at GH:

> *"The research network meets once a year, and we often have group meetings, we do have group meetings..."*

Researcher T4 at TH describes the frequency of face-to-face meetings at TH:

> *"...we meet at least four times a year face to face, um...at conferences or at one site or the other."*

The existence of email and phone communication were also seen during observations (see sample observation notes in Figure 5.8) and in artifacts examined (see sample artifacts in Figure 5.9).

<Call from MI>
On cell phone because the other phone doesn't work, it will be a

MI: Report on Springfield visit (site visit)
- Recruitment dropping
- Referral problems
- Starting
- Update

PF and SG: but they're little different but they did find som
ET: past experiment with website not good so need focus g
SG: context vs. case, shouldn't decouple
B: not everyone thinks computer first
SG: lot likes just phone
ET: volunteered for running focus group because no one w
None of network site wanted it
Finding people to do hard

Figure 5.8
Sample observation notes displaying references to phone communications.

DO: have a great afternoon May 7, 2007 2:04:53 PM CDT

DO: we will stay in touch via email May 7, 2007 2:05:02 F

GD: i will keep in touch via email for the subcontracts

DO: ok May 7, 2007 2:05:09 PM CDT

Upcoming January progress report:
is in the process of emailing deliverable leads for infor
quarterly progress report. Please look for her email and pr
January 3rd. The report covers the period from October to

Figure 5.9
Sample artifacts displaying references to email communications.

The results of the field study reveal many communication related technologies such as electronic mail and phone. As one of the five main processes occurring in MIBC, the communication processes must be studied and the technology necessary to facilitate communication must be supported. Data processes will be further analyzed and described in detail in Chapter 6, Section 3.

Administrative

Administrative processes, like communication processes, affect the entire lifecycle of the collaborative process. Essentially, these are day to day activities that really tie the entire collaborative project together and keep it moving. Administrative activities involve anything from coordinating schedules to disseminating information. Most commonly mentioned administrative processes are schedule coordination and agenda setting.

Project manager G7 at GH mentioned that GH has numerous meetings involving various committees and they all have regular set schedules to accommodate investigators' schedule:

> *"...we have set schedules for all these committee meetings and they all interrelate, um..."*

Researcher T4 at TH mentioned regularly scheduled meetings at TH:

> *"Regular scheduled meeting times, um, with a scheduled agenda."*

This participant also mentioned that at each meeting there was a set agendas associated with appropriate project milestones:

> *"He'll make an agenda and I will make an agenda and we'll go over both persons' agenda."*

Less frequently mentioned administrative activities include information distribution and the standardization of processes. Throughout the lifecycle of a collaborative project, a great deal of information regarding the project needs to be distributed among the investigators. Project manager G7 at GH mentioned how such information is often distributed through online technologies:

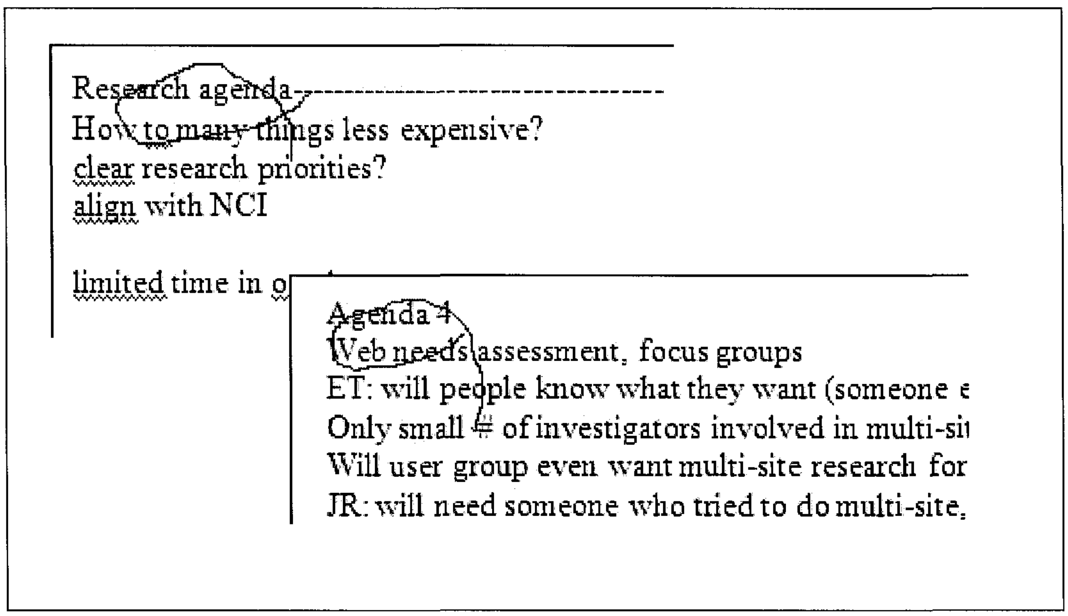
> *"I have a general list, I have a distribution list of people of committee members, site PIs, and I've asked those site PIs have over the time to*

include people that want to know about opportunities, general opportunities…"

Some of the processes involved in collaborative projects are repetitive and needs to be performed often. Project manager G6 at GH mentioned that there are some standardized guidelines in place for such processes:

> *"…trying to come up with best practice…a lot of this is best practices, guidelines…they're not standard operating procedures, so they have to be…"*

Various administrative processes that exist in collaboration were also seen during observations (see sample observation notes in Figure 5.10). The meetings always started with agenda and at a set time. The artifacts also show that the use of agendas, meetings, and calendar schedules are prevalent throughout the collaborative process (see sample artifacts in Figure 5.11).



Research agenda-
How to many things less expensive?
clear research priorities?
align with NCI

limited time in o

Agenda 4
Web needs assessment, focus groups
ET: will people know what they want (someone e
Only small # of investigators involved in multi-sit
Will user group even want multi-site research for
JR: will need someone who tried to do multi-site,

Figure 5.10
Sample observation notes with references to agendas.

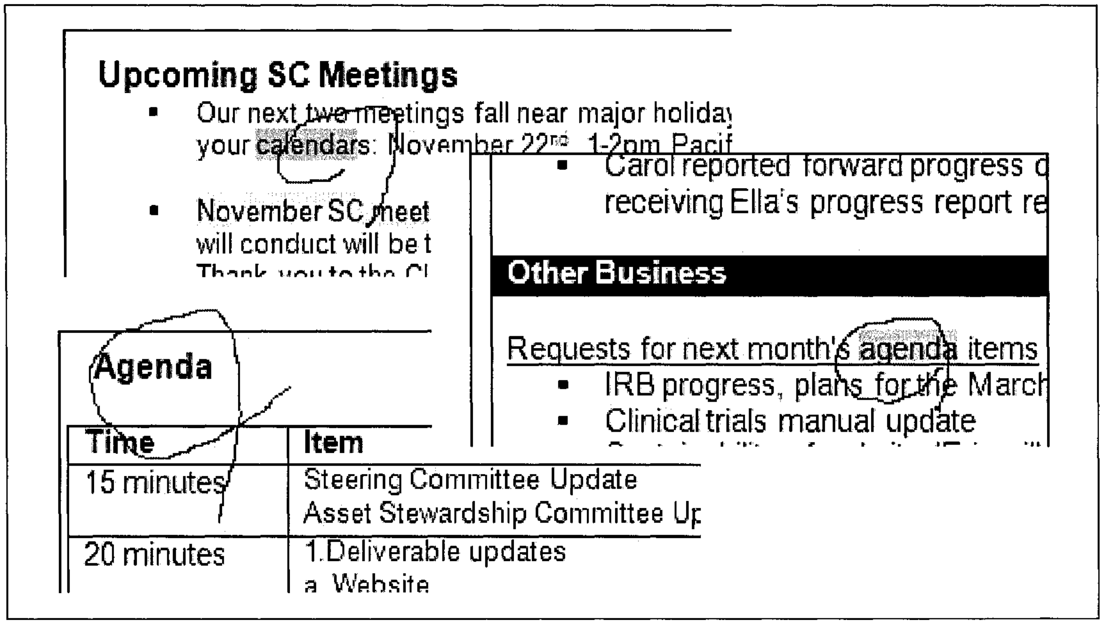


Figure 5.11
Sample artifacts illustrating agendas and scheduling.

The field study brings to our attention many administrative processes such as schedule coordination and information distribution. As one of the five main processes occurring in MIBCs, administrative processes must be recognized as important, supported as part of collaboration and studied in depth. Administrative processes will be further analyzed and described in detail in Chapter 6, Section 4.

Finding Resources and Collaborators

Usually collaborative projects begin with either an interested and idea of a investigator or a funding initiative. However, before any collaborative project can start investigators first need to find resources for the project and other collaborators. The resources are usually data that needs to be obtained for preliminary analysis. Other collaborators are investigators from various fields that need to be contacted to form a collaboration.

When asked what is involved in starting a collaborative project, investigator G2 at GH mentioned the importance of existing relationships in finding collaborators:

"So people who are doing research in these...we all know each other. So if you have an interesting question, you call up somebody and say, hey Tom, you want to work on this thing."

Investigator T4 at TH mentioned that when looking for collaborators he often first asked researchers he knew and trusted:

"I just send email out to everybody and say hey anybody know anybody or ask around at your place to see if we can gather a name..."

Often, investigators began collaborative projects in collaboration with researchers they already knew. If they could not find appropriate investigators they already knew to work with, they then sought referrals.

Existing relationships, however, were not the only means of finding collaborators. Researcher G5 at GH mentioned that online resources are also used to find collaborators:

"He is on our website. He is the contact person, so that's the point of entry."

Investigators also used online resources to find out more about the collaborators they might work with. Research assistant T7 at TH thought it was important to find out more about compatibility of future collaborators. She often did so through online resources:

"...they have their CVs online and their interest and so certainly find out more about the people that I want to collaborate with, um..."

The results of the field study show finding collaborators and resources as one of the main processes of MIBC. Thus, biomedical informatics, in order to facilitate collaboration must find ways to facilitate finding collaborators and resources. Chapter 6, Section 5 discusses this process further.

5.2.2 Socio-technical Factors Surrounding Collaborations

The preliminary framework of collaboration in Chapter 3 introduced some of the social factors that are associated with various collaborative processes. In the preliminary framework, the social factors were interspersed across various collaborative categories. The field study data also supports this trend. Unlike many of the concepts that were in some way directly related to a specific collaborative process, the concepts in this section are overarching to the whole collaboration and not associated with any specific activities or processes. These social and socio-technical factors affect all aspects of collaboration and were thus grouped together.

These social and socio-technical factors are discussed here to give a complete description of all the factors associated with biomedical collaboration that emerged from the study. These factors however, will not be discussed in further detail because they are not factors that the biomedical informatics community in general can tackle and alleviate due to their social nature. Yet the biomedical informatics community must still be aware of these factors if they are to build technology to help alleviate some of the difficulties associated with collaboration.

Developed Relationships, Trust

The investigators in both MIBCs expressed that developed relationships and trust are an integral part of collaboration. When asked for some of the important aspects of collaboration, many participants spoke of how close relationships among researchers are important in maintaining collaboration and forming new collaborative projects. Relationships build trust, understanding, and respect for one another, all of which are also important for success in collaboration.

Investigator G1 at GH mentioned that without shared interests and relationships, collaboration cannot exist:

> *"Um…so, that, in personal kind of collaboration, is critical. We talk to people in the network, why do people collaborate? It's because um,*

because of shared interest and relationship, in a nutshell. If you don't have that, you don't collaborate, or ways to support that."

Investigator G2 at GH also voiced a similar opinion. Trust that has been developed in a long-term existing relationships enables researchers to embark on collaboration without fear:

"...but in my opinion, it's the strength of individual relationship...the sense of trust ad long term commitment...so we did over time..."

Project manager G6 at GH described that the investigators she sees being involved in collaborative projects often have pre-existing relationships:

"People who're into do this kind of research are really developed relationship with people in other institutions who are doing the same type of research."

Existing working relationships were also important to the investigators at TH. Investigator T3 at TH described that the reason their collaboration was able to continue as long as it had was due to good working relationships:

"...for me, technology supplements, what it already a good working relationship. Because I've had other collaborative projects where that was it. That was it. We did one project and no more. And that was because...and that wasn't because technology failed us. It was because of the personal dynamics."

Researcher T4 at TH expressed that in addition to formal working relationships, the informal personal relationships with collaborators were also important:

"we just have a lot of fun and I think all of us would say that we are friends on top of everything else and the other thing that I think is very much the key is that we all pretty much have same work ethic."

Common Ground

The expression "common ground" can have many meanings. It can refer to all of the following: having a common goal or vision, like-mindedness, shared interest, common work style, and/or matching personality. Common ground fosters collaborative spirit among researchers by becoming the connection between people who often have disparate backgrounds. When asked what enables the continuation of collaboration, the most frequently mentioned factor was various aspects of common ground and how they fosters collaboration and enable it to continue.

Investigator G1 at GH mentioned that common research goals and a like-minded spirit helps people to collaborate:

> *"Maybe the nature of research that we do helps us collaborate. We definitely have lot of like minded spirit…"*

Investigator G2 at GH spoke of a common vision:

> *"I think in 10, 15 years, people are saying…we all work together, we have a common vision of how we should do work and we also work in these environments by choice…So they have a collective common vision."*

Researcher G5 at GH directly mentioned common ground and its relationship to trust:

> *"I think the first thing is finding common ground and trust out of gate…"*

The TH participants also expressed similar opinions in regards to a common vision. Investigator T1 at TH spoke of work styles and work ethic as a key to working well together as well as a desire for interdisciplinary research holding collaborators together:

> *"…key factor is the, we have the similar objectives and similar work style and similar work ethic so we, all the team members, want to grow, and we want to, um, you know, first of all we all want to do*

interdisciplinary approach so none of us wants to work just with our own discipline and we have working style match..."

Common ground is an important cornerstone on which participants build to develop relationships which lead to trust and collaborative spirit, both of which are important to collaboration.

Contextual Differences

Collaboration connects various investigators and institutions together. However, the institutions and investigators involved all have differences and that makes successful collaboration more difficult. The institutions have organizational and cultural differences. The investigators have differences in their backgrounds and research interests. Such contextual differences were one of the most frequently mentioned barriers or difficulties that participants faced during collaboration.

Researcher G5 at GH spoke of the institutional differences within GH and how they posed barriers to collaboration:

> *"...some of them are very large and have a very well developed research programs and some of them are newer, their research programs are less mature, their investigator pool is smaller so there's a capacity difference across the sites that really plays into how we collaborate, I think."*

Project manager G6 at GH mentioned that the biggest challenge is institutional differences:

> *"...University, medical centers, technology, like M, primary care, and others Cooperative, stand alone, but everybody's different, that's the biggest challenge...everybody's different."*

Investigator T3 at TH described the difficulties she faced when she moved from one institution to another due to the institutional differences in technology:

"...one problem I come across in this university is that they would not allow me to have that on my desktop, and that was the way this group communicated so I had to very quickly figure out a way around that..."

Contextual differences can result in strained relationships among institutions and/or researchers due to differences in work practices and opinions; therefore, it can be a serious threat to collaboration. Although a difficult task, contextual differences must be reconciled if any meaningful collaboration is to occur.

Incentives for Collaboration

Collaborative projects are often motivated by incentives such as having a better competitive edge for funding as a result of pooling resources, or having access to better data or expertise through collaboration. Without some incentive, it is much harder to overcome the difficulties of collaboration such as contextual differences. When asked what some facilitators of collaboration are, the participants mentioned many different kinds of incentives and how lack of incentives can sometimes prevent collaborative projects from forming.

Some of the incentives mentioned were: better idea generation through multiple perspectives, varied expertise, and synergy. Investigator G1 at GH described one institution within GH that rarely participated in collaborative projects because it had little incentive to do so:

"Some studies don't have to have multi site studies since those are huge sites, KPSC has millions people and they don't need multi studies for a lot of their work...often, organizational structure and efficiency for doing multi site studies with smaller, those folks...they don't have much incentive. That's my impression anyway."

Project manager G6 at GH described intellectual synergy as one of the incentives and discussed how many investigators valued it:

"...some people are really bought into collegiality and kind of intellectual energy that can happen when you get a whole bunch of really smart people in a room together and people that really understand the values of that. Great...when you have that."

According to a research assistant T7 at TH, an opportunity to have different ideas was mentioned as one of the incentives collaboration provided:

"So collaborating has really given us opportunity to have different ideas, but also kind of brought, I don't want to say credibility, but recognition of other profession."

Without some of the incentives mentioned, formation of collaborations would suffer since researchers would find very little reason to collaborate. The advantages of collaboration must outweigh the difficulties associated with it.

Willingness to Collaborate

Often those involved in collaboration, whether they are individual investigators or institutions, are protective of their own work and resources such as data. This protective nature often makes them unwilling to collaborate and share. Many participants acknowledged that unwillingness to collaborate was one of the difficulties they faced during collaboration.

Investigator G3 at GH described that people often are not willing to share and collaborate:

"People tend to be very tribal you know...my way or the highway. And to create shared systems, a lot of ours is collaborative you have to share rights and responsibilities with people from other..."

According to a programmer G11 at GH, people often resisted sharing because they felt threatened:

"We put this together and yet there is a lot of resistance on part of investigators because it allows people at other sites to see data at their

site even though it is just summary data. That's the whole turf issue again... Our problem is getting the users to not be afraid of sharing and turf stuff."

Investigator T6 at TH thought willingness to be open to collaboration was important for starting and maintaining collaborative projects. He felt that collaboration is not something that can be forced and it is a concept that needs to be in the minds of researchers before having to be part of it:

"I don't think you can force people, that you must collaborate...somehow the collaborative must form naturally and something has to hold them together."

Willingness to collaborate is one of the key ingredients in collaboration. Without willingness, no researcher will work with another and collaboration cannot form. Yet it is important to note that unwillingness can occur even when collaboration has already formed. Being able to manage unwillingness within collaboration is an important activity.

Time Consuming, A Lot of Work

MIBCs are often time consuming and require a lot of effort and work compared to non-collaborative research or collaborations within a single institution. The more institutions and/or researchers are involved, the more time consuming the project becomes because it has to reconcile institutional-specific processes or researcher-specific practices and differences in backgrounds.

When asked what some of the difficult aspects of collaboration are, investigator G1 at GH mentioned timelines and the amount of work involved in collaborative projects:

"...all of them have different forms, timeline, different consent forms, and you end up...if you want to please everybody, you end up with a

different project, different sites and then you don't have a multi site study...makes your timeline difficult."

Project manager G6 at GH described that often investigators do not wish to be involved in collaborative projects due to all hassle they have to go through:

"...collaborative research is some people don't want to do because of the hassle, so much more difficult to do even a two site study than one site study. You know, I mean really. Some people it's just not worth the hassle for them..."

Having to put in extra time and effort is a big disincentive for researchers who might otherwise be willing to participate in collaboration. Ways to reduce the additional effort and time required for collaborative research should be researched. Chapter 7, Section 4.2 briefly describes one way to help researchers save time.

Unnecessary Technology

Technology without need and technology that does not support the actual process of collaboration was not only unused, but sometimes made collaborative processes more difficult. Unnecessary technology was mentioned as one of the barriers to collaboration especially because it often wasted already limited resources. Investigator G2 at GH described a situation where much time and resources were wasted on technology that was not useful:

"... fundamentally a bad approach to a whole thing. It was under resourced, you had a group of people that had a vision that they didn't allow to be informed by needs that changed over time and we spent, my feeling is that, all the money spent on that website was wasted...however many thousands of dollars wasted."

Investigator G3 at GH described how tools that were not useful were seldom used:

"...we built a whole bunch of website and they tend not to be used very much..."

The participants also mentioned that in collaborative projects, there is often "reinventing of wheels" where instead of reusing existing technology or an existing process, yet another technology is created, thus wasting resources by doing so. Investigator G2 at GH described a specific example of how existing technology could be used better:

> *"Maybe Google's done for us already or...you know something else. I think people tend to sort of automatically rush to your own when it may not be worth the effort."*

Investigator G3 at GH discussed that it is wasteful to building something that might already exist:

> *"...why would we want to have cost of building a website, why wouldn't we just want to give people instructions for how to find most information on people who would be on the website and use Google or some other device that's already out there that does that."*

Ultimately, investigator G2 at GH mentioned that although technology might enable collaboration, it is not a solution to difficulties existing in collaborative projects unless it is designed support actual collaborative processes:

> *"...it was never about the technology...you know, it was about the process and people and technology could solve problems only if they were addressing appropriate problems. And appropriate problems were always personal problems... it was always about the people."*

Although technology is a necessity in MIBCs, the collective opinion is that technology is not what makes or breaks a collaborative effort. Therefore, biomedical informatics think beyond technology in the effort to support collaboration.

Involvement, Contribution

The feeling of making a contribution or of being involved in collaboration on an individual level is important for the information or continuation of a collaborative effort. When asked what some of the difficulties collaborations face are, investigator G4 at GH mentioned that when a researcher or institute feel as though they are only providing data or are involved in name only, there is no incentive enough for them to continue with the collaboration:

> *"...sometimes we feel like we don't have as much intellectual input... Um, so, sometimes when you're not leading a project, you feel like you're just a data warehouse for someone else's project, but that's fine every once in a while but if you have too many things like that, it can be frustrating, um, so that sometimes feels like it's not always worth it..."*

The importance of feeling involved should be noted and steps should be taken to make everyone involved in collaboration feel that they are contributing.

Inside vs. Outsider

The concept of insider vs. outsider exists both within an organization and between organizations and groups. Participants explained that individuals within an institution set themselves apart from those they consider "outsiders". They often used "we" in reference to "insiders" within the institution and spoke of differences between insiders and outsiders. The "insiders" often felt it was difficult to work with outsiders because of their lack of contextual sensitivity and differences in perspective. This perspective made collaboration more difficult by setting groups or individuals apart from each other.

Investigator G2 at GH explained that such views exist within GH:

> *"...but when you live in this environment and do most of your work in this environment, you develop a view of the world and view of how research is conducted that is different from the university based investigator. And you want to work with people who you don't have to*

spend weeks and months getting to the point of your common understanding."

Programmer G11 at GH explained how the concepts of insider and outsider are constantly changing:

> *"...when I first started at the center almost 25 years ago, there was the in-group, people at the cooperative, and those outside, you know, the people outside were docs and coops, those that are trying to get our data. And then, that changed to where the in-group was anyone in the coop and outsiders were the university or FH, then that changed to where the in-group is now includes a bunch of investigators at the university, a bunch people at FH"*

The insider vs. outsider frame of mind can cause insiders not to collaborate with those who they consider to be outsiders. To promote collaboration, efforts need to be made to promote social connections among researchers that would help them feel allied with others and consider more researchers to be inside their sphere.

5.3 Federal, Institutional Processes External to MIBC

MIBCs cannot occur in isolation from external forces. Many federal and institutional processes influence collaboration. Although these external processes are not factors that can be controlled from within collaboration, they affect each step of the collaborative lifecycle. Therefore, biomedical informatics must be aware of these factors if they are to support MIBC. The two most often mentioned outside influences are funding and regulatory-related processes.

Regulator- related Processes

Regulatory-related processes are not only those internal to collaboration as discussed in Section 2.1.1, but can also be external since every institution's internal review board that reviews biomedical research is guided by federal regulation. The effects of federal regulatory acts on MIBCs however, occur through internal regulatory processes that are guided by external regulatory acts. Thus, the regulatory process was discussed in detail in Chapter 2, Section 2.1.1, under factors internal to MIBC.

Funding related

Source of funding can influence the overall lifecycle of collaboration. Each collaborative project, regardless of institution or investigators involved, is associated with a different funding mechanism or agency. These funding agencies often have different goals and requirements that each project must follow throughout the research process. These differences often create difficulties in collaboration. Investigator G1 at GH explained that each collaborative project with different funding source has a different structure they must follow due to their funders:

> *"...and their relationships with their funders...like CDC and Vaccine Research have very different data management structure then Cancer Research..."*

Investigator G8 at GH explained various structures that exist within GH due to different funding sources:

> *"It's very different...well, it's funding mechanisms are different. They're both cooperative agreements. They are both cooperative agreements, but Breast research, all of us have our own grants, while Cancer research is one grant that has bunch of subcontracts. So, it's very different structure to the way the money flows and administrative burden at one site vs. others."*

The participants also frequently mentioned the difficulties of limited funding. Because funding is almost always limited in collaborative research, only the most crucial collaborative processes are supported. Some of the negative effects of limited funding are: not being able to involve all those who wish to participate in a project, not being able to have more a comprehensive infrastructure, difficulties funding preliminary stages of research before funding, and having to work beyond work hours to finish end of funding deliverables.

Investigator G1 at GH explained that limited funding makes it impossible to build a proper infrastructure:

"There is no budget to develop user centered design and then to...build, create infrastructure to create user centered design and I think the budget, the budget we have had, had not been kind of scoped for allowance for that."

Investigator G3 at GH described the difficulties collaboration face due to limited funding:

"...issues with funding and always, how do you pay for the work that needs to be done. Um...or there is a technical solution and you know, but how do you buy the equipment, but you don't have money..."

Investigator T2 at TH spoke of how they often have to find compromises and lower-cost options due to funding limitations:

"...but for the longest time I don't think we would have had money to do that, that's probably why we went more cheap route..."

The differences among funding agencies combined with the lack of funding can create major difficulties for MIBCs. Some of these difficulties might be alleviated if there are flexibilities in funding structure to allow collaborative projects funded by various funding agencies to work together and to possibly pool resources.

5.4 Summary

In Chapter 5, the overall characteristics of MIBCs that emerged from the contextual field study (described in Chapter 4) were discussed in depth. Included in this chapter are not only descriptions of the different settings themselves (GH, TH) and their similarities and differences, but also core activities and processes that happen throughout the collaborative cycle: a) data, b) regulatory, c) communication, d) administrative, and e) finding resources/collaborators. Also described are some of the social and socio-technical factors that impact overall collaboration (i.e. developed relationships, common ground, contextual differences) and external forces that influence collaboration such as funding.

As a first step to addressing, "How can biomedical informatics best support current multi-institutional, interdisciplinary biomedical collaboration?" the research question posed in Chapter 1, this chapter contributes by first providing an in-depth look into two MIBC settings (see Aim 1 in Chapter 1). It introduces how biomedical informatics can better support biomedical collaboration by shedding insight into some of the unique characteristics of MIBCs (i.e. data and regulatory processes). Some of the contextual data also reveal similarities of MIBCs to other collaborative settings to support which there already exists a large body of research in other fields that biomedical informatics can learn from (see Chapter 2). Furthermore, this chapter illustrates some of the overall social, socio-technical, and external issues associated with the overall process of collaboration which those in support of MIBCs should be aware of. Chapter 6 will delve deeper into the core collaborative activities and processes that biomedical informatics should focus on in support of MIBCs. Chapter 7 will synthesize Chapters 5 and 6 to develop a framework of Multi-institutional Interdisciplinary Biomedical Collaboration (fMIBC) which extends and validates the Preliminary Framework developed in Chapter 3.

Chapter 6: Opportunities for Biomedical Informatics: Collaborative Barriers, Facilitators, and Needs

In Chapter 5, all the factors that influence multi-institutional interdisciplinary biomedical collaboration (MIBC) were illustrated; however, the external and social issues are not aspects of collaboration that can be fully addressed by the biomedical informatics community. Therefore, this chapter delves deeper into only the core collaborative processes and activities, a subset of the factors that can potentially be addressed by biomedical informatics in support of MIBCs.

Better supporting users through assessing their "needs" and reducing "barriers" to system adoption is a well known concept in fields such as Computer Science (CS), Computer Supported Cooperative Work (CSCW), and Information Science (IS) (Lindgaard G). It has long been recognized that systems don't exist apart from their social context. In order to be adopted and used, a system has to fit the needs of the users. Faced with difficulties of too many systems that have been met with user resistance, the concept of needs and barriers has become increasingly important to the biomedical informatics community (Lorenzi 2004; Korjonen-Close 2005; Tang 2006; Tanner 2006; Anderson in Press). This concept of needs and barriers was expanded in this study during the analysis phase and used to group codes generated during the open coding phase.

After the axial coding phase, three distinct groups of codes intuitively emerged within each main category of collaborative processes: a) facilitators, b) barriers, and c) needs. The facilitators are those processes that current participants consider to work well and aid in collaborative efforts. These are the factors that biomedical informatics community should keep in mind and continue to support if already in existence or provide support for if no such facilitator exists. The barriers are factors that currently hinder collaboration from working smoothly or from forming. These are the factors that biomedical informatics can study further and find ways to alleviate these difficulties involved in the process of collaboration. The needs are factors that often arise from the barriers. They are the factors that those involved in collaboration thought could help with the barriers they face during collaborative research. In sections 6.1 to 6.5 of this chapter, each of the core collaborative processes and activities with its facilitators, barriers, and needs are discussed in depth (Table 6.1).

Table 6.1
Core activities and processes of two multi-institutional interdisciplinary biomedical collaborations (MIBCs). Core activities and processes are broken down into those factors mentioned as facilitators, barriers, and needs.

Category	Description	GH	TH	+	-	n
DAT	Data warehouse, repository	x	x	x		
DAT	File, data transfer	x		x		
DAT	Data standards	x		x		
DAT	Data security protocol	x	x		x	
DAT	Project specific data, various data purposes	x			x	
REG	Federal regulations, institutional review board	x	x		x	
REG	Common regulatory process	x				x
COM	Face-to-face meetings: local, conference	x	x	x		x
COM	Email: asynchronous communication	x	x	x		
COM	Phone, conference calls: synchronous communication	x	x	x		
COM	Communication technology limitation		x		x	
COM	Technology enabling face-to-face: video conferencing	x	x			x
ADM	Collaborative technology, infrastructure	x	x	x		
ADM	Support personnel help	x		x		
ADM	Information distribution: websites, online technology	x			x	
ADM	Schedule coordination	x	x			x
ADM	Access to information	x				x
ADM	Collaborative writing technology	x	x		x	x
FRC	Social connections, existing relationships	x	x	x		
FRC	Social connectors	x	x	x		
FRC	Researcher recruitment, promotion	x		x		
FRC	Distance, size of collaboration	x			x	
FRC	Online directory, information resource	x	x			x

Legend
COM - Communication
ADM - Administrative
DAT - Data
FRC - Resource, collaborator location
REG - Regulatory
GH - Exist in non-academic collaboration
TH - Exist in academic collaboration
+ - Facilitator
- - Barrier
n - Needs

6.1 Data

As described in Chapter 5.2.1.1, data-related processes occur throughout the lifecycle of collaborative projects. Due to the importance of data in biomedical collaboration, each of MIBC had at least some data-related technologies and processes in place. Some of these greatly aided researchers while others did not.

6.1.1 Facilitators

Data Warehouse, Repository

Data warehouse or data repository is a technology that facilitates storage of research data. During each step of collaborative biomedical research, a large amount of data is generated. The data then generally needs to be stored in some fashion so that they can later be analyzed and shared. Without storage of data, none of the subsequent steps after data generation can occur. Participants from both MIBCs mentioned the helpfulness of such technology.

When asked to describe some of the technologies used during collaborative activities, investigator G1 at GH immediately mentioned data warehouses. For him, data warehousing was a technology of primary importance and "valuable" during the collaborative process:

> *"So that's huge distributed data warehouse. I think we've talked about that last time. I think that's pretty valuable. That's a valuable technology..."*

When asked the same question, investigator T3 at TH mentioned a common web-based technology that TH uses to store data. He mentioned how "nice" it is to be able to store data at a central location. He felt that being able to access project data at any time made it easier and more convenient for him to participate in the collaboration:

> *"And we also have a sharepoint website, where we keep all of our shared data...and that's real nice because I feel like I'm always*

attached to the project...within an hour of something I'm working on. It works out really well."

The evidence of the importance of data repositories was also seen in the artifacts examined (see Figure 6.1 for a sample artifact). This and other infrastructure documents describe in detail the custom-made virtual data storage tool built by GH. Participants also often discussed the data warehouse during local and national conference meetings (see Figure 6.2 for a sample observational note). Because many of the collaborative projects within GH involved the warehouse, it was almost always a part of the agenda during any meeting.

A data warehouse or repository might not be a necessity in small collaborations where small datasets can be stored in spreadsheets or a local file system. However, in large complex collaborations such as MIBCs, it is almost impossible to manage data otherwise. Technologies such as data warehousing make it possible to store and manage data at a central place and doing so, makes collaborative sharing and analysis possible.

C.4.a. The Virtual Data Warehouse (VDW)

...

standardized data resources that would facilitate more efficient aggregation and use of automated data.

...

The VDW is 1) a series of computerized data sets stored behind separate security firewalls at participating sites that include variables with identical names, formats, and specifications and identical variable definitions, labels, coding, and definitions; 2) a set of informatics tools—hardware and software—that facilitates storage, retrieval, processing, and managing VDW datasets; 3) a set of access policies and procedures governing use of VDW resources; and 4) documentation of all elements of the VDW.

...

Figure 1. Schematic of the Virtual Data Warehouse

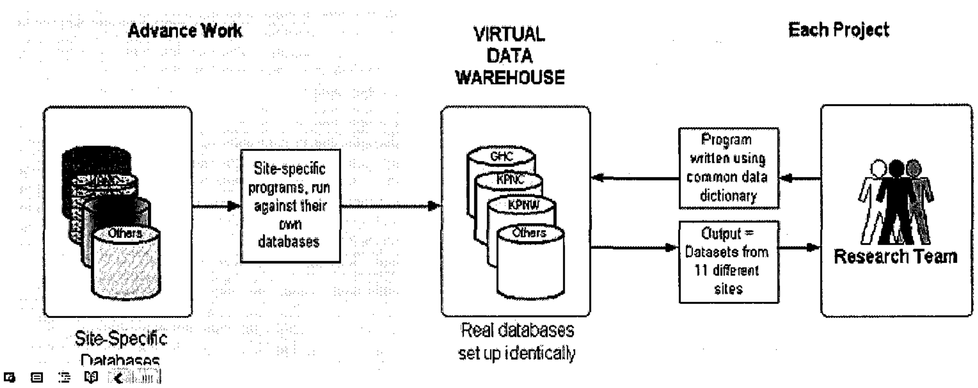


Figure 6.1
A sample artifact at GH describing its data warehouse.

5/22/07 -
Called in

May 22 Noon PT

Review Action items of previous meeting	12:00 – 12:10
Announcements Data warehouse programs underway or planned Obesity HER2 Other Other	12:10 – 12:20
website process Timeline Training process SDMs will have admin role for their site	12:20 – 12:30
QI process	12:30 – 12:40
Data warehouse Implementation issues at new sites	12:40 – 12:50
Other	12:50 – 1:00

Action Items
Topics for another call are: Automating the use of the data warehouse Microdata release

Figure 6.2
Sample observation notes of a collaborative meeting at GH. The red circles indicate where data warehousing technology is mentioned.

File, Data Transfer

Throughout the lifecycle of research, data is often shared and analyzed among various institutions and/or researchers. In order to do so, some secure mechanism for transferring data between locations and/or researchers is needed. Research participants from GH often mentioned data transfer technology as a helpful technology that enables them to transfer data between researchers and institutions, enabling shared analysis and collaboration.

When asked to describe some of the technologies used during collaborative activities, investigator G2 at GH said that the researchers at GH used a web interface to securely transfer data. He felt that it was a good tool that facilitates data sharing during research:

> *"...they developed this web-based interface for secure transfer that's been doing...essentially creating a secure common folder that allows us*

to transfer data back and forth, that satisfies all the HIPAA, all the IRB requirements, so it's been really nice tool..."

Although the participants at TH also mentioned that they frequently share data, they had no specific file transfer technology for doing so:

The evidence of existing file transfer technology enabling data transfer was also seen in the artifacts examined. The proposal for GH infrastructure describe in detail the data aggregation and transfer tool (see Figure 6.3). These MIBCs involve large sets of data that often cannot be transferred over simple tools such as email. A secure file or data transfer technology enables exchange of large data sets among researchers located remotely. To facilitate collaborative research in MIBCs, file or data transfer technology is important.

Aim 2b. Automate Multi-site VDW Data Aggregation.

The current version of the VDW requires independent manual execution of each data processing step at each site. Thus, the number of sites participating in a particular study multiplies the amount of programming labor and the time required for data aggregation. Significant savings in time and programmer labor could be realized if only one programmer was needed to perform all the data processing for all the sites using automated processes. We propose to develop an automated VDW using secure Web links that enable transfer and remote execution of programs at participating sites with appropriate authentications. Thus, a group of collaborating scientists from several sites would develop a well-defined data analysis plan for an approved research project. The lead programmer would write the programs required to retrieve data, create the analysis file, run error checks, perform descriptive statistical analyses, and run inferential hypothesis tests. The programs should execute successfully at all other VDW sites. Under an automated VDW, local programmer responsibilities should be limited to troubleshooting.

Figure 6.3
A sample artifact at GH describing its file transfer technology.

DataSstandards

Sharing data often involves aggregating data across various institutions and/or projects. This aggregation process, however, cannot occur without some data standardization since data at various institutions are collected in different format and stored in various locations. The GH participants noted that having a standardized, agreed-upon format is essential to successfully aggregating data across multiple locations.

Investigator G4 at GH explained the data sharing process that exists within GH. She said that standardization of variables (data) were helpful in making research more effective both in cost and efficiency:

> *"Those standardizing of variables have been extremely helpful in both making these types of studies more cost effective. What else's been helpful…this isn't really technology thing…the biggest thing that's helpful now is virtual data warehouse, because it has standardized data elements, so you know, blood pressure is coded the same way, there's, data is clean before it's essentially put in there so…"*

The evidence of existing data standardization was also seen in the artifacts examined (see Figure 6.4). The infrastructure documents that describe the data warehouse also discuss the standardization that exists to enable aggregation. Since collaborative research cannot occur without shared data and analyses, data standardization facilitates collaboration by making aggregation of data located at multiple places and in different formats possible.

INFRASTRUCTURE

C.4.a. The Virtual Data Warehouse (VDW)
Lack of consensus and the rapid start-up of the core projects slowed progress toward data standardization in [redacted] (see Appendix E).[9, 37] The proposal for CRN2 made clear our commitment to creating standardized data resources that would facilitate more efficient aggregation and use of automated data. The ~~responsibility~~ for this

Figure 6.4
A sample artifact at GH describing its data standardization approach.

6.1.2 Barriers

Data Security Protocol

Although sharing data is a necessity in collaboration, extra steps are needed to protect the data in biomedical research due to the sensitive nature of the data involved.

Rigorous adherence to data security protocols and data de-identification processes is a norm in biomedical research. Therefore, both MIBCs had multiple protocols in place to protect data and all the participants were aware of the necessity of doing so. Nevertheless, the complex data security protocol in place was yet another hurdle participants felt they had to cross. During collaborative projects, the data security protocol often caused extra concern and slowed down the overall research process. Research assistant T7 at TH remarked that although a necessity, the effort required to follow complex data security protocols imposed extra difficulties during data sharing:

> *"And I think just in terms of all of the, not ridiculous things, they are necessary, but all of the kind of ridiculous steps you have to take to de-identify information, and also about e-mailing it ... "*

Information about data security protocols was also seen in the artifacts examined and during observational sessions. The infrastructure documents described security protocols associated with the data warehouse (see Figure 6.5). During the local and national meetings researchers often spoke of security protocols and how important they are to the overall research practice (see Figure 6.6 for a sample observational note). Although the difficulties associated with following data security protocols cannot be removed completely from MIBCs because the protocols are required by federal regulations, there are ways to reduce some of the burden on users. For example, building in functionality that follows security protocols into the data sharing and management tools themselves might shield users from having to deal with them directly

.

meet these multiple needs and named it the ~~Virtual Data~~ Warehouse (VDW). The VDW is 1) a series
iterized data sets stored behind separate security firewalls at participating sites that include
: with identical names, formats, and specifications and identical variable definitions, labels, coding, and
is; 2) a set of informatics tools—hardware and software—that facilitates storage, retrieval, processing,
aging VDW datasets; 3) a set of access policies and procedures governing use of VDW resources;
ocumentation of all elements of the VDW. The VDW serves as the source of standardized data from a

Figure 6.5
A sample artifact at GH describing its data security protocol.

J: wiki be good place for review?
(few agreements)
L: Nice idea but instantaneous feedback might be good
J: supplemental?
X: do in email?
X2: email is good too. Firewall issues, good practice
L: we have listserve ad if we do this, I'll only respond if I can do it 5 sec.
But for code review, then more theoretical
Already have discussion groups for program issues
→ little interest for code review?
We do powerpoint here presentation

Figure 6.6
A sample observation note of a collaborative meeting at GH. The red circle indicates discussion of data security among participants.

Project Specific Data, Various Data Purposes

Although data are shared as part of any collaborative project, each collaborative project involving multiple institutions and even each individual institution can have its own specific way of managing data. The variability in handling data is due to institutional differences, different data needs, and funding related differences. The participants mentioned some difficulty and tension associated with aggregating data across such variable environments.

When asked about sharing data in general, project manager G6 at GH described the various data sources that exist within GH. She mentioned that there are multiple data sources created by multiple projects and multiple institutions. In order to work on a collaborative project together, data must be aggregated across all these entities with their different formats and processes incorporated. She felt that such variability in data made data aggregation difficult and time consuming:

> *"Totally different data sources and people write code to talk to all of those...formatting is different, everything's different."*

Even within a single organization, the purpose of data creation can be different from how they were being used later. Often, the data were initially collected for managerial or business purposes, and was later repurposed to be utilized for research purposes. Because they were not intended for research, data were generally difficult to use and did not live up to scientific research standards and researchers' expectations.

Project manager G6 at GH discussed an instance where existing institutional data that researchers were trying to utilize were not created for research purposes. She felt that due to having an origin that is not geared toward research, the data that were collected for business purposes, although a rich source of information, were difficult to use:

> *"And you know, all of this data were really created for business purposes and not for research and in some cases, not for all there's home grown codes and stuff. It's a challenge, and he can tell you more about that, but that undoubtedly is big area of collaboration."*

Differences in institutional and/or funding practices can also make it difficult for researchers to share data across institutions and projects. Planning prior to the actual collaborative endeavor to build in flexible use agreements among institutions and projects and flexible funding policies can greatly aid current and long-term use of MIBC data.

6.1.3 Needs

Needs related to data process were not explicitly expressed by any participant.

6.1.4 Summary

Section 6.1 described the facilitators, barriers and needs associated with data-related processes. The facilitators that aided data activities included technologies such as data warehouses and data transfer tools and associated data standards. These tools facilitated collaborative research by enabling data sharing and analysis. The barriers included organizational issues such as security protocols and data purposes. Although the difficulties of data security and differences in institutional and funding agency requirements cannot be removed completely, steps can be taken to ease some of the burdens that falls on the researchers. There were no specific needs associated with data process that emerged from the analysis.

6.2 Regulatory

As described in Chapter 5, Section 2.1.1, there are regulatory processes that must be in place for biomedical research to take place. Hence, both MIBCs dealt regularly with federal and institutional regulatory processes. Although the participants understood the necessity for having such regulatory processes in place, the difficulties associated with the processes often hindered or stopped collaborative projects altogether.

6.2.1 Facilitators

There were no facilitators related to regulatory process expressed by any participant. In general, regulatory processes posed major frustrations for the participants in their day to day research activities; therefore, none of the existing regulatory elements were thought to greatly aid MIBC.

6.2.2 Barriers

Federal Regulations, Institutional Review Boards

Federal regulations implemented and controlled by an institutional review board (IRB) were one of the primary concerns participants had when being involved in multi-site collaborative projects. The IRB process is time consuming and difficult enough even for investigators involved in a single biomedical research project. The investigators involved in MIBCs had to go through multiple separate IRBs, usually one for each institutions involved, and each of them as difficult and time consuming as the others.

An aspect of multi-site collaboration that was mentioned as a barrier most frequently was the IRB process. Investigator G1 at GH mentioned that IRB is one of the biggest rate limiting and difficult steps. The frustration of having to go through multiple IRBs, each one a time consuming process, and having to reconcile the differences among these IRBs often deterred participants from initiating collaborative projects:

> *"People get hung up on IRB. Yes...multi site projects, IRBs can't even agree on what's wrong...all of them have different forms, timeline, different consent forms, and you end up...if you want to please everybody, you end up with a different project, different sites and then you don't have a multi site study...makes your timeline difficult."*

Investigator G2 at GH described in detail the difficulties of multiple IRB processes investigators in collaborative projects must go through:

> *"So there's an example of where the barriers to...I was involved in one network study where there were 7 sites involved in study, and every IRB had to approve and then every IRB had modifications which had to go back to every other IRB and IRBs are by definition, conservative bodies. But SG is leading the effort to actually say, can we have basic trust that if we know in advance the concerns IRB have, you are going to give them...and they may accept the...northwest group...it will be people like...people who will be receptive."*

The evidence differences in IRBs was also seen in the artifacts examined (see Figure 6.7) and during observational sessions (see Figure 6.8 for a sample observational note). An artifact at GH describing the result of a survey regarding IRB processes illustrates differences that exist among various institutional IRBs. During meeting observations, IRBs were frequently mentioned in association with any type of research. In general, the regulatory processes required by the institutional review board were considered to be rate limiting and frustrating by most participants. It would greatly aid MIBC if there was a way to relieve some of the difficulties of the regulatory process and make it less time consuming.

Through the auspices of the , a survey of IRB policies and practices was conducted in 2000 to assist investigators in planning IRB reviews. The survey underscores the wide-ranging procedural differences across 10 research institutions including: differing IRB applications forms; conflicting requirements for supplemental information; variability in the use of expedited review; conflicting policies regarding consent waivers; and variability in how modifications and amendments are handled. IRBs also differ in their meeting frequency (10-24 times per year), and the elapsed time between submission and notification of committee deliberations (3-8 weeks). The CRN inventory of IRB procedures will be expanded and updated as part of these deliverables.

Figure 6.7
A sample artifact at GH describing the differences in IRB processes.

Will contact for volunteers (laugh)
<general collaborative spirit>
P4: [Any questions or problem with something sent out, let me know
Verify IRB before data transmit]
P5: Didn't transmit because of problem.
Advice on do this, and that about the problem
L: Is this specific to you or...?
P5: Not clear about this code (laugh)
(More responses in advice)
L: [Other data warehouse stuff?

Figure 6.8
Sample observation notes taken during a meeting at GH. The red circle indicates where participants talked about IRB processes.

6.2.3 Needs

Common Regulatory Process

The participants involved in MIBC were keenly aware of the difficulties of multiple IRBs when doing multi-site collaborative projects. Multiple IRBs were often the primary roadblock to multi-institutional collaboration and many participants felt that difficulties can be alleviated by having a common IRB process. When asked what could help make collaboration easier, investigator G4 at GH mentioned a common IRB process as a primary need:

> *"…if there was one IRB, it would a lot less man power, and also take a lot less time."*

Knowing how difficult it would be to arrange for a single common IRB process, project manager G7 at GH suggested that even having a common IRB template that is acceptable to all institutions would help reduce the IRB time:

> *"…having things, standards at each organization so that we would accept each other's IRB and for the programmers, it would be if we all had same systems, um…if we all…probably 90% of us do, if we had same standards where we code, probably together…um…same template, if we just had Cancer research office that could help organization be more efficient along collaboration…"*

Evidence of the need for a common IRB process was also seen in the artifacts examined (see Figure 6.9). Recognizing the need, GH has been actively pursuing a way to implement a common IRB process. The common IRB process was often discussed during meetings (see Figure 6.10 for a sample observational note). A common IRB was thought to be one of the main ways to relieve difficulties of multiple IRB processes associated with MIBCs. The participants also felt that a common IRB process would save great deal of time. A possible way to facilitate common regulatory processes is described in Chapter 7, Section 4.2.

3. Deliverable updates
   - The closeout and survey training guides will be posted on website in the next few days.
   - The IRB meeting in San Francisco in October is coming together nicely. It will include a presentation by creating a centralized IRB. IRB administrators and/or scientists from all but one ~~site will attend.~~ Follow-up meetings are planned at the December 06 Roadmap Steering Committee meeting and the March 2007 meeting. It was suggested that a follow-up discussion also occur at the November 2006 meeting. Sarah announced a possible upcoming opportunity for the to participate in an IOM study on the impact of HIPAA on research.

Figure 6.9
A sample of an artifact at GH describing a plan for centralized IRB.

Qs?
Work mostly done?
- yah (all around agreement)
- deliverables on track
- IRB had ambitious gold but, having one IRB for at least few common studies
- Notion of common IRB will be incremental
- NCa: Changing local one will have regional implication, will let university know as I read it, isn't it getting more difficult?
- Contextual importance
- SG: changes always difficult
- NCa: U should know this, local context
- SG: okay, thanks

Figure 6.10
A sample observation note taken during a meeting at GH. The red circle indicates where participants talked about common IRB.

6.2.4 Summary

Section 6.2 described the facilitators, barriers and needs associated with regulatory processes. No facilitators associated with regulatory processes were mentioned since the entire regulatory process was burdensome to the researchers. Therefore, none of the existing regulatory elements were thought to facilitate collaborative processes. The whole IRB process itself was most often discussed as a barriers, especially since multi-institutional collaboration involves multiple IRBs. Because multiple IRBs processes were time consuming and frustrating for participants, their primary need was a common IRB process that is acceptable across multiple institutions. The participants felt that a common IRB process would save time and help remove some of the difficulties of having to deal with not only single, but multiple IRBs associated with MIBCs.

6.3 Communication

As described in Chapter 5, Section 2.1.2, communication is an essential process during collaborative efforts. Both MIBCs used many different methods of communication. Most existing communication methods facilitated collaboration; however, some technical aspects of the communication process were still burdensome to the researchers.

6.3.1 Facilitators

Face-to-Face Meetings: Local, Conferences

MIBCs involve a lot of face-to-face meetings. The researchers involved in MIBCs thought that face-to-face meetings were helpful and essential throughout the collaborative processes from the formation of ideas to the continuation of existing relationships. These meetings brought a "personal aspect" to the otherwise formal collaborative process; thus, face-to-face meetings were thought to facilitate stronger relationships among participants and to strengthen collaboration.

Investigator G1 at GH mentioned that face-to-face meetings were one of the most important communication strategies in collaboration. Although many mechanisms for communication exist, he thought that face-to-face meetings were essential to building personal relationships and fostering collaboration:

> *"...sit down and chat, it's really hard to do collaboration without that personal aspect, um..."*

Investigator T3 at TH explained that the face-to-face meeting is also important to generat ideas for forming new collaborative projects. He felt that other ways of communication were not as effective in this area:

> *"…there's nothing better than sitting down in person and generating ideas that way too."*

The evidence of face-to-to face meetings was also seen in the artifacts examined (see Figure 6.11). More than a quarter of artifacts were based on the local meetings and many more described regular national meetings that occur among investigators at GH. Although distance makes face-to-face meetings difficult, a majority of participants at both MIBCs felt collaboration could occur without some face-to-face meetings.

generating new collaborative relationships and propos-
The topics have expanded to include obesity, end-of-life care,
ity of cancer care, the development of biorepositories, and
il disparities. Interest groups meet regularly by conference
or in-person during annual meetings.

Figure 6.11
A sample artifact at GH describing the regularity of face-to-face meeting.

Email: Asynchronous Communication

Email was a technology often used to exchange information, and relay introductions, and project announcements. Participants at both MIBCs spoke a great deal about email communication that occurs throughout the collaborative project. Email was perceived to be an essential technology that collaborative projects cannot do without. It was frequently mentioned by participants when asked about some of technologies used to communicate during collaboration. Project manager G6 at GH described how important email is to her. She felt she could not do her job of managing collaborative projects without email:

"Email…that's the thing. My kind of job is email so…technology. Seriously, my life evolves around email. I'm sort of the point person who sort of connects all the pieces and you know, got a few people here…almost everybody I interact with is somewhere else in the country and I almost never talk on the phone with anyone. Everything I do is sort of through email. Without email, without technology…I'm going to be lost. So, even just technology we take for granted these days is absolute essential."

The evidence of email usage was also seen during observation sessions (see Figure 6.12). Whenever any type of follow up to an action item was needed, participants spoke of doing so over email. In MIBCs, communication occurs primarily through a technical medium. Email was one of the primary technologies participants used for this purpose. Without this communication technology, participants felt they could not do any of their routine collaborative activities.

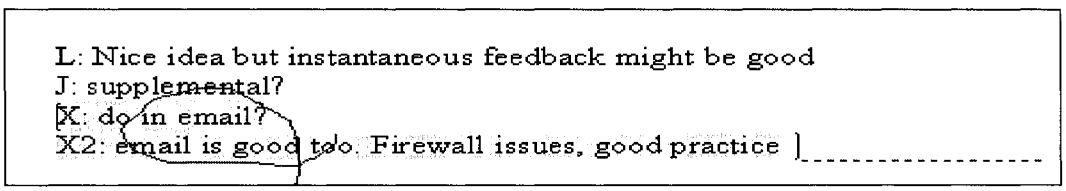

L: Nice idea but instantaneous feedback might be good
J: supplemental?
[K: do in email?
X2: email is good too. Firewall issues, good practice]

Figure 6.12
A sample observation note taken during a meeting at GH. The red circle indicates talk of email usage.

Phone, Conference Calls: Synchronous Communication

Although much communication can be done over email, this medium has a limitation of not being able to support a real-time conversation. Hence, the participants at both MIBCs felt that real-time synchronous communication methods such as phone or conference calls were also important to the overall collaborative process. It was generally accepted that phone and email complement each other. Phone communication enabled real time communication which was sometimes necessary, while email enabled a different type of exchange. When asked about some of the ways she communicated as part of collaboration, investigator T3 at TH mentioned the phone as a primary tool:

> *"...followed by the phone, and that includes cell phones too because all my collaborators have my cell phone number as well as my office number. It's not uncommon for them to call me at my cell phone number while I'm out doing things."*

Evidence of phone usage was also seen in the artifacts examined (see Figure 6.13). Artifacts describing collaborative meeting sessions often described conference calls. Usage of the phone was also often observed during meetings (see Figure 6.14). Due to the distance separating institutions, most of the non-local meetings were done through phone conferencing. Without both email and phone technologies or equivalent technologies, MIBCs would not be able to function.

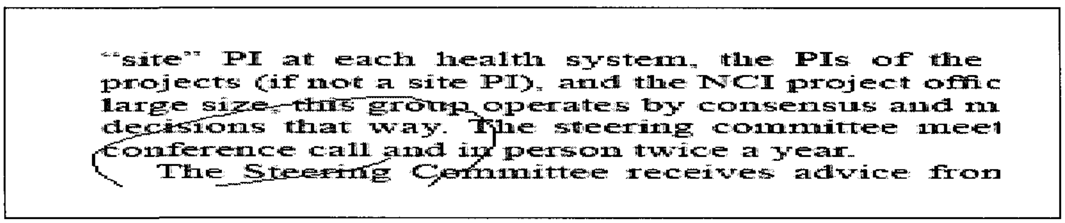
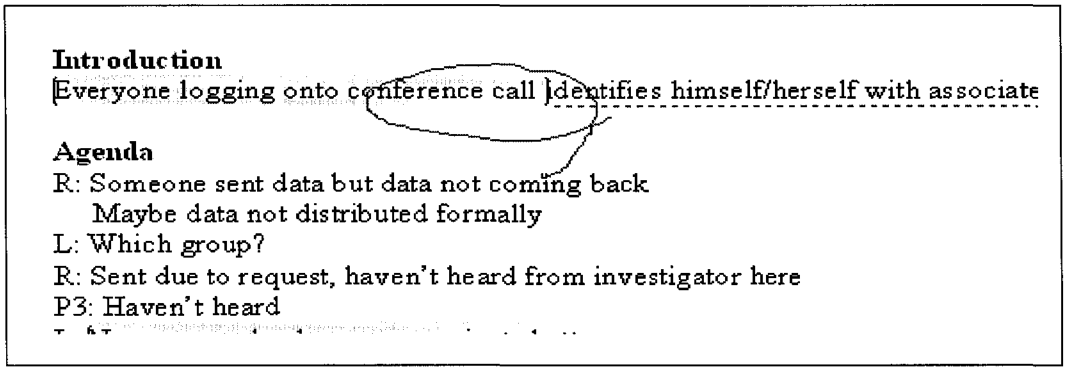

"site" PI at each health system, the PIs of the
projects (if not a site PI), and the NCI project offic
large size, this group operates by consensus and m
decisions that way. The steering committee meet
conference call and in person twice a year.
The Steering Committee receives advice fron

Figure 6.13
A sample artifact at GH describing phone usage.



Introduction
Everyone logging onto conference call identifies himself/herself with associate

Agenda
R: Someone sent data but data not coming back
Maybe data not distributed formally
L: Which group?
R: Sent due to request, haven't heard from investigator here
P3: Haven't heard

Figure 6.14
A sample observation note taken during a meeting at GH. The red circle indicates where participants used conference calling to conduct a meeting.

6.3.2 Barriers

Limitations of Communication Technology

Many participants at TH spoke of the difficulties they had using a video-conferencing technology. Due to budget limitations, the video-conferencing technology available to them was limited in capacity. Although video-conferencing was desired because it simulated face-to-face interaction, TH participants found the difficulty of use and limitations of the technology frustrating and soon abandoned its use. Investigator T1 at TH considered video-conferencing to be one of the main difficulties he faced during collaboration. He found it frustrating because it was limited in capacity and did not provide the type of features needed to support collaborative meeting:

> *"We had videoconferencing when we were using video phones where it was a big group, but the camera was focusing on only two members, and for me, it was somewhat frustrating because you hear voices and you wouldn't know where it was coming from. It was no way to zoom in or move the camera. So those are kind of cases where we make note and say next time let's not do that. But we haven't had any technology, in a sense the technology was frustration."*

Evidence of difficulties caused by poor communication technologies was also seen during observational sessions (see Figure 6.15). During one conference call meeting, problems with the phone delayed the meeting by over 20 minutes. Although technology is essential to communication in remote collaboration, limitations of technology were often seen to frustrate the participants. Although more powerful and complex communication technologies such as video conferencing were desired, they were often abandoned if they did not appropriately meet the needs of the researchers.

  - Great at recording
  - Document who's doing why
- 11 a.m. 1st Mondays of month do another meeting (*Request transcript)
- WA: not always ~~call on~~ time, always little early or late
- Speaker phone problem? (usually never use for chat session so little diffi
- [Who's going 1st? (going right to the business)]
- R21

Figure 6.15
A sample observation note taken during phone meeting at TH. The red circle indicates an example of difficulties with communication technology.

6.3.3 Needs

More Face-to-Face Meetings

MIBC by nature requires researchers to work with each other remotely. Although communication technologies such as email or phone help bridge the distance, the participants at both MIBCs felt that the face-to-face communication could not be replaced by any technology. The participants felt that even more opportunities for face-to-face time would greatly improve the collaborative research process.

When asked what would make collaboration easier are, investigator G2 at GH mentioned face-to-face meetings:

> *"I mean, I'm trying to think of...again I think what will make easier is the ability to be together more often...that's not technology."*

Due to budget constraints and the need to work remotely, participants were not able to meet as often as they would have liked. They felt having more face-to-face meetings strengthened relationships and helped generate more ideas; thus, strengthening the collaboration:

> *"...I don't know how, but if we could meet face to face and more talk, in more cost effective way so that we don't have to charge our grants for all these meetings as well as time scheduling and so forth, that would be great..."*

Although the importance of face-to-face meetings was an acknowledged fact, the limitations of time and funding made it difficult for participants to regularly meet face-to-face. Participants felt that allowing more funding for travel thereby allowing more face-to-face meetings, would greatly aid collaborative research.

Technology Enabling Face-to-Face Meetings: Video Conferencing

Although face-to-face meetings are essential to collaborative projects, it is difficult to have a lot of face-to face interaction in MIBC due to distance and financial constraints. If face-to face meetings were not possible, participants of both MIBCs felt that the next best solution would be a technology that can enable face-to-face like meetings. Participants wanted some way to experience an actual human-to-human interaction such as being able to see the other person and their facial expressions as part of the interaction. Investigator G1 at GH mentioned video-conferencing technology as one of the wish list items he wanted to make collaboration easier:

> *"...super easy, in my office, video connection. So, if I had a big white board there and I had a big screen there, and live video connection to TP and his office, or wherever TP was on his cell phone or wherever he was, TP and I could actually walk through the same thing that we try to do here with blackboard...I could look at TP and hear what he has to say and he has white board in his room and he can talk to me about that, there's something about the video feed, something about having a presence of a person there, talking to him, or like this, you're kind of*

talking and doing that. It's just something about that that you can do, that you can't do over the phone."

Investigator T4 at TH mentioned that video-conferencing would probably be one of the most useful things to have during collaboration:

"We certainly wish we had videoconferencing that would work for everybody. Um, that would probably be single biggest thing that would be helpful."

If more time and funding to enable face-to-face interaction area not possible, participants felt that any technology that would simulate face-to-face interaction such as videoconferencing would be the next best thing.

6.3.4 Summary

Section 6.3 described the facilitators, barriers and needs associated with communication processes. The facilitators associated with communication were face-to-face meetings, and communication technologies such as email and phone. Although face-to-face meetings are essential, in remote collaboration, distance often makes them difficult to conduct. Participants felt that collaboration could occur without communication technology to compensate for the lack of face-to-face communication and to enable face-to-face like communication. The biggest barrier associated with communication was in the limitations of low-cost communication technology. When technologies that were supposed to enable communication did not meet communication needs, participants felt no incentive to continue to use them and soon abandoned their use. Although participants felt that technology could help facilitate remote collaboration, they consistently indicated that face-to-face meetings were still an integral part of collaboration and as much face-to-face interaction was desired. If face-to-face meetings were not possible, the next best choice was thought to be technology like videoconferencing that simulated such interaction.

6.4 Administrative

In Chapter 5.2.1.1, the administrative processes were introduced as processes that apply to and affect every part of the lifecycle of collaboration. Administrative processes include all day-to-day collaborative activities aside from communication. Many of these processes existed within both MIBCs and most were vitally important to maintaining collaboration. For that reason, members of both MIBCs spoke of many facilitators, barriers, and needs associated with administrative processes.

6.4.1 Facilitators

Collaborative Technology, Infrastructure

Participants at both MIBCs felt that tools and infrastructure that supported collaborative processes in general made collaboration easier. These technologies facilitated collaboration by enabling communication, data sharing and coordination. Essentially collaborative technologies helped to virtually bridge the distance between remotely located researchers.

Investigator G2 at GH readily admitted that the technical infrastructure the technical tools that support collaboration are useful and are used daily as part of collaboration:

> *"I think people using technology, web based communications, I think they're useful…Technology has made a lot easier…secure web transfers, FTP, a lot easier than shipping tapes."*

According to G2, technology facilitates collaboration:

> *"…technology can facilitate how we social network…"*

Investigator T3 at TH thought that technology helped sustain relationships among collaborators:

> *"…but I think the technology helps to sustain that relationship…"*

Evidence of the importance of collaborative technology was seen in the artifacts examined (see Figure 6.16). The infrastructure documents examined described technological infrastructure as the thing that makes collaboration possible. Most participants felt that without technology, remote collaboration would not be possible. Due to the distance separating researchers, most interaction among researchers has to be done through a technical medium. Any technology that helped to easily facilitate remote collaborative interaction was considered valuable.

> The extraordinary opportunities for research in these settings *(1)* led to refunding and expansion of the through 2007. This monograph is more than simply a set of papers arising from a multisite research program. It addresses substantive and methodologic issues on screening, disparities, translation of research into practice, cancer care, risk reduction, and health services, as well as the expanding infrastructure that makes it possible to examine key research questions across multiple

Figure 6.16
A sample artifact at GH describing the importance of technical infrastructure.

Support Personnel Assistance

All administrative processes associated with collaboration of any size are difficult and they become especially difficult when the collaboration spans multiple institutions. Participants indicated that when a collaborative project does not have sufficient support personnel to help with administrative processes, the research part of collaboration suffers. When support personnel do exist, they ease the burden of time consuming administrative processes and enable the research part of a collaboration to happen. Without them, collaboration becomes much more difficult to maintain. During the interview, investigator G8 at GH described how integral administrative support is to collaboration:

> *"I think having somebody who is keeping track of minutes, agendas, and conference calls is unbelievably important. We are trying to start up a new project within cancer research and the amount of time it takes in these sort of task is pretty vast…Having adequate administrative*

support to keep track of minutes and agendas, be the point person of communication about availability is really important"

Evidence of the importance of support personnel to help with administrative tasks was seen during observational sessions (see Figure 6.17). Often, administrative support was in charge of coordinating meetings and was called upon to keep track of open agenda items. MIBCs involve many administrative processes such as scheduling and documentation that are necessary for the daily operation of collaboration. Administrative support can manage administrative processes so that the researchers can devote their time to research and not be bogged down by administrative tasks.

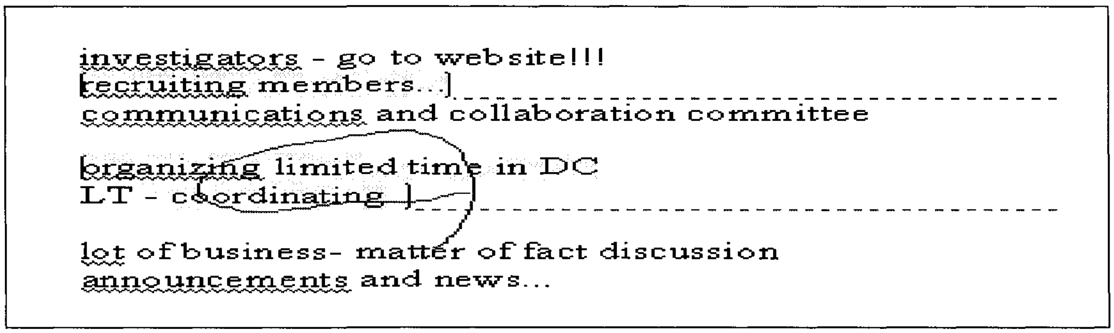


Figure 6.17
A sample observation note taken during a meeting at GH. The red circle indicates an example of the importance of support personnel.

6.4.2 Barriers

Information Distribution, Websites, Online Technology

Aside from research data, much administrative information needs to be distributed within a collaborative project. Participants mentioned that websites and other online technologies were often used to disseminate such information and felt they were appropriate mediums. However, they had mixed opinions regarding the effectiveness of such technologies since they often were not built to support the processes. The lack of support for the process made them not very helpful, if not, useless.

Project manager G7 at GH mentioned that a website that was supposed to help her do much of her work tasks was not appropriate for her and not helpful:

> *"...a lot of extra work that goes through me as admin specialist, are...when is this meeting, what time is it, what is call information, agenda, and it would be, we're working on a website that should be very user friendly to...The current one you can't really do this..."*

Project manager G6 at GH mentioned that a website that was supposed to be used to help disseminate information was often ineffective:

> *"And you know...that's an issue too...another good point is, you can make all this stuff, create all these resources, but you know, to help your collaboration along, but if you don't know how to disseminate them, and how to actually put them in the hands of people who need them..."*

Participants felt that existing technology for disseminating information did not fit their needs; thus, it was not helpful. And because the technology that was often not helpful it was not used.

6.4.3 Needs

Schedule Coordination

Busy schedules and distance make coordinating schedules within MIBCs much more difficult than within a co-located collaboration. Participants from both MIBCs mentioned the difficulties of schedule coordination. They felt that having a common calendar system might alleviate such difficulties. Project manager G7 at GH felt a common scheduling system across multiple institutions would greatly help the collaborative process by helping to schedule activities and coordinate schedules more effectively:

> *"...yah, and new website will also be able to connect through people's outlook emails if they have outlook calendars so, it will be helpful in reminding."*

Investigator G8 at GH mentioned that her institutional calendaring system greatly helped with schedule coordination and wished for network-wide implementation of the tool:

> *"...what we have internally, uh, would be available for collaborators, would really move things, more rapid, but that hasn't been implemented outside here, but that does help collaborative projects internally...common used calendar searching function across people, so we have that internally, where people's calendar's are all supposed to be up to date and online, so somebody setting up a meeting can see when people are free..."*

Although many technologies for schedule coordination such as exchange servers exist, institutions involved in MIBCs often do not have a common intra-institutional platform and cannot apply their existing coordination systems to cross-institutional collaboration.

Access to Information

Participants often mentioned that they would like to have an easier access to the information associated with the collaboration. Information they needed was often spread out over multiple sites and could only be accessed from certain designated locations. Participants felt that not having to track down information from various places and being able to readily locate what they needed would help collaborative research.

Investigator G5 at GH mentioned that one of the most needed items on her wish list was an easier any time access to files she needed:

> *"I guess, the first thing that comes to mind would be a super easy mobile accessible anywhere of accessing any of the files any time I want, whether it's on my cell phone..."*

Project manager G6 at GH wished for a common administrative information resource:

"...having a place where people could just go...get their resources, download helpful information, resources and always know where they're going is most up to date information. Right now, that would be great."

Regardless of the type of information that needed to be accessed, the participants voiced a need for some way to easily access information at a common location. It did not matter to them what type of interface (i.e. web, central file system) they were given to the common information access technology.

Collaborative Writing Technology

A huge part of collaborative research involves collaborative writing. In MIBCs, collaborative writing involves researchers located at various institutions. Although a number of collaborative writing tools exist, participants felt that there were no satisfactory tools to facilitate the collaborative writing process. Participants at both of MIBCs voiced a wish for a good tool to help with the writing process. Project manager G7 at GH mentioned that she has heard of a tool that might be helpful, but GH did not currently have:

"I noticed some people are using google doc which is free software to do collaborative writing. And that's doing collaborative brainstorming...putting ideas down and other people can edit..."

Investigator T3 at TH felt her current method of collaborative writing was dysfunctional:

"...when we have one document that I've worked on and she's worked on at the same time. So now we have no idea who's changes are what. Or what's changed or what isn't and that can be frustrating so one way we try to get around that is use track changes or which we did originally is track changes. Now what we do is, it's yours and I'm not going to do anything to it until I get it back from you. Because we have

> *simultaneously worked on same document at the same time and that's horrible."*

Although there are already a number of existing collaborative writing tools, they are often either too expensive or not satisfactory. For example, using Microsoft Word documents with tracked changes often resulted in losses of edits by others. The coordination among multiple researchers using this tool to work on the same document was impossible. Participants voiced a wish for a better technology to assist collaborative writing.

6.4.4 Summary

Section 6.4 described the facilitators, barriers and needs associated with administrative processes in collaboration. The facilitators expressed were a collaborative infrastructure and support personnel. Most participants felt that remote collaboration could not happen without technology because due to the distance separating researchers most interactions had to be done through a technical medium. Administrative support was seen as a valuable asset to MIBC since support personnel removed the burden of unnecessary administrative tasks researchers had to perform. The barriers associated with administrative processes were information distribution through websites and schedule coordination. Participants felt that although technologies existed to support these two processes, they were either inadequate or not implemented properly due to institutional differences. The needs mentioned by participants included a common access to information and better collaborative writing technologies.

6.5 Finding Resources, Collaborators

As mentioned in Chapter 5.2.1.1, finding resources and collaborators is the first step to starting a collaborative project. Although it seems to be a simple task, due to the distance between researchers and the size of the organizations involved, finding collaborators and resources has an extra layer of complexity in MIBCs. It is, therefore, very important to support the processes that facilitate finding resources and collaborators.

6.5.1 Facilitators

Social Connections, Existing Relationships

Many collaborative projects start through existing relationships and affiliations. Participants at both MIBCs said that when a project idea forms, they often look for potential collaborators among their known network and affiliations first. They felt that they could trust and depend on their existing relationships. Therefore, participants felt that existing relationships were very important in finding collaborators and resources. Investigator G4 at GH mentioned how much easier it was to work with researchers she already had relationships with:

> *"...you know there should be a better betting process where everyone here has an opportunity, but the reality is you often call people who are your friends or you've worked together before or someone says, hey you should call so and so because she has an expertise in that area."*

Investigator T2 at TH thought social networking is what enabled collaboration to work:

> *"Uh, so you know it's one of those things really comes down to social network...I think...I mean how are you going to tell somebody oh he's a senior colleague, he's got lots of grant, he's got lots of publication, but he's not going to do any work for you, he's just going to stick his name on your stuff."*

Evidence of the importance of existing relationships was seen during observation sessions (see Figure 6.18). One of the participants of the observation session mentioned that he needed to rely on personal relationships to recruit investigators to use certain tools they have developed. Participants often turned to people they already knew and trusted to find collaborators and resources. They often did not want to take the risk of seeking out unknown researchers or resources.

PF: data warehouse project need pressure DB
Problem is once funding gone, no more interest, now need to rely on personal relationships]
SG: unsuccessful search for cancer researcher, posting <laugh>
Other site having some problems retaining investigators]
RR: some organizational changes of research director, maybe some separation of groups?
FT: really need one data warehouse deliverable

Figure 6.18
A sample observation note of a meeting at GH. The red circle indicates where participants talked of the importance of personal relationship.

Social Connectors, Matchmakers

Even if a researchers relies on existing connections, finding collaborators and resources can still be a difficult process. Many participants felt that the difficulty of finding collaborators can be alleviated by having a social connector, a "matchmaker" who can connect researchers together. These "matchmakers" have knowledge of the context (i.e. organizational structure, researchers and their backgrounds) and are well-known and trusted. They are thus able to help form collaborative projects by connecting researchers together as well as pointing them to appropriate resources.

Project manager G7 at GH mentioned that matchmakers are often needed to facilitate the initial stages of collaborative projects:

> *"...every week or so, someone will come up with concept of a proposal and I'm the liaison between all the investigators to do....I call researcher match.com... "*

Research assistant T7 at TH used an agency that helped her to find right resources

> *"...they kind of have a national collaboration of hospice agencies that are interested in research. So that was very helpful because I was able*

to e-mail and get in touch with them and see if they'd be willing to give me some contact people and they've been amazingly helpful..."

Researchers often faced difficulties of not knowing who among institutions to choose to start a new collaborative project or where to go to find proper resources. They felt a person or an agency that can help with these tasks would tremendously help researchers wishing to start a collaborative project.

Researcher Recruitment, Promotion

Collaborative research was sometimes limited by the lack of experts working in certain research areas within MIBC. Promoting the institutions within the MIBCs and recruiting researchers through more publications, lectures, and conferences were some of the ways that helped MIBCs to find new expertise. Researcher G5 at GH had to actively recruit researchers so that she could find appropriate expertise for her project. Without such recruitment some collaborative projects fail to happen:

"So there's a lot of...to call it, dog and pony show, just going out there being visible to the partners of community. D, my colleague down the hall might give a lecture at FH and do that as a way to, drum up interest and let people know we're out here and we have data resources, and essential numbers of cancer cases and then hopefully from there, ideas will flow. Hopefully..."

Evidence of the importance of recruitment and promotion was seen during observational sessions (see Figure 6.19). One of the participants in the meeting mentioned that there are several job postings for GH that have yet to be filled. It was often necessary to reach out to the community and actively recruit in order to find missing expertise within the collaboration. Participants felt that active recruitment was a way to alleviate the difficulties of finding the right expertise among a limited pool of researchers within the an organization.

relationships |
SG: unsuccessful search for cancer researcher, posting <laugh>
Other site having some problems retaining investigators |
RR: some organizational changes of research director, maybe some sepa
ET: really need one data warehouse deliverable
<all agree> |
Need feedback on my doc
Everyone: date? Will do
ET: admin stuff how to do things is what people want

Figure 6.19
A sample observation note of a meeting at GH. The red circle indicates where participants talked about researcher recruitment.

6.5.2 Barriers

Distance, Size of Collaboration

Due to the remote nature and large size of the MIBC, it was difficult for researchers to keep track of new opportunities, collaborators or resources available within the MIBC. Several participants mentioned that they often had no idea what types of resources were available to them or who they could connect with when looking to collaborate. Iinvestigator G2 at GH felt that not knowing what is going in a large collaboration was a primary difficulty investigators faced when trying to find collaborators or resources:

> *"...the question was how do you know if there's, you say I want to work with someone at place X, and I'm going to go find somebody there at place X, but the real problem is, I don't know that this might be a good opportunity because this might be somebody who thinks exactly as I am and they're a good place to work with...how do I know those kinds of things..."*

Investigator G4 at GH felt that investigators were often not well informed of opportunities and were unaware of other researchers in other institutions within the MIBC due to the distance separating them:

> *"I think there's a lot of work out there that you don't know what could potentially be used for preliminary studies for grant, um, and you just don't know data exist, you don't know what areas people worked in, that kind of thing."*

This lack of information was often due to the distance and size of the collaboration. Because MIBCs are composed of multiple large institutions located far apart, accidental or chance meetings among researchers were minimal. Thus, social interactions were limited between only those investigators who had developed relationships. Such limited social interactions made discovery of new collaborators and resources difficult.

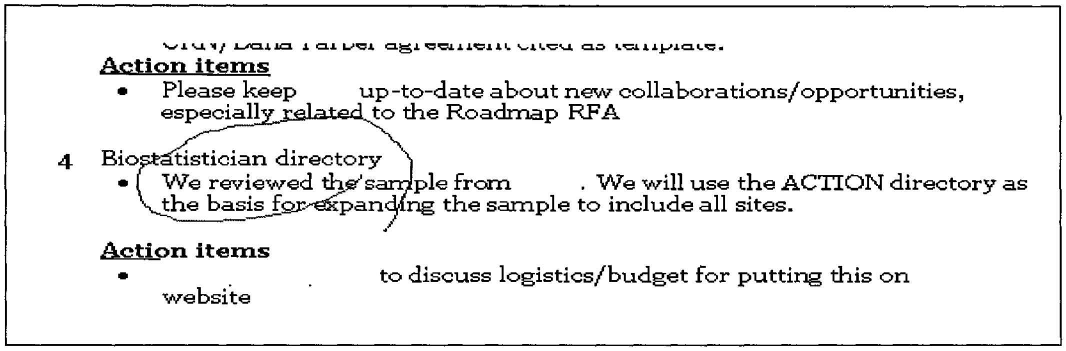
6.5.3 Needs

Online Directory, Information Resource

In Section 6.5.2, the difficulties MIBCs faced due to the distance and size of the collaboration were described. Participants at both MIBCs felt an online directory of researchers or online information on resources and opportunities might help alleviate some of these difficulties. Investigator G4 at GH mentioned that an online document that was frequently updated that she could use to find out about opportunities within the network would greatly help her in her collaborative efforts:

> *"....it's really hard to keep track of what everyone's doing, multiply that by 15 sites, not that everyone's as big as us and it's really hard to know who's working on what area and who has what expertise so having a big web document that would easily be updated would be wonderful, but you know, trying to organize that and stuff is not a small task."*

Evidence of the need for an online directory was often seen in the artifacts examined (see Figure 6.20 for a sample artifact). One of the action items during several meetings was the creation of an online directory. Although it was felt not to be effective within GH, there is a directory of information that was mentione during the observation sessions (see Figure 6.21). Although information about collaborators and resources exists, it is often scattered across many institutions' websites and is difficult to locate. Participants thought that a frequently updated single source of information about opportunities and collaborators would help them when initiating collaborative projects.



Action items

- Please keep up-to-date about new collaborations/opportunities, especially related to the Roadmap RFA

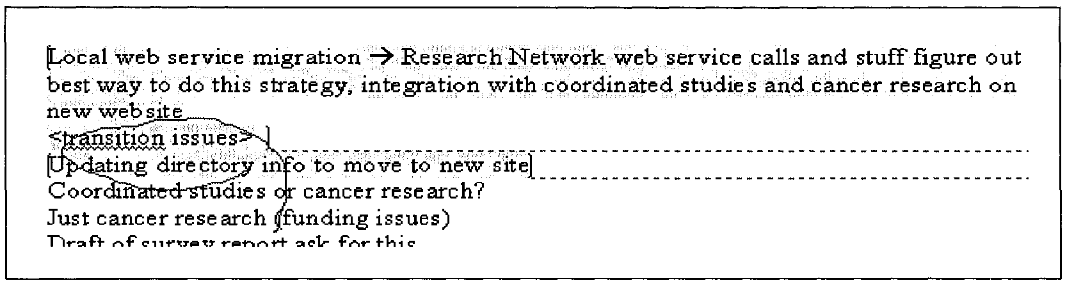
4 Biostatistician directory

- We reviewed the sample from . We will use the ACTION directory as the basis for expanding the sample to include all sites.

Action items

- to discuss logistics/budget for putting this on website

Figure 6.20
A sample artifact at GH discussing a directory of researchers.



Local web service migration → Research Network web service calls and stuff figure out best way to do this strategy, integration with coordinated studies and cancer research on new web site
<transition issues>
Updating directory info to move to new site
Coordinated studies or cancer research?
Just cancer research (funding issues)
Draft of survey report ask for this

Figure 6.21
A sample observation note taken during a meeting at GH. The red circle indicates participants discussing an online directory of researcher information.

6.5.4 Summary

Section 6.5 described the facilitators, barriers and needs associated with the processes of finding collaborators and locating resources. The facilitators included existing social connections, social connectors such as matchmakers between researchers, and researcher recruitment and promotion. Participants often turned to trusted sources such a current collaborators or institutional agencies to help them find reliable collaborators and resources. Institutions often needed to actively recruit outside researchers to have an available pool of researchers. The barriers that were most often discussed were the distance and size of MIBCs. Although a necessity, the remote nature of the collaboration contributed to the segregation of the institutions involved in MIBC. Information regarding existing collaborators and resources was also scattered across multiple institutions. The need mentioned by participants to address the distance barrier to locating collaborators and resources was an online directory and information resource that they could go to, to readily find the information they needed.

6.6 Summary

In Chapter 6, the core collaborative activities and processes (data, regulatory, communication, administrative, and finding collaborators and resources) that emerged from the field study were discussed in further detail. For each of these processes, associated facilitators, barriers, and needs were discussed (see Table 6.1). Unlike Chapter 5, where the overall characteristics of MIBCs were listed, this chapter focused only on those factors that are core to MIBCs that the biomedical informatics community can potentially address.

As a second step to addressing the question of "How can biomedical informatics best support current multi-institutional, interdisciplinary biomedical collaboration?" posed in Chapter 1, Section 2, this chapter contributes by providing an in-depth look at core activities and processes that are integral to every MIBC (see Chapter 1, Section 3, Aim 2). The facilitators are those factors that biomedical informatics should continue to support in current MIBCs if they are already in place, or should initiate, if not. The biomedical informatics community should study the barriers further and look for ways to alleviate them. As first step in research, they should turn to other fields such as Computer Supported Cooperative Work (CSCW) that have already studied collaborations in great detail. The needs mentioned by the participants of

MIBCs are the first items that biomedical informatics can tackle first to remove or alleviate some of the existing barriers. Chapter 7 will discuss future directions that biomedical informatics can take in support of MIBCs.

Chapter 7: Contributions, Future Work, and Conclusions

In this chapter, contributions this study has made to the field of biomedical informatics and to theory and methods in collaboration research are discussed in Section 7.1. Section 7.2 discusses possible future directions the study can take to further support collaborative efforts in biomedical research. Section 7.3 closes chapter with concluding remarks.

7.1 Contributions

7.1.1 Contributions to the Biomedical Informatics Field

Characterization of Multi-institutional Interdisciplinary Biomedical Collaboration

To date, very little has been known about multi-institutional interdisciplinary biomedical collaboration. The research question posed in Chapter 1 was, "How can biomedical informatics best support current multi-institutional interdisciplinary biomedical collaboration?" As a first step to addressing the question, this research identifies general characteristics of Multi-institutional Interdisciplinary Biomedical Collaboration (MIBC) through a contextual field study. Chapter 5 describes these characteristics in detail. As described, MIBCs are composed of core activities and processes that happen throughout the life-cycle of collaboration, social and socio-technical factors that impact the overall collaboration and external forces that also influence collaboration but cannot be controlled from within (See Chapter 5, Figure 1 and Table 1 below). Some of these characteristics are similar for both MIBCs and other collaborative settings, in support of which there already exists a large body of research in other fields (Finholt 2003; Olson 2005). However, some of the processes related to data and regulatory elements of collaboration are unique to MIBCs. The biomedical informatics community should pay particular attention to these processes. In addition to core activities and processes, there are also many social and socio-technical issues associated with MIBCs. Although the biomedical informatics community alone might not be able to address themcompletely, these are factors that must be taken into account as they try to support Multi-institutional Interdisciplinary Biomedical Collaboration (MIBC).

Identification of Facilitators, Barriers, and Needs

As a second step to addressing the research question, this research takes a more in-depth look at the core activities and processes that are integral to MIBCs. The core collaborative activities and processes include: a) data, b) regulatory, c) communication, d) administrative, and e) finding collaborators, and resources. Chapter 6 describes them in detail with associated facilitators, barriers, and needs. These core activities and processes are those that the biomedical informatics community can potentially address to support MIBCs. The facilitators are factors that currently support or help in the collaborative process. Biomedical informatics should continue to support facilitators in current MIBCs if they are already in place, or should look for ways to implement them, if not already in place. The barriers are factors that negatively influence collaboration. These must be studied further and ways to alleviate these difficulties should be found. The needs are the first items that the biomedical informatics can focus on in order to remove some of the existing barriers. Ways to alleviate difficulties in complex collaborative research can be looked for from both within and outside the field of biomedical informatics. As described in Chapters 2 and 3, a large body of knowledge about collaboration already exists in fields such as Computer Science (CS) and Information Science (IS) and in the sub area of Computer Supported Cooperative Work (CSCW). The biomedical informatics community should leverage some of the existing research in collaboration.

The Framework for Multi-institutional Interdisciplinary Biomedical Collaboration (fMIBC)

Although general collaborative frameworks such as TORC already existed, no biomedical context-specific framework existed to support MIBCs. Therefore, a framework for multi-institutional interdisciplinary biomedical collaboration (fMIBC) was developed through the contextual study described in Chapter 4 (to address Aim 3 in Chapter 1, Section 3). The fMIBC is represented graphically (see Figure 7.1) and the factors internal to MIBCs are looked at in further detailed in Table 7.1. It lists all the factors that exist within fMIBC and adds to Table 6.1. The fMIBC is grounded on the contextual field research that was guided by the preliminary framework built upon existing research on collaboration.

In this framework, federal and institutional factors external to collaboration were grayed out because they are not issues that the biomedical informatics community can readily influence. The factors internal to collaboration are composed of social and socio-technical issues, and core collaborative activities and processes. Social and socio-technical issues are factors that affect overall collaboration. Data, regulatory, administrative, communication, and location of resources and collaborators are identified as core collaborative activities and processes.

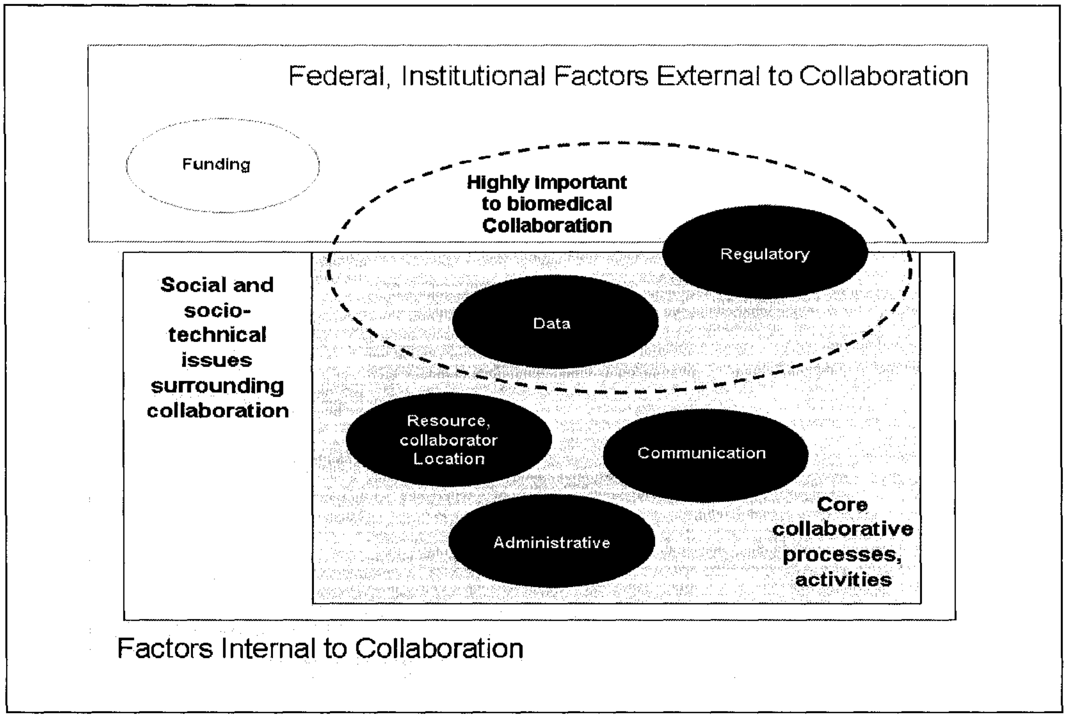


Figure 7.1
A graphical representation of framework of Multi-institutional Interdisciplinary Biomedical Collaboration (fMIBC).

Table 7.1
Tabular and categorical representation of the fMIBC corresponding to the graphical representation above.

Category			Description
Ext, Int	Core	REG	Federal regulations, Institutional review board
Ext, Int	Core	REG	Common regulatory process
Int	Core	DAT	Data warehouse, repository
Int	Core	DAT	File, data transfer
Int	Core	DAT	Data standards
Int	Core	DAT	Data security protocol
Int	Core	DAT	Project specific data, various data purposes
Int	Core	COM	Face-to-face meetings: local, conference
Int	Core	COM	Email: asynchronous communication
Int	Core	COM	Phone, conference calls: synchronous communication
Int	Core	COM	Communication technology limitation
Int	Core	COM	Technology enabling face-to-face: video conferencing
Int	Core	ADM	Collaborative technology, infrastructure
Int	Core	ADM	Support personnel help
Int	Core	ADM	Information distribution: websites, online technology
Int	Core	ADM	Schedule coordination
Int	Core	ADM	Access to information
Int	Core	ADM	Collaborative writing technology
Int	Core	FRC	Social connections, existing relationships
Int	Core	FRC	Social connectors
Int	Core	FRC	Researcher recruitment, Promotion
Int	Core	FRC	Distance, size of collaboration
Int	Core	FRC	Online directory, information resource
Int	SSt		Developed relationship, trust
Int	SSt		Common Ground
Int	SSt		Contextual differences
Int	SSt		Incentive for collaboration
Int	SSt		Willingness
Int	SSt		Time consuming, lot of work
Int	SSt		Un-needed technology
Int	SSt		Involvement, contribution
Int	SSt		Insider vs. outsider

Legend
Ext - Factors external to collaboration
Int - Factors Internal to collaboration
Core - Core collaborative activities/processes
SSt - Social and Socio-technical issues
COM - Communication
ADM - Administrative
DAT - Data
FRC - Resource, collaborator location
REG - Regulatory

A Checklist for the Evaluation of MIBC Infrastructure

The fMIBC can be used to inform the design and evaluation of informatics infrastructure to better support MIBC. For example, the components of core processes and activities in fMIBC can be readily turned into an evaluation checklist. One example of such a checklist might be a list of the minimal set of infrastructure requirements for MIBC (see Table 7.2 and Table 7.3 for sample evaluation checklists). The evaluation checklist below is composed of the minimal set of infrastructure requirements that MIBC needs to satisfy at the level of individual institutions (Table 7.2) and at the level of the MIBC as a whole (Table 7.3). The "must have" items in the checklist need to be present in order for a MIBC to form and function successfully. Factors in the desirables column are those that might not be essential for collaboration to form and continue, but might help save a great deal of time and effort if they are present. This is a directly useful tool that can guide the evaluation of any MIBC.

Table 7.2
A sample infrastructure checklist for multi-institutional interdisciplinary biomedical collaboration infrastructure for individual institutions

Must haves	Examples	Desirables
Asynchronous communication	**Electronic mail, discussion or message board** Researchers must be able to continue to communicate regardless of the availability of a person on the other end at any particular time. Especially in remote collaboration, the ability to asynchronously communicate is important due to time zone differences and distance separating researchers.	

Table 7.2 (continued)

Synchronous communication	**Phone, cell phone, online chat, messenger** Researchers at remote locations are often unable to conduct meetings in person. Synchronous communication tools enable researchers to discuss research with each other in real time.	**Videophone, videochat** Although synchronous communication can be accomplished with voice or text-only tools, having the ability to see each other can greatly enhance the communication experience. Non-verbal cues can only be obtained through mediums such as video.
	Face-to-face meetings Face-to-face meetings are indispensible even though much of the communication in remote collaboration occurs via communication technology	
Data management	**Databases, file servers, spreadsheets** Biomedical research generates a large amount of data. Tools to effectively manage data should be provided.	**Customized data management tools** Although basic spreadsheets and databases can minimally meet needs of the researchers, having a data management system that is customized for the need can benefit researchers.
Data security	**De-identification protocols** Before data use, there must be at minimum a data de-identification protocol that complies with federal regulations.	**Encryption, firewall, authentication** Having more comprehensive encryption and authentication protocols can add an additional level of data security.
Schedule coordination	**Personal calendar (paper or electronic)** Having a personal calendar, paper or electronic enables researchers to coordinate their schedules.	**Microsoft exchange, other common calendaring tools** Having a common calendaring system better facilitates schedule coordination (i.e. project managers can set meetings without having to ask every individual for their availability).

Table 7.2 (continued)

Support personnel	**Project managers** Project managers ensure that the overall collaboration is running smoothly and removes some of the administrative burden from researchers.	**Technical support** Remote collaborations require researchers to use many technical tools. Technical support can aid researchers through this process.
Regulatory structure	**Local IRB approval** By federal regulation, any biomedical research involving human subjects must be reviewed by an institutional review board.	

Table 7.3
A sample infrastructure checklist for multi-institutional interdisciplinary biomedical collaboration infrastructure for the MIBC as a whole

Must haves	Examples	Desirables
Synchronous communication tools	**Phone conferencing, online chat** Phone conference enables meetings among collaborators located remotely.	**Video conference** Although remote meetings can be accomplished with voice or text-only tools, having the ability to see eah other can greatly enhance the user experience. Non-verbal cues can only be obtained through a medium such as video.
	Annual conferences, meetings Face-to-face meetings are indispensible. Regular face-to-face meetings and/or annual conferences help researchers get more face-to-face time.	
Data storage	**Data repositories, data warehouses** Biomedical research generates a large amount of data. Tools to effectively manage and store data should be provided.	

Table 7.3 (continued)

Data sharing	**Data standards across institutions** Institutions involved in collaboration often have an institution-specific way of storing data. In order for data sharing to occur, at least some portion data must meet standards so that data across multiple institutions can be aggregated and shared.	**Standardized data across institution** At minimal, data sharing can occur by standardizing portions of already existing data; however, having a standardized format across all institution facilitates more ubiquitous sharing.
Data transfer	**Electronic mail, postal mail** Since data is de-identified at the institutional level, sending data through electronic or postal mail can be a way to minimally support data transfer.	**Secure data transfer tools** Even though data is de-identified at the institutional level, securely transferring data between institutions through technology can greatly aid collaborative research.
Information distribution	**Listserv** Much information (i.e. administrative, news, etc.) need to be distributed as a part of collaboration. Listserv can help distribute information across multiple institutions.	**Website resource** A well-designed and easy to use website with resources and information can better facilitate information dissemination across multiple institutions.
Schedule coordination	**Calendaring tools** Calendaring tools can minimally help researchers schedule meetings.	**Common calendaring system across institutions** A common calendaring system across institutions can help make schedule coordination faster and easier.
Collaborative writing	**Word documents** A minimal level of collaborative writing can be supported through tracking changes in Microsoft Word documents.	**Collaborative writing tools** Collaborative writing technologies that enables tracking changes and support simultaneous edits can greatly benefit the collaborative writing process.
Regulatory Structure	**IRBs working together** Multi-institutional collaboration involves approval from multiple IRBs. IRBs must work to get a collaborative project approved.	**Common inter-institutional IRB** Whether IRBs work together or not, multiple IRB is time consuming and difficult. Finding a way to have a common inter-institutional IRB would greatly aid collaborative research

Table 7.3 (continued)

Collaborator and resource locator	**Web resource of researchers, opportunities across collaboration** Web resource of researchers and opportunities help researchers locate potential collaborators and opportunities	**Social connectors** Although a web-based information source can minimally help researchers connect with collaborators and find resources, researchers often prefer and knowledgeable social connectors to get information.

7.1.2 Contribution to Theory

Verification of the Theory of Remote Collaboration (TORC)

The Theory of Remote Collaboration (TORC) described in detail in Chapter 2.2.1 is based on an evaluation of prior collaboratory studies and a survey about remote collaboration (Olson 2005). The TORC was developed as a theoretical framework to guide general research about collaboration; however, it has yet to be verified in a real collaborative setting. This research began with a comprehensive literature review, described in Chapter 3, which developed into a preliminary framework that extends the TORC into the context of the biomedical setting. Through the contextual field research, the preliminary framework was refined and verified as a framework for multi-institutional interdisciplinary biomedical collaboration (fMBIC) (see Section 7.1.1, Figure 7.1, and Table 7.1). Because this preliminary framework extended TORC, some of the concepts in TORC were also verified (see Table 7.2, detailed description available in Chapter 6).

The TORC and fMIBC align well in many areas. Although each collaboration is unique, there are some underlying qualities that can be generalized across many collaborations. Table 7.2 maps fMIBC concepts to TORC concepts. There are fMIBC concepts that map directly to TORC and fMIBC concepts that are implied in TORC. The concepts that map directly to concepts in TORC are more general and social in nature (see Table 7.2 under fMIBC Concepts That Map Directly to TORC). These are collaborative concepts such as trust, common ground, and support for collaborative processes that are generally applicable to all collaborations. The fMIBC concepts that are implied in TORC are more specific and technical in nature (see Table 7.2 under column fMIBC Concepts That are Implied in TORC). These are specific technologies such as communication and data sharing tools that support collaboration.

Table 7.4
Verification of TORC by mapping with fMIBC developed through field data.

TORC Concepts		**fMIBC Concepts That Map Directly to TORC**		**fMIBC Concepts That are Implied in TORC**	
Nature of the Work	- Understanding of work - Tightly vs. loosely-coupled work	FRC	Distance, size of collaboration	SOC	Insider vs. Outsider
Common Ground	- Mutual knowledge - Common vocabulary - Past relationships - Existing common ground - Common work style	SOC SOC SOC	Developed relationship, trust Common ground Contextual differences	SOC	Insider vs. Outsider
Collaboration Readiness	- Motivation: monetary, skill - Benefit for all participants - Trust or contractual agreement - Goal alignment	SOC	Common ground	SOC	Willingness
		SOC SOC FRC	Incentive for collaboration Developed relationship, trust Social connections		

Table 7.4 (continued)

Management, Planning, Decision Making	- Time and commitment - Face-to-face time - Sense of contribution - Critical mass at each site - Project management - Being informed - Communication plan - Oversight, advisory committee - Policy compromises - Data management plan - Fair decision making	SOC	Time consuming, a lot of work	COM	Asynchronous communication
		COM	Face-to-face meetings	COM	Synchronous communication
		SOC	Involvement, contribution	ADM	Information distribution technology
		ADM	Support personnel help	ADM	Schedule coordination
				ADM	Access to information
Technology Readiness	- Adoption of tools - Appropriate technology - Beneficial technology - Interoperability - Technical support - Data standards	SOC	Un-needed technology	COM	Asynchronous communication
		DAT	Data standards	COM	Synchronous communication
		ADM	Support personnel help	COM	Video conferencing
				ADM	Information distribution technology
				ADM	Collaborative writing technology
				DAT	Data warehouse, repository
				DAT	File, data transfer
				FRC	Online resource

Extension of the Theory of Remote Collaboration (TORC)

The concepts in Table 7.3 (detailed descriptions are available in Chapter 6) are those that were not included in TORC that extend the TORC in the context of the biomedical field. The majority of concepts in Table 7.3 are regulatory in nature. As described in Chapter 3, Section 2.2, unlike collaborations in other fields, biomedical research collaborations are highly controlled by federally mandated regulations (i.e.

HIPAA). Although not a direct regulatory element, data security protocols are also related to the regulatory process (as described in Chapter 5 Section 2.1). Since the regulatory aspect is pervasive in all biomedical collaborations, these collaborative elements are novel extension to the TORC in the context of the biomedical field.

Social connectors (described in Chapter 6, Section 5.2) is a concept especially important in MIBCs since as mentioned in Chapter 1, Section 1, biomedical researchers are highly competitive and MIBCs involve researchers from very different backgrounds. Many researchers do not know who in other fields to connect with, even if they have desire and reason to collaborate with another researcher with different expertise. This competitiveness also hinders biomedical researchers from readily seeking out new collaborators. Rather, they would be much more likely to collaborate if they had a trusted social connector who referred them to the other researcher.

Table 7.5
fMIBC concepts that extend the TORC to the context of the biomedical domain

fMIBC Concepts Not in TORC (fMIBC Extension to TORC)	REG	Institutional review board
	REG	Common regulatory process
	DAT	Data security protocol
	FRC	Social connectors

7.1.3 Contribution to Methodology

Use of Qualitative Methods in Studying the Collaborative Biomedical Setting

More and more researchers have been using qualitative methods in the biomedical setting since Diana Forsythe's foundational works using qualitative methods in this field (Forsythe 1992; Forsythe 1996; Forsythe 1998). Yet the use of qualitative methods in biomedical research has been very narrow in scope. Thus far, the qualitative methods in biomedical field have mostly been used to study implementation of a single system or a tool (Ash 2003; Ash 2005; Anderson in Press), or to develop a single protocol for a system design (Bartlett 2002; Bartlett 2005). This study was one of the first studies to broadly use qualitative methods to characterize a large complex collaborative settings in the biomedical field. The framework resulting from this study was able to extend an existing theory and aid in the design and development of an informatics infrastructure. The result of this study illustrates that the use of qualitative methods is a valid and practical approach to inform system design and development of theory to describe a complex biomedical environment.

7.2 Limitations

This study has limitations that are inherent to any qualitative study. Field studies are best used to capture complex settings; however, the broad nature of MIBCs made it difficult for this study to gain more than a localized and constrained view of MIBCs. Nevertheless, as described in Chapter 4.3.8, this study undertook steps to minimize the limitations by closely following the methodological steps in Chapter 4.2.2 and 4.2.4. Multiple sources of data (interviews, observations, artifacts) were used to triangulate and get a better understanding of the settings. Multiple sites were studied to obtain results that are applicable to a broader range of settings.

A single researcher also lacks resources and time required to properly study the multi-institutional setting. Due to the limited time in the field, this study does not attempt to answer the questions of how relationships among themes make a given MIBC better. Although the framework developed through this study is able to show relationships among concepts related to MIBCs, only a longitudinal study that examines a full set of MIBCs are able to establish what factors make MIBCs more or less effective. Such longitudinal study is beyond the scope of this research.

7.3 Future work

7.3.1 Refinement of the Framework for Multi-institutional Interdisciplinary Biomedical Collaboration (fMIBC)

The fMIBC is based on a pool of existing research on collaboration and through this research was refined and verified in a real context. However, MIBCs are also complex social settings. No single MIBC is exactly like another and MIBCs like the studies they are involved in are constantly changing. In order to support such a diverse and changing setting, the fMIBC should be constantly refined through re-evaluations in real settings. One of the future extensions to this study is to continuously gather and evaluate needs and barriers of various existing MIBCs and refine fMIBC as MIBCs evolve. Although generalizing such diverse and changing settings into a single framework might be difficult, through continuous study and refinement, commonalities can be gathered into a single framework that will guide the how biomedical informatics' supports MIBCs. The resulting framework will not only help evaluate the existing MIBCs, but also assist initial formation and continuation of MIBCs.

7.3.2 Ways to Facilitate Inter-institutional Processes

Inter-institutional Processes

Inter-institutional processes occur when institutions engage in a collaborative project where each institution has its own institutional processes that need to be reconciled across multiple institutions. For example, every institution involved in MIBC has its own regulatory practice as described in Chapter 6 Section 2. Federal regulations (i.e. HIPAA) require that an institutional review board (IRB) reviews, approves and monitors any biomedical research involving humans (HIPAA). To comply with federal regulations, each institution has its own IRB that monitors this regulatory process.

One of the main barriers in MIBCs is in the difficulty of reconciling differences in institutional practices to better facilitate inter-institutional processes. For example, due to each institution having its own IRB, multi-institutional collaborative projects need approval from multiple IRBs. As described in Chapter 6 Section 2.2, in order to proceed with a multi-site collaborative project, a separate approval has to be received from each IRB. Getting a research project approved by even a single IRB is not a small task. Facing multiple IRBs, each with its own interpretation of federal regulations, poses a great a difficulty for the researchers involved in MIBCs.

Yet every institutional IRB is regulated by common federal regulation. The National Institutes of Health (NIH)'s information on IRB and HIPAA privacy rule (NIH HIPAA) states that "an IRB must determine that specified criteria have been satisfied". Thus, although interpreted slightly differently by each, each IRB must follow common HIPAA regulations and standards (HIPAA). The HIPAA regulations and standards explicitly states the privacy rule, HIPAA statute, security rule, and standards that each institutional IRB must follow. Thus, the IRB process for MIBCs is composed of multiple individual IRBs all operating under common regulatory guidelines, but with each IRB having its own institutional specific differences.

Dynamic Workflow System

Workflow research initially started with the problem of office automation, attempting to model the flow of work through an organization to support collaborative activity (Nutt 1996). However, first workflow systems were criticized for their rigidity and lack of ability to tolerate changes and exceptions associated with the actual work processes (Grudin 1994). Abandoning the rigid workflow model completely, the Computer Supported Cooperative Work (CSCW) placed emphasis on the importance of real life practices. CSCW research led to numerous groupware (i.e. video conferencing, collaborative desktops) to support real-life, often unstructured collaborative activities. However, due to the ad hoc nature of these tools, modeling, identifying exceptions and managing evolution of the processes became impossible (Kammer). Later workflow research, influenced by the concepts from both views, resulted in the concept of dynamic workflow. Dynamic workflow systems encapsulate processes within an organization where each encapsulated process is composed of a set of sub-processes with explicit exceptions to the main processs (Nutt 1996). Unlike traditional workflow systems, dynamic workflow systems are flexible and are able to support changes and unexpected occurrences within the processes they model (Kammer).

Encapsulating Inter-institutional Process as Dynamic Workflow System

As described in Chapter 6, Section 2, one of the main difficulties MIBCs face is the difficult IRB process that is composed of multiple institutional IRBs. Many participants desired a common system that incorporates multiple institutional IRBs into one IRB system. One of the ways to combine multiple institutional IRB into a single IRB system is to develop a dynamic workflow system for the inter-institutional IRB process. The IRB process in MIBC is composed of a large common set of rules determined by federal regulation that must be followed by institutions along with institution-specific differences in interpretation. The inter-institutional IRB therefore, lends itself well to being encapsulated into a dynamic workflow system. The large common set of rules that each institutional IRB must follow can be modeled into main workflow processes and the institutional specific practices can be modeled as exceptions to the main workflow. One of the future research directions stemming from this study is to identify the main steps of the inter-institutional IRB processes (rules) that are common across all individual institutional IRBs and also to identify differences (exceptions). The common rules and exceptions can then be used to design a dynamic workflow system (a common inter-institutional IRB system across multiple institutions) that can automate an inter-institutional IRB process for MIBC. Such system can help save a large amount of time and effort and financial resources for those involved in MIBCs.

7.4 Concluding Remarks

The results of this research do not provide solutions for all the difficulties that exist in MIBCs nor do they provide a single magical formula for how to create a successful MIBC. This study is only the first few steps towards answering the question, "how can biomedical informatics better support the multi-institutional interdisciplinary biomedical collaboration?" In order to alleviate the difficulties of the MIBC, the biomedical informatics community must continue to study and improve upon new and existing research. Furthermore, biomedical informatics should get involved early in the process of MIBC formation and continue to be involved at every step. During the initial planning and implementation, biomedical informatics should set up a collaborative infrastructure that fits the characteristics of the collaboration. Throughout the collaborative process, biomedical informatics should continuously support, evaluate, and make appropriate adjustments to the ever-changing needs and technologies of the MIBC. Only through these efforts can the biomedical informatics community help alleviate the difficulties current MIBCs face and help these and new collaborations to discover answers to the extremely important and extremely complex research questions they strive to answer.

Bibliography

Ackerman, M. S., and McDonald, D. W. (2000). "Collaborative support for informal information in collective memory systems." Information Systems Frontiers **2**(3-4): 333–347.

Addis, M., Ferris, J., Greenwood, M., Li, P., Marvin, D., Oinn, T., and Wipat, A. (2003). Experiences with e-Science work-flow specification and enactment in bioinformatics. In Proceedings of UK e-Science All Hands Meeting 2003, Nottingham, UK.

Allison, D. B., Cui, X., Page, G.P., and Sabripou, M. (2000). "Microarray data analysis: from disarray to consolidation and consensus." Nat Rev Genet. 2006: 55-65.

Altman, R. (2006). "Share and Share Alike: A proposed set of guidelines for both data and software." Biomed Comp Rev. **1**.

Altman, R. B., and Klein, T.E. (2002). "Challenges for biomedical informatics and pharmacogenomics." Annu Rev Pharmacol Toxicol. **42**: 113-33.

Anderson, N., Ash, J., and Tarczy-Hornoch, P. A. (in Press). "Qualitative Study of the Implementation of a Bioinformatics Tool in a Biological Research Laboratory." International Journal of Biomedical Informatics.

Appelt, W. (1999). WWW Based Collaboration with the BSCW System. Proceedings of SOFSEM'99, Milovy (Czech Republic), Springer Lecture Notes in Computer Science 1725.

Arnstein, L., Borriello, G., Consolvo, S., Hung, C., and Su, J. (2002). "Labscape: A Smart Environment for the Cell Biology Laboratory." IEEE Pervasive Computing Magazine **1**(3): 13-21.

Arnstein, L. F., Grimm, R., Hung, C, Hee, J., LaMarca, A., Sigurdsson, S. B., Su, J., and Borriello, G. (2002). Systems Support for Ubiquitous Computing: A Case Study of Two Implementations of Labscape. Proceedings of the First International Conference on Pervasive Computing, Springer-Verlag, Germany.

Ash, J. S., Fournier, L., Stavri, P.Z., and Dykstra, R. (2003). Principles for a successful computerized physician order entry implementation. AMIA Annu Symp Proc.

Ash, J. S., Gorman, P. N., Lavelle, M., Payne, T. H., Massaro, T. A., Frantz, G. L., and Lyman, J. A. (2003). "A cross-site Qualitative study of Physician Order Entry." Journal of American Medical Informatics Association **10**(2): 188- 200.

Ash, J. S., Sittig, D.F., Seshadri, V., Dykstra, R.H., Carpenter, J.D., and Stavri, P.Z. (2005). "Adding insight: a qualitative cross-site study of physician order entry." Int J Med Inform **74**(7-8): 623-628.

Atlas Ti. "Atlas Qualitative Analysis Software." from http://www.atlasti.com/index.php.

Bafoutsou, G., and Mentzas, G. (2002). "Review and functional classification of collaborative systems." International Journal of Information Management **22**(4): 281-303.

Baker, K., and Bowker, G. (2004). "Information Ecology: Open System Environment for Data, Memories and Knowing." Journal of Intelligent Information Systems **BDEI Special Series.**

Baker, K. S. (2004a). Ecological design: an interdisciplinary, interactive participation process in an information environment. Proceedings of the workshop on Requirements Capture for Collaboration in e- Science, Edinburgh, UK.

Baker, K. S., Benson, B. J., Henshaw, D. L., Blodgett, D., Porter, J. H., and Stafford, S. G. (2000). "Evolution of a Multisite Network Information System: The LTER Information Management Paradigm." BioScience **50**(11): 963-978.

Baker, K. S., Jackson, S.J., and Wanetick, J.R. (2005). Strategies Supporting Heterogeneous Data and Interdisciplinary Collaboration: Towards an Ocean Informatics Environment. Proceedings of the 38th Hawaii International Conference on System Science. HICSS38, Big Island, HI, US.

Ball, C. A., Sherlock, G., and Brazma, A. (2004). "Funding high-throughput data sharing." NatureBiotechnology **22**(9): 1179-1183.

Bannon, L., and Bodker, S. (1997). Constructing Common Information Space. Proceedings of the European Conference on Computer-Supported Cooperative Work ECSCW'97, Lancaster, UK, Kluwer Academic Publishers.

Barga, R. S., Andrews, S. and Parastatidis, S. (2007). A Virtual Research Environment (VRE) for Bioscience Researchers. International Conference on Advanced Engineering Computing and Applications in Sciences, 2007. ADVCOMP 2007..

Bartlett, J. (2002). The oral tradition of bioinformatics expertise. Proceedings of the 65th annual meeting of the American Society for Information Science and Technology.

Bartlett, J. C., and Toms, E. G. (2005). "Developing a protocol for bioinformatics analysis: an integrated information behavior and task analysis approach." Journal of the American Society for Information Science and Technology **56**(5): 469–482.

Bernard, H. R. (1994). Research Methods in Anthropology: Qualitative and Quantitative Approaches, 2nd Ed. . WalnutCreek, CA, AltaMira Press.

Beyer, H., and Holtzblatt, K. (1998). Contextual Design: Defining Customer-Centered Systems Morgan Kaufmann Publishers.

Bhandarkar, M., Budescu, G., Humphrey, W., Izaguirre, J. A.m Izrailev, S., Kalé, L. V., Kosztin, D., Molnar, F., Phillips, J. C. and Schulten, K. (1999). BioCoRE: a collaboratory for structural biology. In Proceedings of theSCS International Conference on Web-Based Modeling and Simulation. A. G. Bruzzone, A. Uchrmacher, and E. H. Page, editors. San Francisco, CA.

BioCoRE. "Biological Collaborative Environment,." from http://www.ks.uiuc.edu/Research/biocore.

BIRN. "Biomedical Informatics Research Network." from http://www.nbirn.net/.

Birnholtz, J. P., and Bietz, M.J. (2003). Data at work: supporting sharing in science and engineering. Proceedings ofthe 2003 International ACM SIGGROUP Conference on Supporting Group Work (GROUP'03) Sanibel Island, FL, US.

Birnholtz, J. P., and Bietz, M.J. (2005). When do researchers collaborate? Toward a model of collaboration propensity in science and engineering research. Ann Arbor, MI, University of Michigan.

Bly, S. K., K.M., and Henline, P.A. (1997). The Work of Scientists and The Building of Collaboratories. Proceedings of the Group 97 International Conference on Supporting Group work Phoenix, AZ, US, General Atomics Report GA-A22619.

Bødker, S., Grønbæk, K. and Kyng, M. (1993). Cooperative Design: Techniques and Experiences from the Scandinavian Scene. Participatory Design: Principles and Practices. . Hillsdale, NJ, Lawrence Erlbaum Associates: 157 - 175.

BSC. "Biological Sciences Collaboratory." from http://www.pnl.gov/biology.

BSCW. "Basic Support for Cooperative Work." from http://bscw.fit.fraunhofer.de/.

Buetow, K. (2005). "Cyberinfrastructure: Empowering a "Third Way" in Biomedical Research." Science **308**(5723): 821-824.

caBIG. "Cancer Biomedical Informatics Grid." from http://cabig.cancer.gov/.

Campbell, E. G., Clarridge, B.R., Gokhale, M., Birenbaum, L., Hilgartner, S., Holtzman, N.A. and Blumenthal, D. (2002). "Data Withholding in Academic Genetics: Evidence From a National Survey." Journal of the American Medical Association **287**: 473-480.

Cao, J., Jarvis, S.A., Saini, S. and Nudd, G.R. (2003). GridFlow: Workflow Management for Grid Computing. Proceedings of the 3rd IEEE/ACM International Symposium on Cluster Computing and the Grid (CCGRID' 03).

Casper, T. A., Meyer, W. M., and Moller, J. M. (1998). "Collaboratory Operations in Magnetic Fusion Scientific Research." Interactions **3**(56): 56-64.

Chin, G., and Lansing, C. S. (2004). Capturing and supporting contexts for scientific data sharing via the biological sciences collaboratory. Proceedings of the 2004 ACM conference on Computer supported cooperative work, Chicago, IL.

Chin, G., Myers, J., and Hoyt, D. (2002). "Social networks in the virtual science laboratory." Communications of the ACM **45**(8): 87-92.

Consolvo, S., Arnstein, L., and Franza, B.R. (2002). User Study Techniques in the Design and Evaluation of a Ubicomp Environment. Proceedings of the Fourth International Conference on Ubiquitous Computing.

Crabtree, B., and Miller, W. (1999). Doing Qualitative Research. 2nd ed, Sage Publications.

CSIEDG. "Commission on Systematic Interoperability: Ending the Document Game, ." from http://endingthedocumentgame.gov.

CTSA RFA. "Clinical and Translational Science Awards RFA." from http://www.ncrr.nih.gov/clinicaldiscipline.asp.

Cummings, J. N., and Kiesler, S. (2005). "Collaborative research across disciplinary and institutional boundaries." Social Studies in Science (to appear) **35**(5): 703-722.

David, P., and Spence, M. (2003). Towards Institutional Infrastructures for e-Science: The Scope of the Challenge, Final Report of the Oxford Internet Institute project on "The Institutional Infrastructure of e-Science: The Scope of the Issues."

Denzin, N., and Lincoln, Y. (1994). Handbook of Qualitative Research. Thousand Oaks, CA, Sage.

Dourish, P., and Bellotti, V. (1992). Awareness and coordination in shared workspaces. In Proceedings of CSCW'92, Toronto, ACM Press.

EMSL. "Environmental Molecular Sciences Laboratory, ." from http://collaboratory.emsl.pnl.gov.

Finholt, T. A. (2002). Collaboratories: Science over the Internet. AAAS Science and Technology Policy Yearbook. In Teich, Albert H., Nelson, Stephen D., and Lita, Stephen J. (editors), Washington, DC: American Association for the Advancement of Science: 339-344.

Finholt, T. A. (2003). "Collaboratories as a new form of scientific organization." Econ. Innov. New Techn. **12**(1): 5–25.

Finholt, T. A., and Olson, G. M. (1997). "From Laboratories to Collaboratories: A New Organizational Form for Scientific Collaboration." Psychological Science **8**(1): 28-36.

Fisher, D., and Dourish, P. (2004). Social and temporalstructures in everyday collaboration. In Proceedings of the SIGCHI conference on Human factors in computing systems '04, Vienna, Austria.

Forsythe, D. (1998). "An anthropologist's viewpoint: observations and commentary regarding "Implementation of nursing vocabularies in computer-based systems"." J Am Med Inform Assoc **5**(4): 329-331.

Forsythe, D. E. (1992). Using ethnography to build a working system: rethinking basic design assumptions Proc Annu Symp Comput Appl Med Care.

Forsythe, D. E. (1996). "New bottles, old wine: hidden cultural assumptions in a computerized explanation system for migraine sufferers." Med Anthropol Q **10**(4): 551-574.

Foster, I., Kesselman, C., Nick, J. M., and Tuecke, S. (2002). "Grid Services for Distributed system integration." IEEE Computer: 37-46.

Foster, I., Vockler, J., Wilde, M., and Zhao, Y. (2003). The Virtual Data Grid: ANew Model and Architecture for Data-Intensive Collaboration. Proceedings of the First CIDR - Biennial Conference on Innovative DataSystems Research, Asilomar, CA.

Glaser, B. (2001). The Grounded Theory Perspective I: Conceptualization Contrasted with Description Mill Valley Sociology Press.

Glaser, B. G., and Strauss, A.L. (1967). The Discovery of Grounded Theory: Strategies for Qualitative Research. New York, NY, Aldine

Google scholar. "Google Scholar." from http://scholar.google.com.

Grudin, J. (1994). "Computer-Supported Cooperative Work: History and Focus." IEEE Computer **May 1994**: 19-26.

Hagstrom, W. O. (1974). "Competition in Science." American Sociological Review **39**(1): 1-18.

Halkola, E. (2004). Scientific collaboration and information infrastructures – Information practices in LIDETexperiment.

Hara, N., Solomon, P., Kim, S. L. and Sonnenwald, D. H. (2003). "An Emerging View of Scientific Collaboration: Scientists' Perspectives on Collaboration and Factors that Impact Collaboration." Journal of the American Society for Information Science and Technology **54**(10): 952-965.

Herbsleb, J. D., Mockus, A., Finholt, T. and Grinter, R.E. (2000). Distance, dependencies, and delays in a global collaboration. Proceedings of Computer Supported Cooperative Work 2000, New York, NY ACM.

HIPAA. "Health Insurance Portability and Accountability Act." from http://www.cms.hhs.gov/HIPAAGenInfo.

Hoffelner, W. Applicabilityof of Groupware for Communication in Different Project Environments - A Case Study, Report from ETH Zürich.

Hollan, J. D., and Stornetta, S. (1992). Beyond Being There. Proceedings of ACM SIGCHI'92, Monterey, CA.

Horwitz, A. R., Watson, N. and Parsons, J. T. (2002). "Breaking barriers through collaboration: the example of the Cell Migration Consortium." Genome Biology **3**(11): 2011.

Jeffrey, P. (2003). "Smoothing the waters: Observations on the process of cross-disciplinary research collaboration." Social Studies of Science **33**: 539-562.

Kammer, P. J. B., G. A. Taylor, R. N. Hitomi, A. S. Bergman, M. (1998). "Techniques for Supporting Dynamic and Adaptive Workflow." Computer Supported Cooperative Work **9**(3/4): 269-292.

Kaplan, B., and Shaw, N.T. (2004). "Future directions in evaluation research: People, organizational, and social issues." Methods Inf Med **43**(3): 215-231.

Katz, J. S., and Martin, B. R. (1997). "What is Research Collaboration." Research Policy **26**: 1-18

Korjonen-Close, H. (2005). "The information needs and behaviour of clinical researchers: a user-needs analysis.". Health Info Libr J. **22**(2): 96-106.

Kouzes, R. T., Myers, J. D., and Wulf, W. A. (1996). "Collaboratories: Doing science on the internet." IEEE Computing Practices **29**: 40-46.

Kraut, R., and Egido, C. (1998). Patterns of Contact and Communication in Scientific Research Collaboration. Proceedings of the 1988 ACM conference on Computer-supported cooperative work, Portland, OR.

LaCoursier, S., and Sarkar, M. (2004). Communication and Information needs and barriers: An international collaboratino model. Medinfo 2004.

Latour, B. (1979). Laboratory Life: The Social Construction of Scientific Facts. Princeton, NJ, Princeton University Press.

Lee, C., Dourish, P. and Mark, G. (2006). The Human Infrastructure of Cyberinfrastructure. Computer Supported Cooperative Work (CSCW '06), ACM Press.

Lee, E. S. McDonald, D.W., Anderson, N. and Tarczy-Hornoch, P. (2008 (in Press)). "Incorporating collaboratory concepts into informatics in support of translational interdisciplinary biomedical research." Int. J. Med. Inform.

Lincoln, Y. S., and Guba, E.G. (1985). Naturalistic inquiry. Newbury Park, Sage.

Lindgaard G, D., R, Trbovich, P, White, R, Fernandes, G, Lundahl, S, and Pinnamaneni, A. (2006). "User Needs Analysis and requirements engineering: Theory and Practice. ." Interacting with Computers **18**(1): 47-70.

Lorenzi, N. M. (2004). "Beyond the gadgets: Non-technological barriers to information systems need to be overcome too." British Medical Journal **328**: 1146-1147.

Louie, B., Mork, P., Martin-Sanchez, F., Halevy, A. and Tarczy-Hornoch, P. (2007). "Data Integration and Genomic Medicine." Journal of Biomedical Informatics **40**(1): 5-16.

Malin, B. A. (2005). "An Evaluation of the Current State of Genomic Data Privacy Technology and a Roadmap for the Future." J Am Med Inform Assoc **12**(1): 28-34.

Mankoff, S. P., Brander, C., Ferrone, S., Marincola, F.M. (2004). "Lost in Translation: Obstacles to Translational Medicine." J Transl Med. **2**(1): 14.

Maojo, V., Kulikowski, C.A. (2003). "Bioinformatics and Medical Informatics: Collaborations on the Road to Genomic Medicine? ." J Am Med Inform Assoc. **10**(6): 515-522.

Maxwell, J. A. (1996). Qualitative research design Thousand Oaks, CA, Sage.

Melin, G. (2000). "Pragmatism and self-organization: Research Collaboration on the individual level." Research Policy **29**: 31-40.

Morse, J. M., and Field, P.A. (1995). Qualitative Research Methods for Health Professionals. 2nd ed. Newbury Park, CA, Sage.

MS Exchange. "Microsoft Exchange." from http://www.microsoft.com/exchange/default.mspx.

MS Sharepoint. "Microsoft Sharepoint Server." from http://office.microsoft.com/en-us/sharepointserver.

Muller, M. J., Wildman, D. M., White, E. A. (1993). "Taxonomy of PD Practices: A Brief Practitioner's Guide." Communications of the ACM **36**(4): 26 - 27.

Myers, J. D., Allison, T.C., Bittner, S., Didier, B., Frenklach, M. Green, W.H., et. al. (2004). A Collaborative Informatics Infrastructure for Multi-Scale Science. In the proceedings of the Challenges of Large Applications in Distributed Environments (CLADE) Workshop, Honolulu, HI, IEEE Computer Society.

NCRR. "National Center for Research Resources." from http://www.ncrr.nih.gov/biotech/collabmtg2002.asp.

Neches, R., Fikes, R., Finin, T., Gruber, T., Patil, R., Senator, T., et al. (1991). "Enabling Technology for Knowledge Sharing." AI Mag. **12**(3): 36-56.

NHII. "National Health Information Infrastructure, ." from http://aspe.hhs.gov/sp/NHII.

NIH. "National Institute of Health." from http://www.nih.gov.

NIH data. "NIH data sharing policy." from http://grants.nih.gov/grants/policy/data_sharing.

NIH HIPAA. "NIH HIPAA." from http://privacyruleandresearch.nih.gov/irbandprivacyrule.asp.

NIH interdisc. "NIH Roadmap Interdisciplinary." from http://nihroadmap.nih.gov/interdisciplinary/.

NIH roadmap. "National Institute of Health roadmap." from http://nihroadmap.nih.gov.

NMEDCTS. "National Meeting on Enhancing the Discipline of Clinical and Translational Sciences, ." from http://www.ncrr.nih.gov/ncrrprog/roadmap/Executive_Summary.pdf.

Nutt, G. J. (1996). "The evolution towards flexible workflow systems." Distrib. Syst. Engng **3**: 276–294.

O'Day, V., Annette, A., Kuchinsky, A., and Bouch, A. (2001). When Worlds Collide: Molecular Biology as Interdisciplinary Collaboration. Proceedings of the Seventh European Conference on Computer Supported Cooperative Work.

Olson, G. M., Atkins, D. E., Clauer, R., Finholt, T. A., Jahanian, F., Killeen, T. L., Prakash, A., and Weymouth, T. (1998). "The Upper Atmospheric Research Collaboratory." Interactions **5**(3): 48-55.

Olson, G. M., Teasley, S., Bietz, M. J. and Cogburn, D. L. (2002). Collaboratories to support distributed science: The Example of international HIV/AIDS research. Proceedings of SAICSIT 2002, Port Elizabeth, South Africa, ACM International Conference Proceedings.

Olson, J. S., and Teasley, S. (1996). Groupware in the Wild: Lessons Learned from a Year of Virtual Collocation. In Proceedings of the 1996 ACM Conference on Computer-Supported Cooperative Work (CSCW'96), ACM Press.

Olson, J. S., Olson, G. M., and Hofer, E. C. (2005). What makes for success in science and engineering collaboratories? Workshop on advanced collaborative environments, Redmond, WA.

Patton, M. Q. (2002). Qualitative Research & Evaluation Methods. 3rd ed. . Thousand Oaks, CA, Sage Publications.

PubMed. "PubMed Central." from http://www.pubmedcentral.nih.gov.

Rosetta. "Rosetta BioSoftware." from http://www.rosettabio.com/.

Schleyer, T. (2001). "Collaboratories: leveraging information technology for cooperative research." J Dent Res **80**: 1508–1512.

Schur, A., Keating, K. A., Payne, D. A., Valdez, T., Yates, K. R., and Myers, J. D. (1998). "Collaborative suites for experiment-oriented scientific research." ACM Interactions **3**(5): 40-47.

Shachak, A., Schuval, K. and Fine, K. (2007 prepring). "Barriers and enablers to bioinformatics acceptance: a qualitative study." Journal of the Medical Library Association.

Simmhan, Y., Plale, B., Gannon, D. (2005). "A survey of data provenance in e-science." SIGMOD Record **34**(3).

Sonnenwald, D. H. (2003a). Expectations for a scientific collaboratory: a case study. Proceedings of the 2003 international ACM SIGGROUP, Sanibel Island, FL, ACM Press.

Sonnenwald, D. H., Bergquist, R. E., Maglaughlin, K. L., Soo, E. K., and Whitton, M. C. (2001). Designing to support scientific research across distances: The nanoManipulator environment. Collaborative Virtual Environments. D. S. A. M. E. E. Churchill. London, Springer-Verlag.

Sonnenwald, D. H., Maglaughlin, K. L., and Whitton, M. C. (2004). "Designing to support situation awareness across distances: an example from a scientific collaboratory." Information Processing and Management **40**: 989-1011.

Sonnenwald, D. H., Whitton, M. C., and Maglaughlin, K. L. (2003). "Evaluating a scientific collaboratory: Results of a controlled experiment." ACM Transactions on Computer-Human Interaction **10**(2): 150-176.

Star, S. L., and Ruhleder, K. (1994). Steps towards an ecology of infrastructure: complexproblems in design and access for large-scale collaborative systems. Proceedings of the 1994 ACM Conference on Computer Supported Cooperative Work, Chapel Hill, NC, ACM Press.

Strauss, A. a. C., J. (1998). Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory. 2 ed, Sage Publications.

Sumner, T., Stolze, M. (1997). Evolution, Not Revolution: Participatory Design in the Toolbelt Era. Computers and Design in Context. M. K. a. L. Mathiassen. Cambridge, MA, MIT Press: 1 - 26.

Tang, P. C., Ash, J., Bates, D.W., Overhage, J.M. and SAnds, D.Z. (2006). "Personal Health Records: Definitions, Benefits, and Strategies for Overcoming Barriers to Adoption." Journal of the American Medical Informatics Association **13**(2): 121-126.

Tanner, C., Eckstrom, E., Desai, S.S., Ririe, M.R. and Bowen, J.L. (2006). "Uncovering Frustrations. A qualitative needs assessment of academic general internists as geriatric care providers and teachers." J Gen Intern Med **21**(1): 51-55.

Teasley, S., and Wolinsky, S. (2001). "Scientific collaborations at a distance." Science **292**: 2254-2255.

Wagner, C. S. (2004). "Wiki: A Technology for Conversational Knowledge Management and Group Collaboration." Communications of the Association for Information Systems **13**: 265-289.

Wagner, C. S. (2005). "Six case studies of international collaboration in science." Scientometrics **62**(1): 3-26.

WCS. "Worm Community System." from http://www.canis.uiuc.edu/projects/wcs/index.html.

Weng, C., McDonald, D. W., Sparks, D., McCoy, J. and Gennari, J. H. (2006). "Participatory design of a collaborative clinical trial protocol writing system." International Journal of Medical Informatics **76**(S1): 245-251.

Winograd, T. (1994). "Categories, disciplines, and social coordination." CSCW **2**: 191–197.

Zhao, J., Goble, C., Stevens, R. and Turi, D. (2007). "Mining Taverna's Semantic Web of Provenance." Concurrency and Computation: Practice and Experience.

Appendix A. Recruitment Email

Dear ________________________,

I am a doctoral student in Biomedical Health Informatics at University of Washington conducting field research for my dissertation, which focuses on interdisciplinary collaborative researchers in biomedicine. I am seeking to recruit individuals who are involved in collaborations who will allow me to observe and interview them regarding their collaborative research practices.

Modern biomedical research has become increasingly interdisciplinary and collaborative. To improve support of research collaboration, I am seeking to understand complexities of the collaborative biomedical research setting. The <u>purpose</u> of this research is to better enable and support interdisciplinary collaborative biomedical research.

<u>What I need from you:</u>

Your participation can include any part of following:

- <u>Observations:</u>

  I would like to observe you involved in daily group activities such as group meetings or group work within the institution, as well as meetings that occur with groups external to the institution. During this process, field notes will be taken that characterize the group collaborative processes (i.e. communication tools being used during collaborative process, types of meetings). I <u>will not</u> record your identity during the observation.

- <u>Document examination:</u>

  I would be interested in examples of documents directly related to research collaboration. Notes will be taken while examining documents involved in the collaboration (i.e. text of research proposal (without budget information) for mention of any informatics infrastructure, text of any collaborative activities that you are willing to share or catalog of any collaborative informatics infrastructure involved in the project.

- Interviews:

  I may ask to interview you regarding your perspective and involvement on collaborative research processes. You might be involved in either a short interview (lasting 15-20 min.) or a longer interview (lasting 40-45 min.) or both. I would like to audio-record the interviews. I will provide a consent form for you to sign if you agree to being interviewed.

I will schedule all appointments in advance at your convenience. Your participation is completely voluntary and you can choose to stop at any time. Information about you is confidential.

If you are interested in participating in the study or have any questions about the study, please contact:

Eunjung Sally Lee, Ph.D. student (sallylee@u.washington.edu)

Thank you,

E. Sally Lee

Biomedical Health Informatics

University of Washington

**Please note, we cannot ensure the confidentiality of information sent via e-mail.*

Appendix B. Consent Form

UNIVERSITY OF WASHINGTON CONSENT FORM

Developing a framework for collaborative biomedical informatics infrastructure to facilitate interdisciplinary biomedical research

Investigator:

Name: Eunjung Sally Lee	Academic Affiliation: University of Washington	UW College of : Medicine Biomedical Informatics
Telephone: 206-579-6570	E-mail:* sallylee@u.washington.edu	

**Please note, we cannot ensure the confidentiality of information sent via e-mail.*

INVESTIGATOR'S STATEMENT

I am asking you to be in a research study. The purpose of this consent form is to give you the information you will need to help you decide whether or not to be in the study. Please read the form carefully. You may ask questions about the purpose of the research, what I would ask you to do, the possible risks and benefits, your rights as a volunteer, and anything else about the research or this form that is not clear. When all your questions have been answered, you can decide if you want to be in the study or not. This process is called 'informed consent.'

PURPOSE OF THE STUDY

I want to better enable and support interdisciplinary collaborative biomedical research through developing a framework for design and evaluation of informatics infrastructure supporting collaboration. I would like to interview researchers involved in biomedical collaborations about their experiences of collaboration and collaborative infrastructure.

STUDY PROCEDURES

If you choose to be in this study, I would like to interview you about your perspectives and involvement on collaborative research processes. The interview will last about 1 hour and will focus on collaboration. For example, I will ask you, "What do you think you would need to make collaboration easier or better?" "Please describe some experiences you have had with technologies used in the collaboration process?" and "What type of communication tools or methods do you use?" You do not have to answer every question.

I would like to audio record your interview so that I can have an accurate record. I will keep the audio recordings in a password protected computer. I will transcribe your interview tape within 8 weeks of your interview, assign a study code to the transcript, and destroy the recording.

RISKS, STRESS, OR DISCOMFORT

Some people feel that providing information for research is an invasion of privacy. I have addressed concerns for your privacy in the section below. Some people feel self-conscious when they are audio-taped.

ALTERNATIVES TO TAKING PART IN THIS STUDY

If you do not choose to be in this study, I will not interview you.

BENEFITS OF THE STUDY

I hope the results of this study will help us better enable biomedical collaboration and thereby solving complex biomedical research questions to help broader community. You may not directly benefit from taking part in this research study.

OTHER INFORMATION

Taking part in this study is voluntary. You can stop at any time. Information about you is confidential. I will code the study information. I will keep the link between your name and the code in a separate, secured location until September, 2008. Then I will destroy the link. If the results of this study are published or presented, I will not use your name unless you have given us permission to do so.

Although I will make every effort to keep your information confidential, no system for protecting your confidentiality can be completely secure. It is possible that unauthorized persons might discover that you are in this study, or might obtain information about you. Government or university staff sometimes review studies such as this one to make sure they are being done safely and legally. If a review of this study takes place, your records may be examined. The reviewers will protect your privacy. The study records will not be used to put you at legal risk of harm.

I may want to re-contact you to clarify information from your interview. In that case, I will telephone you and ask you for a convenient time to ask you additional questions closely related to your interview. Please indicate below whether or not you give your permission for me to re-contact you for that purpose. Giving your permission for me to re-contact you does not obligate you in any way.

Eunjung Sally Lee____

Signature of investigator Printed Name Date

Subject's statement

This study has been explained to me. I volunteer to take part in this research. I have had a chance to ask questions. If I have questions later on about the research I can ask one of the investigators listed above. If I have questions about my rights as a research subject, I can call the University of Washington Human Subjects Division at (206) 543-0098. I give my permission for the researchers to audiorecord my interview as described above in this consent form. I will receive a copy of this consent form.

I give my permission for the researcher to re-contact me to clarify information.

Yes _______ No _______

____

Signature of subject Printed name Date

Appendix C. Informational Interview Guiding Questions

1. Could you briefly tell me about the collaboration?
2. Could you tell me your reasons for being part of the collaboration?
   a. How did you get involved?
3. Could you describe a general management or leadership structure of the collaboration as well as any existing oversight committees or advisory board?
   a. Could you tell me how institutions relate to this structure?
   b. Could you tell me about any known long term plans for this collaboration? (i.e. data produced, infrastructure)
4. Could you describe some of the members of the collaboration you closely work with (i.e. Roles, Backgrounds) and how?
   a. Where and how do you usually meet with your collaborators?
   b. How important do you think trust is to maintaining the collaboration? (maybe)
5. Could you describe some of the communication, data or other collaborative systems or tools you use?
   a. Could you describe me how these systems work together?
   b. Could you tell me about some systems that help collaboration?
   c. Could you tell me about some systems that hinder collaborative activities?
   d. Do you have any technical support for these systems?
   e. Could you describe me any standards associated with these systems? (maybe)
6. Could you describe some examples of what you wish you had in terms of tools and support for the collaboration?
7. Could you describe anything that comes to your mind that makes your day to day collaborative activity easier or more difficult? (Focus on tools/systems/support)

Appendix D. In-Depth Interview Guiding Questions

1. Could you describe some of the technologies that you use during collaborative activities?
   (if not mentioned)
   a. Could you tell me how information is shared?
   b. How do you coordinate sharing and activities? (maybe not)
   c. Standards??
   d. Can you tell me anything about security issues associated with collaboration?
2. Could you describe me some of experiences where technologies were really helpful during any collaborative activities?
3. Could you tell me about experiences where you were either frustrated with technology or felt they were inadequate?
   a. Could you tell me if any type of support was available? (maybe)
4. Thinking of during initial stages of collaboration, could you tell me your experience using some technologies that helped during initial stages or collaboration?
   a. What would be helpful?
   b. Frustration?
5. Could you tell me about any known long term plans for this collaboration? (i.e. data produced, infrastructure)
   a. What comes to your mind when you think trust and collaboration?
6. Could you describe some examples of what you wish you had in terms of tools and support for the collaboration?
   a. During initial stages?

7. Could you describe anything that comes to your mind that makes your collaborative activity easier? (maybe not)

Appendix E. Observation Guide Sheet

Location description:

Number of people present:

Description of different types of roles people present play:

Collaboration tools descriptions (meeting medium, communication systems, etc.):

Any mention of management structure associated with collaboration?

Any mention of common collaboration goal?

Any mention of mechanism for sharing data or analyzing data?

Any specific good or bad experiences related to collaboration technology?

Any mention of technical support or lack there of, related to collaboration?

Appendix F. Themes that Guided 2nd interviews

Below are roughly analyzed themes emerged from 1st interviews that guided 2nd interview questions.

F.1 Factors contributing to the continuation of collaboration

Some of the concepts emerged from the interviews as factors that contribute positively to the continuation of collaboration. These are factors that give participants a sense of focus for the collaboration they are part of, or a sense of belonging and understanding for the other researchers in collaboration they are part of.

Code	Description	Setting
Project group identity	Participants often identified himself/herself with certain projects (i.e. Cancer research, Breast research), separating themselves apart from other projects within the Cooperative or Network structure. Participants often would use "we" for their own project as supposed to "they" when they were referring to others they might be marginally involved in, but not feel part of.	GH
Institution and Network identity	Participants often identified themselves with the Institution (Cooperative) or Network. They often used "we" in both referring to Institution (Cooperative) and Network since their relationships with others existed both within and outside Cooperative itself (which could have been another layer of Identity).	GH
Research area identity	Participants also identified themselves specifically to the type of research area they are involved in. This identity did not have the boundary of project group or network/cooperative identity. The researchers in same research area interest often collaborated together.	GH
Identity for competitive edge	There was also an idea of how identity could improve a way for the researchers to work together and make them more competitive.	GH
Common vision, Understanding, Like-mindedness	Participants talked of having some sort of common vision, understanding or like-mindedness within Network/Cooperative. This is related to team spirit and collaborative spirit. This concept associated with identity in a sense that they feel they belong through certain collective identity.	GH, TH
Trust, commitment	Participants spoke of trust and commitment as something that is part of collaboration.	GH, TH
Existing relationships	The participants often spoke of how there are many existing relationships among researchers in their field and within network.	GH, TH

	Many of their collaborative projects start through existing relationships and affiliations. Whether it is participant initiated or initiated by others, the collaboration starts due to participant knowing that person. When the project idea forms, participant often looks for those among his/her known network and affiliations to choose to participate with since they know the other researcher and can depend on existing relationship.	
Communication importance	Participants acknowledge that communication is very important in collaboration, for solving problems or just keeping the collaboration alive.	GH
Face to face meeting	Face to face meetings were thought to be a nice addition to any type of e-mail or phone conversation.	GH, TH
Close relationships	The participants often spoke of how close relationships among researchers are important in maintaining collaboration. Both formal and informal relationships among members of collaboration were thought to be important. Close relationships build trust, understanding, respect for one another, and honesty, all of which are important for collaborative effort.	TH
Importance of a team	A good team, especially of interdisciplinary in nature, is thought to be important for individual researchers. Team structure helps to see new perspectives, generate new ideas, connect to other researchers, create opportunities, and support researchers in their research.	TH
Willingness	Willingness to be open to collaboration is important in starting and maintaining collaborative projects. It is thought that collaboration is not something that can be forced and it is a concept that needs to be in the minds of researchers before having to be part of it.	TH
Importance of focus	Having a focus and related agenda was key to maintaining tight collaboration.	TH
Mentorship	For junior investigators, it was important to have a good mentorship/leadership to support them through research process.	TH

F.2 Factors contributing to start of collaboration

Some of the concepts emerged from the interviews as factors that contribute positively to starting collaboration. These are resources that help participants to connect to other researchers to collaborate, and reasons for doing so.

Code	**Description**	**Setting**
Outside network collaboration	The participants often collaborated with academic institutions and outside community researchers.	GH
Connecting through mentors	The participants often became involved in collaborative projects through their mentors.	GH
Existing relationships	The participants often spoke of how there are many existing relationships among researchers in their field and within network. Many of their collaborative projects start through existing relationships and affiliations. Whether it is participant initiated or	GH, TH

	initiated by others, the collaboration starts due to participant knowing that person. When the project idea forms, participant often looks for those among his/her known network and affiliations to choose to participate with since they know the other researcher and can depend on existing relationship.	
Promoting network	There were lot of promotions of network and collaborative projects, giving visibility via publication, lectures, and being part of the community.	GH
Reason for collaboration: Data	One of the main reasons for collaborating with other network researchers was the need for extra or different set of data.	GH
Reason for collaboration: Expertise	One of the reasons for collaborating with other network researchers was the need for an expertise in certain area.	GH
Reason for collaboration: Funding	One of the reasons for collaborating with other network researchers was the need for funding.	GH, TH
Finding collaborators: cold call, online resources	When there aren't already established relationships existing within network researchers or affiliates, outsiders connect through online resources such as websites or through funding resources. There resources need to be set up properly if collaborations are to happen.	GH, TH
Connecting through mentors	The participants often became involved in collaborative projects through their mentors.	TH
Institutional support	Having support of certain institutional or department for collaboration was helpful in creating collaborative opportunities, both in terms of funding and connecting people.	TH

F.3 Barrier to collaboration: both initiating and continuing

Some of the concepts emerged from the interviews as factors that might be possible barriers to either initiating or continuing collaborative research. These are factors that often promote isolation among researchers, such as local, institutional, and funding differences, and difficulty in finding collaborators.

Code	Description	Setting
Institution/Network vs. Outsider	Participants often set themselves apart from outsiders (i.e. university) with this identity of Cooperative/Network. They often spoke of "we", referring to within network. They alluded to how they are separate from outside network researchers. This concept of "outsider" often posed difficulty of network researchers working with outsiders because of the lack of contextual sensitivity and differences in perspective of outsiders to insider.	GH
Project/Grant based funding	The funding was associated with the large projects within each network, where each large project was often funded by a different funding agency. This created a different project having different relationship with its funding agency as well as having to manage project in different way due to funding.	GH

Site and project specific data	Each institute has its own site-specific data and way of managing. This is due to institutional differences, different data needs as well as project funding related differences. Some of the data across these institutions are aggregated for different projects. Each of these projects also has their own specific way of managing data due to their different data needs.	GH
Institutional and project differences	Even though research network is formed, institutions and projects all have their differences, structurally, size and funding-wise. This creates variability in terms of their capacity to unify their infrastructure for collaboration as well as their capacity to participate in various projects. This might result in strained relationship between institutions and projects.	GH
Unwillingness	Participants acknowledged the existence of unwillingness to collaborate due to feeling threatened or due to simple lack of desire to work together. Not all researchers had collaborative spirit.	GH
Insularity	Projects are often insular in that they do not look at other projects who are dealing with same issues	GH
Difficulty of passing on contextual knowledge	It is difficult for a new person to learn the context quickly.	GH
Purpose of data	The purpose of why data was created is different from how they are used later. The data is often created for administrative/business purposes and later, thought to utilize it for research purpose. Maybe because of this, data might be more difficult to use or not live up to their expectations.	GH
Difficulty of disseminating information	It is difficult to disseminate project resources, because it is not easy to know who these resources need to reach among vast project groups.	GH
Difficulty of change	Change of any kind is difficult, especially within large collaboration.	GH
Lack of experts in certain research area	Research collaboration activities are limited in the network due to lack of experts in certain research area. Because collaborative projects require researchers within network to be interested in those projects, without the necessary experts in various areas to be interested in projects, the projects have no chance of starting.	GH
Difficulty of finding collaborator	Participants spoke of how finding collaborator is difficult due to lack of knowledge in how to and difficulty of obtaining information due to size of network.	GH, TH
Difficulty of recruiting new researchers	Recruiting new researchers into certain environment is sometimes difficult.	GH
Difficulties and Limitations of Technology	Participants spoke of some of the difficulties they had using technology to communicate and meet. Because technology is an important part of being able to communicate during remote collaborations, some of the frustrations voiced could have affect on overall collaboration. Various methods were employed by individuals to alleviate some of these technical difficulties.	TH
Difficulties of	Aside from technical difficulties, some difficulties of remote	TH

remote collaboration	collaboration were mentioned, from coordinating time to keeping up with others' work and awareness of the project as whole.	

F.4 Facilitating the process of collaboration

Some of these concepts emerged as the factors that facilitate the process of collaboration. These are factors important to day-to-day operation of collaboration and keeping collaborative projects on track to do what they were intended to do.

Code	**Description**	**Setting**
Data standardization	In order to aggregate data across different sites and/or projects, standards are used. There are some variables common to all studies, but other ones are usually added. Each site has to keep their data in this agreed standard format.	GH
Data sharing, resource protection	Because data are being aggregated and shared, there has to be some sort of data sharing protocol in place (also known as data use agreement). This is for resource protection as well as for institutional review process. Each project/collaboration has to have this in place in order to proceed.	GH, TH
Communications by phone	Collaboration involves a lot of phone communications, both one on one and conference calls.	GH, TH
Communication by email	Collaboration involves a lot of email communication for contact, introductions, project announcements, etc.	GH, TH
Meetings	Collaborations also involve a lot of meetings, both face to face and conference-based scheduled regularly and with set agendas. These meetings are often for updates, to keep agenda on track and are important for collaborative activities.	GH, TH
Committees	Collaborations have lots of committees for set agendas and goals. These committees are in charge of specific goals and are important for collaborative activities.	GH
IRB	IRB was one of the primary concerns for all participants in regards to multi-site collaborative efforts. Due to differences in IRB for all sites, the researchers were burdened with having to go through multiple IRBs to form multi-site collaborative projects. Such difficulty will impact researchers' desire to do collaborative projects.	GH, TH
Collaborative research support	Collaboration is a difficult process and especially one that spans multi-site. Participants indicated that the process of collaboration needed a lot of support. When support or champion for the collaborative process exists, it fosters collaboration while those without are not as involved with collaborative process.	GH
Infrastructure	Participants had mixed opinions about technical infrastructure. There were specific components of infrastructure that was working well for people while others, not so much. The infrastructure didn't seem to be all working well together and due to variability in funding mechanisms and institutional differences,	GH

	finding a common infrastructure scheme was a challenge. However, technical infrastructure was not something voiced as being a big problem for the collaborations. • No good infrastructure exists • Infrastructure doesn't all work together • Infrastructure working fine • Infrastructure variability due to management and project differences • Technology is really not barrier • Building on previous infrastructure • Infrastructure challenge: Adaptability, flexibility	
Infrastructure: data warehouse	One of the most frequently mentioned infrastructure was data warehouse. The data warehouse and online data wiki accessing data warehouse were actively in use by various projects. Although slowness was the problem, the accessibility, flexibility, and frequent updates were noted as good points. • Data warehouse • Data warehouse is working well • Data warehouse being used by multiple projects • Data warehouse slow • Data wiki: Accessible, flexible, frequently updated	GH
Infrastructure: online resources	Online resources such as websites, wiki, and repositories were also mentioned by participants as being use. However, websites still lacked desired capabilities, and were thought by some participants as not very useful. • Websites • Repositories • Wiki • Collaborative writing • Website not so useful • Ease of use needed	GH
Infrastructure: Online directory	Participants mentioned online directory as possible solution to being able to find collaborators better. However, online directory for some projects didn't exist possibly due to reluctance of researchers and lack of ease of use. • Need for online directory • Lack of online directory • Reluctance for online directory	GH
Coordination and facilitation	Due to multiple projects and agendas associated with the team, someone has to coordinate and facilitate each to make sure all are heard equally.	TH
Archiving	Archiving was important to the collaborative group. One of the primary reason why certain technology was selected to be used was the archiving feature.	TH

F.5 Long-term collaborative issues

Often, collaborative projects do not think beyond the funding period. For long-term collaboration, thinking beyond the funding period is important.

Code	Description	Setting
End of funding issue	There was a problem of what to do once the funding ends. The resources that existed during funding might not be available afterwards, so unless there was some planning of what will happen to infrastructure that existed during funding, after funding, it might be lost.	GH

Appendix G. Preliminary Codes and Categories

G.1 Codes and categories emerging from GH site

Category 1: Collaborative Process Related

Finding collaborators/resources

Timeline: Pre-project

Code	Description	Data	Effects
Online directory, information resource	Online directory of researchers or information on resources and opportunities were mentioned as a technology/infrastructure that might help alleviate the difficulty of finding resources/collaborators due to size and distance.	I, A, O	+, w
Distance and size of collaboration	Due to remote nature and large size of collaboration, it is hard for researchers to keep track of all the new opportunities, collaborators or resources available.	I	-
Existing social connections	Many of their collaborative projects start through existing relationships and affiliations. Whether it is participant initiated or initiated by others, the collaboration starts due to participant knowing that person. When the project idea forms, participant often looks for those among his/her known network and affiliations to choose to participate with since they know the other researcher and can depend on existing relationship. The participants also became involved in collaborative projects through their mentors.	I	+
Administrative personnel/Match makers	Difficulty of finding collaborators can be alleviated by having administrative personnel who can connect researchers together. These "matchmakers" have knowledge of the context (i.e. organizational structure, researchers and their backgrounds) and are well-known and trusted. They are thus able to help form collaborative projects by connecting researchers together.	I	+
Researcher recruitment, Promotion	Research collaboration activities are limited in the network due to lack of experts in certain research area. There were mentions network and collaborative project promotions and active recruitment via publications, lectures, conferences and being part of the community. Such active promotion and recruitment might alleviate problem of lack of experts in certain area.	I, A, O	+

Information, Data sharing

Timeline: Pre/During/Post project

Code	Description	Data	Effects
Data security protocol, de-identified data	Due to security concerns, extra steps need to be taken to protect data through de-identification and careful protocol.	I, A, O	-
File transfer technology, Data wiki	File transfer technology is often used to transfer data and information between researchers, enabling sharing.	I, A	+
Data warehouse, repository technology	One technology mentioned that help facilitate storage of research data was data warehouse.	I, A, O	+
Standards for data	In order to aggregate data across different sites and/or projects, standards are used. There are some variables common to all studies, but other ones are usually added. Each site has to keep their data in this agreed standard format.	I, A	+
Project specific data, various data purposes	Each institute has its own site-specific data and way of managing. This is due to institutional differences, different data needs as well as project funding related differences. Some of the data across these institutions are aggregated for different projects. Each of these projects also has their own specific way of managing data due to their different data needs. The purpose of why data was created is different from how they are used later. The data is often created for administrative/business purposes and later, thought to utilize it for research purpose. Maybe because of this, data might be more difficult to use or not live up to their expectations.	I	-
Common, easy access to information	Researchers often mentioned they would like easier access to information or data. Data or information were often spread out over multiple places or were not able to be accessed other than at certain places.	I	w

Communication

Timeline: Pre/During/Post project

Code	Description	Data	Effects
Email, Asynchronous communication	Email was technology often used in every aspect of collaboration and was perceived as an essential technology. Collaboration involves a lot of email communication for contact, introductions, project announcements, etc.	I, O	+
Face-to-face, local, conference meetings	Collaborations involve a lot face-to-face meetings. These meetings perceived as helpful and essential in every step of collaborative process, from the formation of ideas to continuation of relationships. Participants voiced wish for opportunity for more face-to-face time. Collaborations also	I, A, O	+, w

	involve a lot of meetings, both face to face and conference-based scheduled regularly and with set agendas. These meetings are often for updates, to keep agenda on track and are important for collaborative activities.		
Phone, cell phone, conference calls, Synchronous communication	Collaboration involves a lot of phone communications, both one on one and conference calls.	I, A, O	+
Video communication, Technology enabling face-to-face communication	Because the modern day collaboration is often remote and it is difficult economically and time-wise for actual face-to-face communication, researchers often desired technology that would enable such type of communication.	I	W

Administrative processes

Timeline: Pre/During project

Code	Description	Data	Effects
Support personnel	Collaboration is a difficult process and especially one that spans multi-site. Participants indicated that the process of collaboration needed a lot of support. When support or champion for the collaborative process exists, it fosters collaboration while those without are not as involved with collaborative process.	I, A, O	+
Common process, Standardization, Templates	Standardization of processes or guidelines of steps involved in various aspects of collaboration help guide those involved in the difficult processes of collaborative projects.	I, A	+
Information distribution	Although a lot of information has to be disseminated and distributed to researchers involved in collaboration, the methods of doing so was not very efficient or often difficult. Participants wished for a better way of doing this task.	I, O	+, w
Committees, Agenda	Collaborations have lots of committees for set agendas and goals. These committees are in charge of specific goals and are important for collaborative activities.	I, A, O	+
Schedule coordination	Collaborative projects often have to deal with researchers' busy schedule, varied locations and remote distances where coordinating meetings and schedules become very difficult. Having some sort of ability to have a common calendar system across institutions and researchers were viewed as helpful.	I	w, +
Websites, online technology	Websites or other online technologies were often used to disseminate information. Researchers had mixed opinions regarding these technologies. If they were built to support collaborative processes, they could be very helpful, it not, useless.	I, A, O	+, -

Financial

Timeline: Pre/During project

Code	Description	Data	Effects
Insufficient funding	There's often mention of funding limitation. Because funding is always limited and/or under funded, only the most crucial collaborative processes are supported. Some of the negative effects of limited funding are: not being able to involve all those who wish to participate in a project, not being able to have more comprehensive infrastructure, having difficulty funding preliminary stages of projects before funding, and having to work beyond work hours to finish end of funding deliverables.	I, O	-
Project-based funding	The funding was associated with the large collaborative projects within each network, where each project was often funded by a different funding agency. This created a different project having different relationship with its funding agency as well as having to manage projects in a different way.	I	-

Institutional, federal regulatory processes

Timeline: Pre project

Code	Description	Data	Effects
Federal regulations, Institutional review board	Federal regulatory process, IRB was one of the primary concerns for all participants in regards to multi-site collaborative efforts. Due to differences in IRB for different sites, the researchers were burdened with having to go through multiple IRBs to form multi-site collaborative projects.	I, A, O	-
Common regulatory process	To alleviate some of the difficulties of multiple regulatory processes, participants talked of a wish for common regulatory process and tools to support it.	I, A, O	w

Category 2: Social issues surrounding collaboration

Timeline: Pre/During/Post project

Code	Description	Data	Effects
Contextual differences	Even though research network is formed, institutions and projects all have their differences, structurally, size and funding-wise. This creates variability in terms of their capacity to unify their infrastructure for collaboration as well as their capacity to participate in various projects. This might result in strained relationship between institutions and projects.	I, A, O	-
Developed relationships, Trust	Participants spoke of trust and commitment as something that is an integral part of collaboration. The participants often spoke of how there are many existing relationships among	I, O	+

	researchers in their field and within network.		
Time consuming, lot of work	Collaborative project is often time consuming and requires lot of effort and work compared to non-collaborative projects. The more institutions and/or researchers are involved, more difficult it becomes due to the institutional specific processes or researcher-specific quirks and practices.	I	-
Unwillingness	Often those involved in collaboration, whether individual or institution, are protective of their own work/resources/data and are unwilling to share them. Participants acknowledged the existence of unwillingness to collaborate due to feeling threatened or due to simple lack of desire to work together. Not all researchers had collaborative spirit.	I	-
Incentive for collaboration	Collaborative projects are often motivated by incentives such as gaining new knowledge, resources/funding, data, and expertise. Without incentive, it is difficult to overcome the difficulties of collaboration.	I	+
Common ground	Common ground referring to: common goal/vision, like-mindedness, shared interest, common work ethics/style, personality match and is related also to team and collaborative spirit. Participants talked of having this common ground with those they collaborate with.	I, A	+
Involvement, contribution	Feeling of contributing to project or feeling of being involved in collaborative processes is important. If a researcher or institute feels as though it is only providing data or is involved in name only, there is not incentive enough for them to continue with the collaboration.	I, A	+
Inside vs. Outsider	Participants often set themselves apart from outsiders (i.e. university). They often spoke of "we", referring to insiders. They alluded to how they are separate from outside network researchers. This concept of "outsider" often posed difficulty of network researchers working with outsiders because of the lack of contextual sensitivity and differences in perspective of outsiders to insider.	I, A	-
Change management	Changes in structure, technology, processes are difficult, but collaborative projects often involve changes due to contextual differences across collaborators.	I, A, O	-

Category 3: Technical issues surrounding collaboration

Timeline: Pre/During/Post project

Code	Description	Data	Effects
Technology not based on need	Technology without need, technology that does not support the process were not used or made collaborative processes more difficult.	I, O	-
Technology is not a solution	Many participants voiced their opinion that though technology might enable collaboration, it is not solution to difficulties existing in collaborative projects.	I	

Collaborative technology, infrastructure	Collaborative technologies or infrastructure to help collaborative processes were generally thought to facilitate collaboration through bridging distances, enabling communication, and enabling collaborative processes.	I, A	+

G.2 Codes and categories emerging from TH site

Category 1: Collaborative Process Related

Finding collaborators/resources

Timeline: Pre-project

Code	Description	Data	Effects
Existing social connections	Many of their collaborative projects start through existing relationships and affiliations. Whether it is participant initiated or initiated by others, the collaboration starts due to participant knowing that person. When the project idea forms, participant often looks for those among his/her known network and affiliations to choose to participate with since they know the other researcher and can depend on existing relationship. The participants also became involved in collaborative projects through their mentors.	I, A	+
Institutional help	Having support of certain institutional or department for collaboration was helpful in creating collaborative opportunities, both in terms of funding and connecting people.	I	+

Administrative processes

Timeline: Pre/During project

Code	Description	Data	Effects
Setting Agenda	Collaborative projects require agendas to keep track of what needs to be done and who needs to be heard.	I, O	+
Coordinating schedule	Collaborative projects often have to deal with researchers' busy schedule, varied locations and remote distances where coordinating meetings and schedules become very difficult.	I	-

Information/Data sharing, dissemination

Timeline: Pre/During/Post project

Code	Description	Data	Effects
Data warehouse, repository technology	One technology mentioned that help facilitate storage of research data was data warehouse.	I, A, O	+
Data security	Due to security concerns, extra steps need to be taken to	I, O	-

protocol, de-identified data	protect data through de-identification and careful protocol.		

Communication

Timeline: Pre/During/Post project

Code	Description	Data	Effects
Phone, cell phone, conference calls, online chat, Synchronous communication	Collaboration involves a lot of phone communications, both one on one and conference calls.	I, A, O	+
Email, Asynchronous communication	Email was technology often used in every aspect of collaboration and was perceived as an essential technology. Collaboration involves a lot of email communication for contact, introductions, project announcements, etc.	I, A, O	+
Face-to-face, local, conference meetings	Collaborations involve a lot face-to-face meetings. These meetings perceived as helpful and essential in every step of collaborative process, from the formation of ideas to continuation of relationships. Participants voiced wish for opportunity for more face-to-face time. Collaborations also involve a lot of meetings, both face to face and conference-based scheduled regularly and with set agendas. These meetings are often for updates, to keep agenda on track and are important for collaborative activities.	I, A, O	+, w
Limitations of communication technology	Participants spoke of some of the difficulties they had using technology to communicate and meet. Because technology is an important part of being able to communicate during remote collaborations, some of the frustrations voiced could have affect on overall collaboration. Various methods were employed by individuals to alleviate some of these technical difficulties.	I, O	-
Video communication, Technology enabling face-to-face communication	Because the modern day collaboration is often remote and it is difficult economically and time-wise for actual face-to-face communication, researchers often desired technology that would enable such type of communication.		w

TH codes – Social issues surrounding collaboration

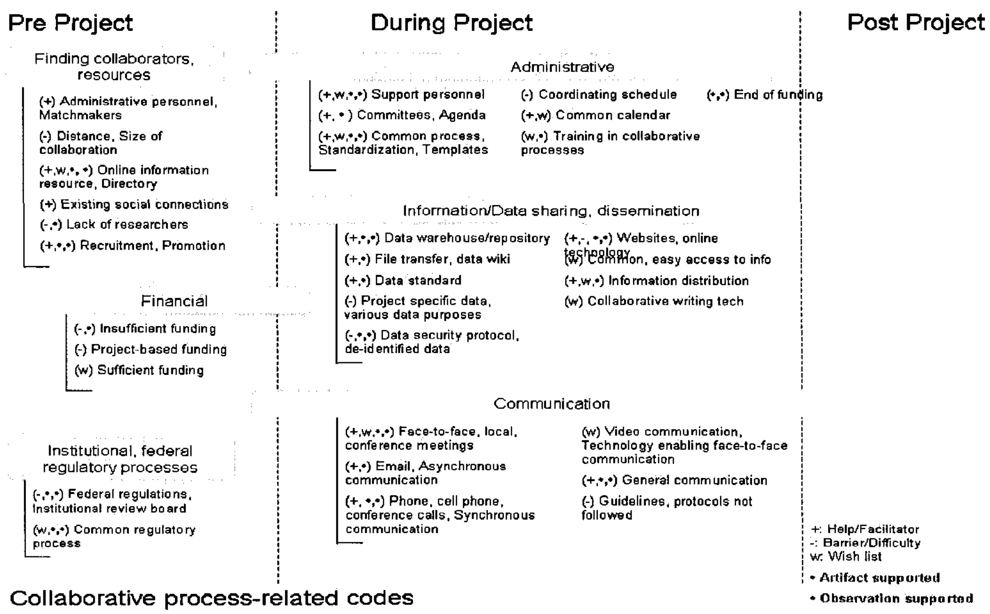
Timeline: Pre/During/Post project

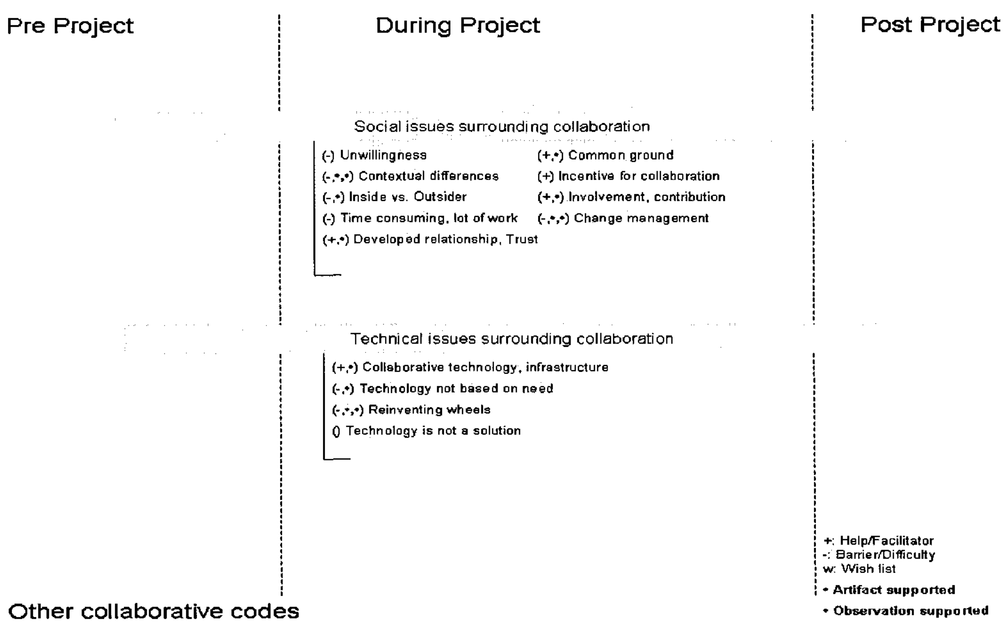
Code	Description	Data	Effects
Common ground	Common ground referring to: common goal/vision, like-mindedness, shared interest, common work	I	+

	ethics/style, personality match and is related also to team and collaborative spirit. Participants talked of having this common ground with those they collaborate with.		
Developed relationships, Trust	Participants spoke of trust and commitment as something that is an integral part of collaboration. The participants often spoke of how there are many existing relationships among researchers in their field. The participants often spoke of how close relationships among researchers are important in maintaining collaboration. Both formal and informal relationships among members of collaboration were thought to be important. Close relationships build trust, understanding, respect for one another, and honesty, all of which are important for collaborative effort.	I, A, O	+
Team benefits	A good team, especially of interdisciplinary in nature, is thought to be important for individual researchers. Team structure helps to see new perspectives, generate new ideas, connect to other researchers, create opportunities, and support researchers in their research.	I, A	+
Mentorship	For junior investigators, it was important to have a good mentorship/leadership to support them through research process.	I, O	+
Willingness	Willingness to be open to collaboration is important in starting and maintaining collaborative projects. It is thought that collaboration is not something that can be forced and it is a concept that needs to be in the minds of researchers before having to be part of it.	I, O	+

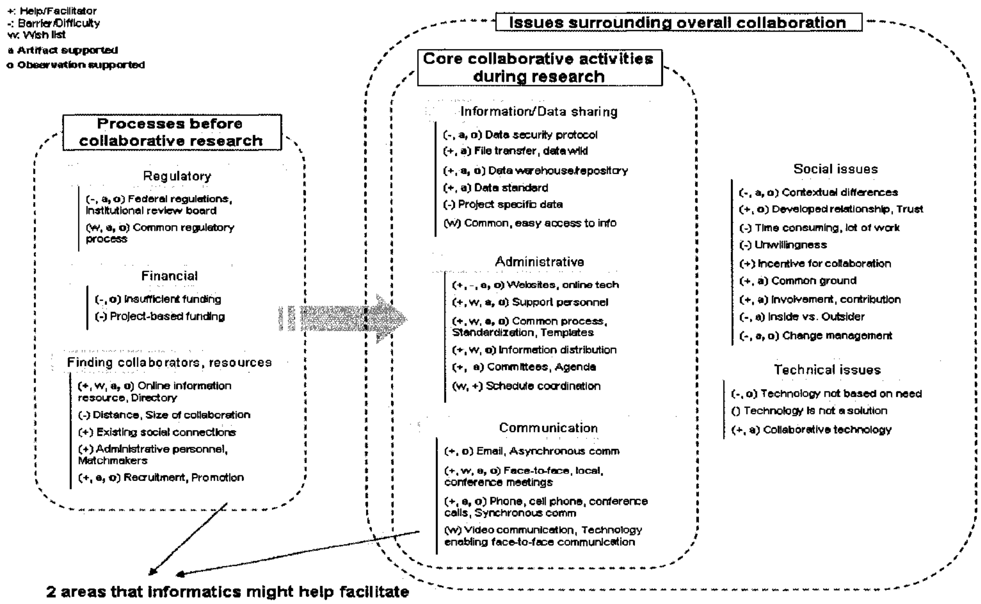
Appendix H. Progression of Preliminary framework Development

H.1 First version of framework after few iterations of refinement

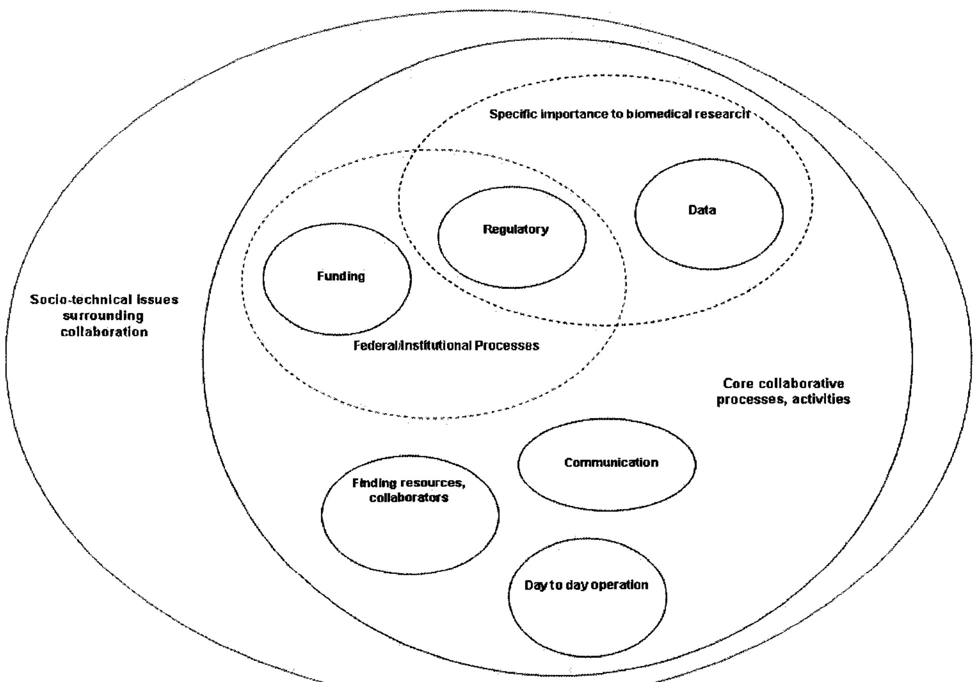




H.2 Second version of framework after more iterations of refinement



H.3 Third version of framework after more iterations of refinement



Curriculum Vitae

Eunjung Sally Lee

EDUCATION:

- **2004-present Ph.D. in Biomedical Health Informatics**
  University of Washington
- **2001 M.S. in Computer Science**
  California State University of Long Beach, 3.68 GPA
- **1998 B.S. in Biology**
  University of California, Los Angeles, 3.65 GPA
  Cum Laude, Departmental Highest Honors, College Honors

PUBLICATIONS:

- **Lee, E.S.**, McDonald, D.W., Anderson, N.R., and Tarczy-Hornoch P. *Incorporating collaboratory concepts into informatics in support of translational interdisciplinary biomedical research.* International Journal of Medical Informatics (in print)
- Turner, A.M., Ramey, J., and **Lee, E.S.** *Connecting Public Health IT Systems with Enacted Work: Report of an Ethnographic Study.* American Medical Informatics Association Fall Symposium 2008 (accepted)
- **Lee, E. S.** *Facilitating collaborative biomedical research.* In Proceedings, Conference on Supporting Group Work. Group '07 Doctoral Consortium papers. November 4-7, 2007 Sanibel Island, FL.
- Anderson, N.R., **Lee, E.S.**, Brockenbrough, J.S., Minie, M.E., Fuller, S., Brinkley, J., and Tarczy-Hornoch, P. *Issues in Biomedical Research Data Management and Analysis: Needs and Barriers.* Journal of American Medical Informatics Association. 14(4) 478-488.
- Bales, N., Brinkley, J., **Lee, E.S.**, Mathur, S., Re, C., and Suciu, D. 2005. *A Framework for XML-based Integration of Data, Visualization and Analysis in a Biomedical Domain.* In Proceedings, XSym 2005, pages pp. 207-221, Trondheim, Norway.

- **Lee, E.S.**, Suciu, D., and Brinkley, J.F. 2005. *Automatic XQuery Generation and Generalized Visualization for an XML Interface to a Relational Database.* In Proceedings, American Medical Informatics Association Fall Symposium

- **Lee, E.S.**, Lewitus, A.J., and Zimmer, R.K. 1999. *Chemoreception in a marine cryptophyte: Behavioral plasticity in response to amino acids and nitrate.* Limnol. Oceanogr. 44(6): 1571-1574

RESEARCH EXPERIENCES:

- **1/06 – present**
  Dissertation: Facilitating multi-institutional interdisciplinary biomedical collaboration
  - Development of a framework on collaboration through interviews, observations and examination of work-related documents

- **7/07-3/07 Public Health IT Systems Workflow**
  - Development of IT system workflow through interviews, observations

- **3/06-7/06 Needs Assessment of Clinical Research Lab**
  - Assessment of needs of clinical research lab interviews and observations

- **4/05-3/07 Needs Assessment of Biomedical Researchers**
  - Assessment of needs of biomedical researchers through interviews and survey

- **7/04-3/05 Human Brain Project**
  - Development of system for query and generalized visualization of heterogeneous data sources using xQuery and Java

- **8/96-9/98 Honors Undergraduate Biology Research**, UCLA
  - Analysis of swimming behavior on a computer-assisted video motion analyzer (CAVMA) interfaced with a Sun Microsystems SPARC 2 workstation

HONORS/AWARDS:

- NLM Training Grant, 2004-2007
- Undergraduate Research Award, UCLA, 1997
- Dean's Honor List Fall 94, Fall 96, Spring 97, Fall 97
- NSF Grant, University of South Carolina, Summer 1997

WORK AND RELATED EXPERIENCES:

- **7/07-present** **Senior Computer Specialist**
  Evans School, University of Washington
  - Support faculty, students, staff with any computer needs
  - MS SQL, MySQL, Access database and website maintenance

- **5/01-5/04** **Software, Web Developer/Project Manager,** Urban Science Inc.
  - Data Analysis, Data mining using MS SQL and MS Access
  - Development of strategic planning and marketing website and report products in VB, C++, C#
  - Consulting clients on Data analysis for target marketing

- **10/99-2/01** **Web Developer, Netovations**
  - ASP, Cold Fusion web/e-Commerce sites design and development
  - MS SQL, MS Access database administration, design

- **5/03-5/04** **Education Docent Volunteer**
  Long Beach Aquarium of the Pacific
  - Educating the public about invertebrates at the touch lab
  - Presenting facts about pacific marine life to the public

- **9/03-present** **Recreation Class Instructor**, Parks and Recreation
  - Teaching public about middle eastern dancing and movements

- **4/96-9/97** **Laboratory Assistant**, Department of Molecular Biology, UCLA
  - Maintenance of lab Equipment and supplies
  - Assisted with molecular biology research – cloning, maintaining and managing of cultures, manual sequencing