"Greater than the Sum of its Parts:" Coordinating Centers as Facilitators of Network-Level Work in Cancer Epidemiology Coordinating Center Enabled Networks

Betsy Rolland

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2013

Reading Committee:

Charlotte P. Lee, Chair

Matthew J. Bietz

John D. Potter

Mark Zachry

Program Authorized to Offer Degree:

Human Centered Design & Engineering

University of Washington

Abstract

"Greater than the Sum of its Parts:"

Coordinating Centers as Facilitators of Network-Level Work in Coordinating Center Enabled Networks

Betsy Rolland

Chair of the Supervisory Committee:

Assistant Professor Charlotte P. Lee

Human Centered Design & Engineering

As collaborative research has increased in size and scope, the overhead of managing such large endeavors has also increased. In cancer epidemiology, one tool used to address the challenges of working on multi-institutional research projects is the employment of a Coordinating Center (CC) charged with the facilitation of the project's scientific objectives. Yet little research has been done on how CCs work or how their work affects the work of the networks in which they are employed, here called "Coordinating Center Enabled Networks," or CCENs. This study addresses the research question, How do CCs facilitate the network-level work of CCENs? Employing ethnographic methods and qualitative data analysis, this study incorporates seven months of observation and 9 interviews of CC staff and Principal Investigators, as well as 8 interviews with funding agency representatives and CCEN collaborators. This dissertation presents a definition of the CCEN form, describes the work of the CCEN and develops a typology of work practices, providing an analytical lens through which it is possible to get a better sense of precisely what a CCEN is and does. Finally, this study proposes a definition of facilitation in collaborative cancer

epidemiology as the application of the CC's collective and individual knowledge and experience, amassed over years of experience supporting collaborative, multi-institutional research projects, to the development of systems and processes to address the challenges of networked science. This study highlights the need for more research in the area of coordination of collaborative work and the use of CCs in collaborative scientific research.

TABLE OF CONTENTS

List of Figures

List of Tables

Acknowledgements

I want to thank my amazing participants, many of whom are my colleagues and friends, for their generosity and willingness to open their professional lives to scrutiny. Their support and wholesale embrace of this project was invaluable.

I am also grateful to the National Cancer Institute for their support via grant R03CA150036 and to the Fred Hutchinson Cancer Research Center for their continued support.

Dedication

To my husband, Stephen Flaherty, whose kindness, generosity and willingness to lend a helping hand are the stuff of legends and have made this whole endeavor possible.

To my son, Henry John Flaherty, whose unflagging curiosity and dedication to and passion for learning inspire me every day.

To my mother, Nyla Rolland, whose devotion to teaching and learning made me the student I am today.

To my father, Dennis Rolland, whose sacrifices for his country taught me that nothing worth doing is easy.

Finally, to Dr. John D. Potter, my mentor and friend, whose encouragement and enthusiasm for my scholarly pursuits gave me the courage to take the first step on this path.

Chapter 1: Introduction

In recent years, biomedical research has become increasingly collaborative (Falk-Krzesinski et al., 2011; Wuchty, Jones, & Uzzi, 2007). Development of information and communication technologies (ICTs) has allowed scientists to work together in larger numbers, on increasingly complex problems, over ever-greater distances. Such large collaborative projects bring together scientists from different labs, different disciplines, and different institutions, generally managing to bring all these disparate elements together into a functioning whole. Yet this collaboration comes at a cost. Coordinating large numbers of dispersed researchers working on complex questions such as global warming or early detection of cancer across geographic and institutional boundaries requires a significant commitment of time and resources (Cummings & Kiesler, 2007). This administrative burden often falls to the lead Principal Investigator (PI) and his/her staff.

In the field of cancer epidemiology, multi-site research projects are increasingly employing Coordinating Centers (CCs) as a tool to ease that administrative burden by offloading it onto a group with substantial experience in the coordination of such projects (Rolland, Smith, & Potter, 2011). A CC is a central body tasked with coordination and operations management of a multi-site research project. I call the groups that the CC coordinates "Coordinating Center-Enabled Networks" (CCENs). CCENs are research networks comprised of scientists, representatives of funding agencies, and CC staff, all of whom are focused on the overarching goals of the collaborative project, goals that can be achieved only within a network structure. This research represents the first step in identifying, defining and understanding this specific type of collaborative research structure.

Yet we know very little about either how such CCENs function or how best to facilitate them. In fact, there is no definition of what facilitation means in the context of CCENs. CCs receive very little guidance as to how to go about their tasks, beyond the expectations laid out in the funding agency's Request for

Application (RFA), the document that details the requirements of the project. Few CCs write about their work, leaving new CC PIs and managers to devise their practices anew without evidence of efficiency or efficacy. NIH spends millions of dollars each year supporting such CCENs and their CCs, yet little research has been done on how the CCs work, how to structure these CC or precisely which aspects of the research project should be allocated to the CC. This dissertation research seeks to rectify that deficiency by investigating and documenting the work practices of two current CCs involved in CCENs.

In this project, I set out to answer the question of how a CC facilitates collaboration within a CCEN. Through seven months of meeting observations and 17 interviews, I took a rare look inside the operations of two CCs and their role within their respective CCENs. By doing so, I hope to identify areas of the collaborative process that are enhanced by the work of the CC. The areas on which study participants choose to focus, along with their tools and techniques, are the result of collective decades of experience coordinating multi-site projects. As such, they represent crucial sources of knowledge, which, in turn, could be used to improve the process of collaboration in other networked science projects.

Research Questions:

RQ: How do Coordinating Centers facilitate the network-level work of Coordinating Center-Enabled Networks (CCENs)?

- **RQa:** What are the distinguishing characteristics of a CCEN?
- **RQb:** What work do each of the CCEN components perform?
- **RQc:** What does facilitation mean in a CCEN?

The two CCs that were the focus of this research are housed at the Fred Hutchinson Cancer Research Center (FHCRC) in Seattle, WA, and are run by a group at FHCRC that specializes in the management of multi-site research projects, the Science Facilitation Team (SFT). As such, the two CCs share many staff and PIs, making them an ideal case study in which to explore work practices as applied to two CCENs with very different scientific objectives. The first, the Biomarker Network (BN), has been in operation for approximately 12 years and has as its overarching scientific objective the discovery and validation of biomarkers for cancer diagnosis and prognosis. Biomarkers are biologic markers that can be detected in

the body via biologic samples such as blood or urine, and are used to detect cancer, measure its progression or monitor treatment response. The aim of this project is to prove the efficacy and reliability of such markers such that they may be used in clinical practice. The Screening Network (SN) is a relatively new project, having been funded approximately 4 months before I began my fieldwork (Fall 2012). The SN seeks to improve the screening process in the United States by developing a deeper understanding of the process and by searching for ways to personalize screening recommendations for patients, based on their risk profiles. The specific aim of this project is the creation of a data repository of screening events across the populations at seven different Research Centers (RCs). Three of these RCs are focused on breast cancer, three on colorectal cancer and one on cervical cancer.

This research involved seven months of observations of meetings of the SFT, as well as the larger, in-person meetings of the CCENs themselves. I also conducted interviews with 17 CCEN members, including nine CC staff and PIs, two funding agency representatives, three BN PIs and three SN PIs. These interviews were semi-structured with questions focused on the work of the CCEN and the CC. Interviews were digitally recorded and transcribed, then coded using qualitative analysis software according to interview questions and themes. The data analysis process is described in more detail in Section 3.3.3.

One of the main contributions of this dissertation is the identification, definition and description of the CCEN form, which is defined as a research network comprised of scientists, representatives of funding agencies and CC staff, all of whom are focused on the overarching goals of the collaborative project, goals that can only be achieved within a network structure. As the name implies, one of the defining characteristics of the CCEN is its inclusion of a CC the charge of which is to facilitate the work of the CCEN. Yet CCs received little guidance in how to translate that mandate from the vague language of the RFA into day-to-day activities, process and systems that help the collaborative work progress toward achievement of the CCEN's scientific goals.

A second main contribution of this dissertation is the description of a typology of work in which the CCEN engages; these work practices provide an analytical lens through which we can get a better sense of precisely what a CCEN is and does. Some of these work practices involve all three types of CCEN participants – the CC, RCs, and funding agency representatives – and are here called "network-level" work. The CCEN's formation in service of a scientific objective that can be achieved only within the network structure lends special importance to this network-level work, as it is at the heart of why the CCEN exists. And it is upon this network-level work that the CC spends the majority of its time. By focusing on this specific set of work practices, I begin to craft a definition of what facilitation means in collaborative CCEN work and describe precisely how the CC goes about this work, a third contribution of this dissertation.

Thus, we find the answer to the central research question of this dissertation in the CC's application of their experience and expertise to the challenges of collaborative research. The CC plays a distinct role in the CCEN, facilitating the network-level work of the project, with the aim of making the work of the CCEN go more smoothly and generating high-quality data. This facilitation involves the application of the CC's collective and individual knowledge and experience, amassed over years of experience managing and supporting collaborative, multi-institutional research projects. In service of this goal, the CC has developed systems and processes to address the challenges of networked science. When the CC is allowed to play this distinct role as a facilitator, as in the BN, the network-level work of the CCEN moves toward the achievement of its scientific goals with little resistance. However, when the role of the CC is limited, as in the SN, weaknesses and conflicts in one area of work spill over into other areas and the SN is not in a position to counteract these negative forces.

Chapter 2: Background

Collaboration has been the subject of study for decades and many lessons have been learned about how collaborative groups work. Yet few hard and fast rules have been developed that can be applied to situations beyond the original context of the study. Thus, one of the great lessons of studies of collaborative science has been that context matters. A strategy for successful collaboration among graduate students in one professor's lab does not necessarily transfer intact to a large, interdisciplinary group building the Large Hadron Collider, for example. Over time, the goal of studies of collaboration has shifted from seeking universal, prescriptive rules for how to collaborate to description of factors that can impact a group's success [e.g., Falk-Krzesinski et al., 2011; Wuchty et al., 2007]. Such an approach has not met with great success, however, again because of context. Finding that leadership is crucial to the success of a consortium, for example, does not actually help a consortium turn their current PI into a great leader. Furthermore, examples of successful consortia with mediocre leadership can also be found, further confusing the matter.

As these multi-site projects are developed via funding awards, it appears that little thought is given to the social and organizational connections required for a successful collaboration. While great attention is paid to the scientific components of the project when composing the Request For Applications (RFA), nowhere in the NIH peer review system is attention given to how the varying applicants will work together; it is simply assumed that they will because they must. This *laissez faire* approach to designing the collaboration is also evident in the construction of most collaborative organizational structures. A Steering Committee (SC), generally conceived as a group providing scientific direction, is a de facto requirement, even when it consists of all participating PIs. Working Groups (WGs), the "duct tape of collaboration,"[1] are called for and established each time a thorny issue is raised and can't be solved immediately by the larger group. Already overworked PIs are drafted to participate in these WGs, even though it means more time away from their science. Yet these WGs are rarely granted full decision-

[1] Charlotte P. Lee, personal communication, 2012

making power, instead being asked to make a recommendation back to the full SC, which then has to once again address the issue.

The amount of overhead inherent in this structure is staggering, though little evidence exists that it contributes to the achievement of the group's scientific goals. It is simply how collaborative biomedical research is done. While collaborative research may be the only way to address large, complex problems such as early detection of cancer or improved screening procedures for at-risk populations, the manner in which such research is currently being conducted has itself become a large, complex problem in need of a solution.

Today's large research challenges such as global climate change and cures for cancer can only be addressed in large, multi-site, multi-disciplinary collaborative efforts, as they require the input of scientists from disciplines as disparate as epidemiology, ecology, sociology, clinical medicine, molecular biology, population genetics, and veterinary medicine. Advances in communication technologies have allowed for collaboration over ever greater distances and for larger projects than ever before. Email, video chat software, secure file transfers, more nimble databases that can communicate with one another – all these technologies have increased the abilities of scientists to communicate and work together even when they are far apart geographically and culturally. Collaborative research is increasingly the standard method of doing science (Wuchty et al., 2007), yet working in multi-site projects is still difficult and carries substantial administrative costs (Cummings & Kiesler, 2007). In collaborative research, as in many other areas of our lives, the social process in which we engage have not quite maintained pace with the technical capabilities available to us. Put another way, the technologies upon which we rely to support large, distributed, multi-site projects do not currently support the social processes of doing science, processes that have evolved over many generations of scientists. Ackerman calls this the social-technical gap and defines it as "the divide between what we know we must support socially and what we can support technically" (Ackerman, 2000). While significant research has been done in the field of Computer Supported Cooperative Work (CSCW) on the work practices of scientists, these discoveries have not yet

been translated into technological solutions that can seamlessly support these practices, leaving scientists to muddle along.

So, if one size does not fit all, if each collaboration requires and consists of different elements in need of interconnectedness, then it makes sense to shift our attention to how those elements work together by looking at groups specifically charged with coordinating diverse collaborators. In this chapter, I will review what we know about facilitation and coordination in collaborative research and the current state of research on CCs.

2.1 The Social-Technical Gap

Ackerman argues for making the social-technical gap an "explicit intellectual focus" in CSCW research and suggests the rethinking of CSCW as a science of the artificial, as described by Herb Simon. In his book *The Sciences of the Artificial*, originally published in 1969, Simon calls for the creation of new sciences that address the nature of design, loosely defined as designing artifacts to attain goals (Simon, 1996). According to Ackerman, CSCW fits Simon's definition of a science of the artificial, noting "CSCW is at once an engineering discipline attempting to construct suitable systems for groups, organizations, and other collectivities, and at the same time, CSCW is a social science attempting to understand the basis for that construction in the social world (or everyday experience)" (Ackerman, 2000).

CSCW has been at the forefront of research on collaborative work, but its impact on practicing scientists has been minimized by our inability to implement the knowledge gained into easy-to-use, cost-effective and efficient systems. I argue that this is true not only of technological solutions but social solutions. The gap is not just social-technical but social-social, that we have not yet learned how to create more efficient and effective social solutions to support scientific work practices. In short, we have not yet learned how to offer alternative arrangements that solve the problems of collaborative research. To be clear, this is not a

failure of the relatively young field of CSCW but, rather, a continuing challenge. What is needed, then, is a way to translate what we have learned about the problems of collaborating into practical solutions.

2.2 Facilitation and Coordination in Collaborative Research

The fields of CSCW and Science of Team Science (SciTS) have identified a number of factors around the facilitation and coordination of collaborative research that influence the work of collaboration. In this section, I discuss two key themes that emerge from this research; namely: (1) costs and outcomes of coordinative activities; (2) the challenges of doing collaborative research and the strategies people employ to work within such projects. These themes both touch on issues of tension within the collaborative environment between the needs and objectives of the individual sites and their interaction with the collaborative structure.

2.2.1 Costs and outcomes of coordinative activities

Cummings and Kiesler (2005) describe the high costs of coordination of multi-disciplinary, multi-institutional research projects, referencing van de Ven's 1976 definition of coordination as "the integration or linking together of different pieces of a project to accomplish a collective task" (Cummings & Kiesler, 2005). I contrast coordination with collaboration, here defined as working together on the project itself, performing the "collective tasks" being coordinated. In this study, Cummings and Kiesler used the following items to measure coordinative activities via questionnaire: "direct supervision of work; use of special events, such as workshops, to get people together in the same place; travel in order to work together or meet; and regular use of face-to-face meetings, email, and telephone" (Cummings & Kiesler, 2005). They report that multi-institutional collaborations experienced fewer positive outcomes than did multi-disciplinary collaborations when compared to those collaborations with fewer institutions and disciplines, respectively. In other words, integrating scientists across disciplinary boundaries was less burdensome than integrating scientists across institutional boundaries. They noted that coordination mechanisms "slightly reduced" the disparity experienced by the multi-institutional projects. Further, they note that "the existing literature provides no clear guidelines to managing coordination and relationship

development in multi-disciplinary collaborations" (Cummings & Kiesler, 2005). PIs involved in multi-institutional collaborations spent a great deal of time and effort on coordination activities, but felt the funding agency (here, the National Science Foundation (NSF)) did not recognize those activities as important or necessary.

In an extension of this work on a different NSF grant program, Cummings & Kiesler (2007) analyzed 491 multi-university NSF-funded collaborations, again investigating how a collaboration's coordination activities affected its project outcomes. Coordination activities were broken down in to five areas: (1) Division of responsibilities, such as explicit assignment of and responsibility for tasks; (2) Shared resources like web portals and shared instruments; (3) Knowledge transfer, including rotation of post-docs through member laboratories and writing papers or grants together; (4) Meetings at least monthly and face-to-face; and (5) Communication technology use such as video conferencing or email. Project outcomes included: (1) Knowledge outcomes, such as published manuscripts, patents or awards; (2) Tools outcomes, including new software or hardware; (3) Training outcomes, defined as a graduate student from the group finishing a dissertation or receiving an academic job; (4) Outreach outcomes, including forming partnerships or collaborations with non-academic groups like industry or local schools; (5) Collaboration outcomes, such as continuing the collaboration beyond the funding or sharing data; and (6) Leverage outcomes, such as the research itself, not just the collaboration, continuing beyond the funded period or receiving new funding.

Through statistical analyses of all these activities and outcomes, Cummings and Kiesler concluded that a larger number of universities involved in the collaboration predicted fewer knowledge, tools, training and leverage outcomes. They also found that the most powerful predictor of project outcomes was knowledge transfer, which was associated with all types of outcomes. When more universities were part of the project, the lack of coordination activities explained the lack of outcomes, especially in the areas of division of responsibility and knowledge transfer. It appears as though the more universities that were involved, the less defined these coordination activities were. The existence of prior relationships and a

history of joint collaboration predicted better coordination activities, and most multi-university projects in this study lacked that characteristic. However, they also found that a "collaboration plan" as part of the grant was not a substitute for those existing relationships. Cummings & Kiesler's findings lend substantial support to the notion that coordination is a crucial part of any successful collaboration by emphasizing the value of planned and executed coordinative activities. These findings also give us specific areas of collaboration support upon which to focus.

2.2.2 Collaborative Work and Strategies

Collaborative work takes many forms. One of the most common, and of interest in this work, is what are commonly called Virtual Organizations (VOs). In a 2008 report from an NSF workshop on building effective Virtual Organizations, VOs are defined thus:

> A virtual organization (VO) is a group of individuals whose members and resources may be dispersed geographically and institutionally, yet who function as a coherent unit through the use of cyberinfrastructure (CI). A VO is typically enabled by, and provides shared and often real-time access to, centralized or distributed resources, such as community-specific tools, applications, data, and sensors, and experimental operations. A VO may be known as or composed of systems known as collaboratories, e-Science or e-Research, distributed workgroups or virtual teams, virtual environments, and online communities. VOs enable system-level science, facilitate access to resources, enhance problem-solving processes, and are a key to national economic and scientific competitiveness (Cummings & US National Science Foundation, 2008).

Although this definition includes the use of CI as a characteristic of VOs, others consider CIs to be a subset of VOs focused strictly on data-intensive science, while maintaining that VOs can support any type of work. In either case, working within such large, non-centralized organizations brings about many challenges for everyone involved. In this section, we note a few of those challenges; specifically, we discuss those associated with coordination and facilitation.

Katherine Lawrence (2006) introduces five tensions that offer another lens through which to examine collaboration, allowing us to see how collaborative research requires balancing the views and needs of many stakeholders. While studying a collaboration between scientists and software engineers developing a digital environment for atmospheric research, Lawrence identified five tensions as "paradoxical challenges" faced by a distributed collaboration: "First, the team had to allocate their energies to a blend of research and development. Second, the team members needed to encourage harmony while welcoming the benefits of conflict. Third, decisions sometimes were made by consensus and other times required a top-down mandate. Fourth, the project team had to cope with multiple modes of communication among the large number of individuals involved, finding a way to share sufficient information while not overwhelming themselves. Finally, the project had to move forward at a pace that was fast enough to get the job done but slow enough to fit the realities of the project" (Lawrence, 2006). She notes that these tensions are not a matter of choosing one way or the other, but of balancing between the two. Managing these tensions requires a significant amount of coordinative work on the part of the project team. It is important to note that these are not technological or scientific challenges being described by Lawrence, but, rather, organizational and coordinative challenges.

Ribes and Finholt (2007) also address tensions inherent in collaborative research, focusing on the difficulties of sustainability and planning for cyberinfrastructure. Their article uses data they gathered while studying four NSF-funded projects in the earth and environmental studies arena: (1) Geosciences Network (GEON); (2) Linked Environments for Atmospheric Discovery (LEAD); (3) Water and Environmental Research Systems (WATERS); and (4) Long-Term Ecological Research (LTER). What Ribes and Finholt found was that they could identify commonalities across the four projects in terms of the tensions experienced by participants. They developed three "scales of infrastructure" and three "concerns." The scales of infrastructure refer to the range of participants' work when involved in large-scale CI projects. These include: (1) institutionalizing, which refers to the necessity of long-term institutional arrangements for the project's work; (2) organizing work, which the authors describe as

thinking about the "kinds of work" the group must do to ensure long-term sustainability; and (3) enacting technology, or moving from the development of proof-of-concept systems to systems which are stable and support the long-term goals of the project. The concerns, on the other hand, are issues with which participants must grapple when thinking about the long-term sustainability of CI projects. The three concerns are: (1) aligning end goals, or how to balance multiple, often competing goals within a project; (2) motivating contribution, which refers to ways of ensuring long-term commitment of team members; and (3) designing for use, or ensuring that the developed infrastructure actually meets the needs of the end users. The intersection of each scale of infrastructure and concern produces one of the nine tensions (Table 2.1) (Ribes & Finholt, 2007).

Concerns / Scales	Institutionalizing	Organizing Work	Enacting Technology
Aligning End-Goals	Project vs. facility	Planned vs. emergent	Inclusion vs. readiness
Motivating Contribution	Individual vs. community	Development vs. maintenance	Research vs. production quality systems
Designing for Use	Communities vs. constituencies	Research vs. development	Today's requirements vs. tomorrow's users

Table 2.1: Ribes & Finholt Tensions

The authors conclude by noting that their goal in this paper is to make visible the choices CI projects make on an ongoing basis, not to cast those choices as right or wrong ones. By presenting this framework, Ribes and Finholt seek to foreground the tensions inherent in CI projects that might otherwise go unremarked and unexamined.

Lee et al. have described the "human infrastructure" of a CI project as "the arrangements of organizations and actors that must be brought into alignment in order for work to be accomplished" (Lee, Dourish, & Mark, 2006). One of the main contributions of this work was their finding that the overarching organizational structure of the collaboration under study was not at all clear to most of those involved, if to anyone. In fact, most participants had very little understanding of how the collaboration was structured, focusing instead on their own local work and the parts of the structure in which they were direct participants. The authors note that "[collaboration] members know what part of the infrastructure they need to tap into to coordinate, get information, or to perform a task" (Lee et al., 2006). However, this

evolving, “multimorphous” structure was challenging for managers and those who were charged with getting the work done. These managers reported frustration at their lack of a way to hold collaborators accountable.

Bietz et al extended the idea of human infrastructure to include what they called “synergizing,” which is the “work that developers of infrastructure do to build and maintain productive relationships among people, organizations, and technologies” (Bietz, Baumer, & Lee, 2010). The authors found several strategies that CI developers use to accomplish synergy in their projects, including leveraging and aligning. Alignment is the “work that developers do to enact a relationship in a way that enables it to produce, and to function within, the nascent cyberinfrastructure” (Bietz et al., 2010). Alignment takes place among a wide variety of entities within the collaboration, from individual workers to institutions, and helps bring everyone involved into a relationship where all parties are working toward the same goal. Leveraging, on the other hand, is “using an existing relationship with a person, artifact, or organization to build or strengthen a productive relationship with another person, artifact, or organization” (Bietz et al., 2010). These two sub-processes of synergizing both serve to move the collaborators forward by using the human infrastructure.

These articles make it clear that collaborative research involves great balancing acts, yet they don’t make it clear who is responsible for balancing the various interests. With the exception of Bietz et al’s synergizing, these articles also say little about the actual day-to-day work practices that go into balancing the tensions, much of which falls into the realm of coordination and facilitation. The literature addressed in this section covers a variety of understandings about collaborative processes, including explicit research on coordinative activities, as well as strategies for improving collaboration that implicitly require coordination.

2.3 Coordinating Centers

Collaborative research is understandably difficult and can add high overhead to a scientific project, yet scientists are being pushed to do more of it with little extra support. This additional overhead can slow research down, which means wasted money, lost opportunity, and frustration for scientists. A CC is one tool that can help offload some of the administrative burden from investigators. A well-built Coordinating Center can ameliorate some of the overhead and offload some of the burden from researchers by managing the administrative aspects, facilitating collaborative activities, and empowering investigators to focus on the science, thus improving every stage of a study. As a result, funded projects run more smoothly and are more likely to reach their scientific goals, creating a greater return on a funding agency's investment. A good CC will have the available expertise and resources to facilitate protocol development, ensure timely information exchange, and coordinate data management and statistical analysis. CC staff will also take the lead on bringing all parties to the table and ensuring all participants have an equal voice in the areas of the project as appropriate to their expertise. It is these "soft" areas of research that become increasingly important, even mission critical, in collaborative projects and which receive the least attention from research teams.

Although it is tacitly recognized that a good CC is essential to the success of any multi-site collaborative project, very little study has been done on what makes a CC successful, why some CCs fail, or how to build a CC that meets the needs of a given project. Moreover, very little published guidance is available, as few CCs outside the clinical trial realm write about their work [see, for example, (Curb et al., 1983), (C. Meinert, Heinz, & Forman, 1983), (Collins et al., 2003), (Blumenstein, James, Lind, & Mitchell, 1995)]. CC directors are largely forced to reinvent the process through trial and error with each new collaboration. This wastes not only precious funds but also time, delaying the achievement of scientific goals. Without well validated best practices, it is impossible for a CC manager to be sure s/he is not only avoiding the worst mistakes of the past but is maximizing resources by running a CC as effectively and efficiently as possible. It is notoriously difficult to measure and compare effectiveness and efficiency in

scientific projects, as their aim is generally to create new knowledge. This level of difficulty in achieving this aim varies substantially from project to project, and it is challenging to separate the organizational accomplishments from the scientific ones. Thus, for our purposes, we conceive of effectiveness in a CC as its ability to move the CCEN successfully toward its scientific goals. Efficiency is here conceived of as a CC saving the collaborating PIs time by doing work the PIs and their local staff might otherwise need to do themselves.

2.3.1 What We Know about CCs

In the mid-1970s, the National Heart Lung and Blood Institute (NHLBI) began a project called Coordinating Center Models Project (CCMP) in an attempt to better understand CCs in clinical trials (Symposium on Coordinating Clinical Trials). At that time, clinical trials were still a fairly new method of doing research and large amounts of money were being spent to coordinate those trials. Yet very little was known about what made a good CC or how to run a CC most effectively. To address these issues, a CCMP research team was designated. This team was made up of scientists who were interested in the design and implementation of clinical trials. Their methodology consisted of a survey of those involved in six NHLBI-funded clinical trials, as well as interviews with key staff members. The results were reported at a conference in 1978 and published soon after (Symposium on Coordinating Clinical Trials).

One of the key findings of the CCMP was that it was not possible to identify a common set of activities across the CCs (Symposium on Coordinating Clinical Trials). The research group concluded that this meant that there was no one model of a CC. Interestingly, they don't seem to have considered the possibility that the great variation in activities and attitudes stemmed from the fact that CCs represented a new organization model with no blueprint for its activities and that CC leaders were simply creating policies and procedures as they went along in reaction to the events around them. Perhaps the variation could be traced exclusively to the lack of both standards for running a CC and communication among CC leaders.

Soon after the CCMP report was published, several clinical trials published articles about their CCs. These were not empirical studies but, rather, reports written by the CC and clinical trial leadership detailing how their CC worked, including a list of the activities for which the CC was responsible, as well as assessments of any issues or problems and any particularly interesting solutions the CC may have devised for working in a clinical trial. While the articles go into vastly different levels of detail about what a CC should do, all stressed that the primary responsibility is to ensure the quality of the science. Blumenstein et al. (1995) describe the CC's primary mission thus: "to assure the validity of study findings that eventually will be disseminated in publications and public presentation" (Blumenstein et al., 1995). Going into slightly more detail, Mowery and Williams (1979) write that monitoring the implementation and adherence to protocol are the primary responsibility of the CC (Mowery & Williams, 1979). Rifkind (1980) adds delivery of results to the community in a timely and high quality manner (Rifkind, 1980).

The specific responsibilities listed by these authors vary widely, ranging in level of detail from "statistical and content methodological support" (Bangdiwala, de Paula, Ramiro, & Muñoz, 2003) to "ordering study medications" (C. Meinert et al., 1983). Some divided responsibilities into categories, most of which are common in theme, if not specific name. These categories include (1) statistical coordination and management, (2) study coordination, and (3) administrative and secretarial support. The first category of responsibilities include those around data, including data management and analysis, as well as monitoring data collection and performing quality assurance (see, for example: (Curb et al., 1983), (Bangdiwala et al., 2003), (Blumenstein et al., 1995), (Margitic, Morgan, Sager, & Furberg, 1995), (Greene, Hart, & Wagner, 2005), (Lachin, 1980), (Berge, 1980), (C. Meinert et al., 1983), (Winget et al., 2005)). The second category of responsibilities involves study coordination, including duties such as protocol and form development, monitoring adherence to the protocol or performance monitoring, computer system development, training of staff, documentation and archiving of study information, communications, adherence to institutional policies, reporting, allocation of CC resources and manuscript preparation. Administrative and secretarial support included functions like fiscal management, meeting and site visit

organization, budget preparation and management, securing equipment rentals, and personnel management, as well as general secretarial support (Curb et al., 1983), (Bangdiwala et al., 2003), (C. Meinert et al., 1983). These last two categories were sometimes conflated into one, but the duties remained the same.

One of the most interesting perspectives published during this time period was the 1980 publication of "Perceptions of the Coordinating Center," a set of papers written from the perspective of different players in a clinical trial, including a study site PI (Prout, 1980), a clinic coordinator (Overton, 1980), the project officer (Rifkind, 1980), the advisory board (Berge, 1980) and a site visitor (i.e., evaluator) (Sedransk, 1980). These papers paint a vivid picture of the conflicts inherent in running a CC. While these different perspectives led to emphasis on different aspects of the CC's roles, what was most interesting was the problem areas raised by the different parties. Prout, the study site PI, warned CCs to pay close attention to fiscal management and not "featherbed" the budget. He also raised the question, discussed further below, of the long-term sustainability of a CC and emphasized that statisticians need time and data access in order to develop new statistical methods for clinical trial data analysis (Prout, 1980). The clinic coordinator's essay focuses on types of support the CC can give to clinical staff, including usable forms and computer systems, railing against "cold" error messages spit out by the computer systems (Overton, 1980). She also demands someone at the CC be in charge, so clinic coordinators have one person they can contact should they have problems. This focus on the minute details of the study stands in contrast to Rifkind's project officer viewpoint, which demands that the CC focus on the big picture, including an imperative to "assure maximum utilization of the data being collect" (Rifkind, 1980). Rifkind also stressed the importance of communications, including both formal communications like reports and informal communications such as phone calls between staff members. Representing the point of view of the advisory board member, Berge (1980) notes the "complex" relationship an advisory board has with a CC. Berge then goes on to list the accountabilities the CC has to various parties, such as the sponsor, the parent institution, the major committees of the trial, the FDA, the policy board and the scientific

community (Berge, 1980). There can be great difficulties, Berge notes, if one of these entities is requesting action or information that affects the budget and schedule. Finally, the site visitor writes about the difficulties of evaluating CCs (Sedransk, 1980).

One of the overarching themes raised in all of the papers published on CCs is that of the difficulties in balancing the CC's responsibilities to multiple parties. These difficulties are compounded by another theme, that of the difficulties of staffing a CC. CCs are expected to have on-staff expertise in a wide range of activities, including administration, statistics, federal regulations, human subjects, technology and organizational development. At the same time, the CC's organizational structure is expected to evolve over the course of the project in response to changes in the work, while minimizing costs. At a workshop at the CCMP kickoff in 1977, the group reported:

> "One major managerial problem has to do with the establishment of a large, well-trained staff and whether personnel should be retained or transferred out once a study is terminated. Many university-based coordinating centers are locked into the cycle of maintaining these staff positions and have invested much time and effort in staff training in order to fulfill their function. Frequently the only way personnel can be retained is to proceed directly into another study. Since this option is not always available, there is a clear danger in creating too large a coordinating center within a university setting" (Meinert & Coordinating Center Models Project).

This staffing difficulty is even more challenging given the current financial climate and budget cuts at NIH. Finding funding to support the infrastructure of a CC, as opposed to funding a CC for a specific project, is thought to be virtually impossible (Potter, personal correspondence). This leaves CCs with the dilemma of rapidly staffing a new project with experienced staff.

Curb (1983) and Blumenstein (1995) both noted one of the major problems of running a CC is the time crunch inherent in such a project. Once funded, CCs are expected to get the project up and running quickly, with little attention paid to the set-up phase. These papers argued that more time spent on securing agreement on organizational issues such as data sharing agreements, authorship policies and communication expectations, as well as scientific issues such as common data, survey forms and required technologies would have made the project smoother, and, thus, produced better science more quickly (Curb et al., 1983). CC managers also noted that more time for close-out and staff time to support

manuscript writing at the tail end of the projects would have, similarly, led to even stronger outcomes for the project (Blumenstein et al., 1995).

There is a great variety in the organizational models followed by the different CCs described in the literature. Blumenstein (1995) et al. described several different models of clinical trials and several different models of CCs, though no discernible pattern for matching these was described (Blumenstein et al., 1995). Curb et al (1983) noted that "[t]he role of a coordinating center in a multicenter clinical trial varies with the particular design and organization of each trial" (Curb et al., 1983). The implication of this is that the organizational structure of the CC must also be a consequence of the trial it supports. Curb also asserts that responsibilities, and, therefore, the staffing makeup, of the CC must shift as the trial progresses through the phases.

Thus, the literature on CCs is lacking a comprehensive model of what different kinds of CCs look like, how they are formed, how they should be managed or even what impact they have on the projects they are coordinating. Furthermore, the projects that are being coordinated are structured in many different ways, with little understanding of what types of CCs might work best for these different types of projects. In short, we know very little about how CC or the projects of which they are a part function.

2.3.2 Coordinating Center-Enabled Networks

As discussed briefly in Rolland et al. (2011), there are different types of CCs (Rolland et al., 2011), including those that are organized around a lead investigator using project funds set aside for coordination, and those that are separately funded under independent RFAs from the funding agency. The SN and BN CCs studied in this research belong to the latter group, as they are funded entities in and of themselves. Yet they exist only within the concept of their networks; that is, a coordinating center must be coordinating something. In this context, such networks are a special kind of VO, one that I call Coordinating Center-Enabled Networks (CCENs). As a part of these networks, CCs are both separate and integrated at the same time.

As part of these RFAs, the CCs are required to have their own specific aims (BN RFA, SN RFA). These aims are not only organizational, such as facilitating collaboration, but also scientific, such as developing novel biostatistical methods for the discovery of biomarkers. While not the network's "leader" by any stretch, they are charged with taking a "leadership role" in the collaborative efforts of the network. They are expected to facilitate all activities that involve more than one site. In practice, they are often asked (or offer) also to assist with projects being conducted by the individual sites, but this is generally outside of their scope (BN and SN RFAs). These CCENs are not uncommon in the world of biomedical research, yet there are no guidelines that exist for how to build or run a CC or even a common definition of what a CC does. To the best of my knowledge, the research reported here is the first project to address such questions. My aim in doing so is to take some of the guesswork out of building and running a CC by providing some guidance to CC staff and PIs on how best to engage in the coordination of CCENs.

2.4 Articulation Work

Articulation Work (AW) has been commonly defined as "the work of making work go well." Strauss (1988) and Corbin & Strauss (1993) have similar definitions of articulation work which focus on the coordination of tasks to keep work flowing (Corbin & Strauss, 1993; Strauss, 1988). Both articles focused on articulation work as interaction among different parties in service of agreeing on how to align their work to achieve their work goals, as perceived by the workers themselves. Strauss emphasizes that articulation work is but one component of the "articulation process," which includes "putting *all* the work elements together *and* keeping them together" (Strauss, 1988, emphasis in original). He further notes several interactional processes involved in articulation, including persuasion, teaching or negotiating with others about the value of one's project, manipulation and coercion. In their 1993 work, Corbin & Strauss refine this notion of interactional processes to include discussions of "working things out" and "stance." Working things out refers to establishing work arrangements and keeping them going. A worker's stance describes his or her position toward the process of working things out. Interactional strategies can be thought of as the tools workers use to accomplish the necessary articulation work.

One of the key principles of articulation is the agency of workers. Strauss (1988) notes that all workers have responsibility, formal or informal, for negotiating the arrangements of how work fits together and gets done. Corbin & Strauss talk more explicitly about the power that actors have in the workplace to develop arrangements. They note, "[a]ctors have the ability to manipulate, use to their advantage, avoid, or in other ways respond to conditions. It is their action taken in response to conditions that gives conditions their (analytic) meaning" (Corbin & Strauss, 1993). Fundamentally, AW focuses on the process of workers doing things in pursuit of an objective.

Articulation also focuses on the structural and organizational conditions that contribute to the process of making arrangements. Workers have more or less power to affect the arrangements, often based on both historical and anticipated future developments of the organization and of the larger world. These arrangements are never permanent, even when ensconced in formal organization policies, but are subject to continuous reworking due to changing conditions and unexpected incidents. Articulation of work is fundamentally a social process, based on negotiations among parties in response to the world around them.

Gerson (2008) further refines the notion of AW into *metawork* and *local articulation work* (Gerson, 2008). Gerson's focus is on describing the coordinative work involved in distributed organizations, using the term *reach* to refer to "the distribution of tasks across organizational, spatial, and temporal boundaries" (Gerson, 2008). It is within this context that he notes that AW has been used in two key senses, and seeks to refine those definitions. Gerson writes:

> On the one hand, articulation work is about making sure all the various resources needed to accomplish something are in place and functioning where and when they're needed *in the local situation*. This means bringing together everything needed to accomplish a task at a particular time and place, including all the administrative and support functions such as janitorial services, food service, equipment maintenance, and covering for staff out sick or on vacation. The concern and emphasis in this sense are on particular situations rather than classes of activity.
>
> In its second sense, articulation work means "putting together tasks, task sequences, task clusters-- even aligning larger units such as lines of work and subprojects-- in the service of work flow" (Strauss 1991: 100). In this second sense, the focus is not so much on the specifics of work in a particular local situation,

> as it is on making sure that different *kinds* of activity function together well. The two senses, of course, overlap heavily-- especially when all the tasks are part of the same organization and are carried out in the same place ((Gerson, 2008) Emphasis in the original.).

It is to this first usage that Gerson assigns the term local articulation work and to the second, the term metawork. He further emphasizes that the local articulation work is "about a particular situation" (Gerson, 2008).

When this work is all being done within a local organization, Gerson notes, this distinction between local articulation work and metawork is not particularly important. However, when the work is distributed across multiple organizations, it becomes more so, as the work becomes more complex and more reliant on the connections being made between different kinds of work. "The importance of increasing reach however, means that we must make the distinction clear and understand its implications. The work of specifying the work to be done is one thing; the work of ensuring performance in specific circumstances is another" (Gerson, 2008). In other words, as the work becomes spread out over multiple organizations, the work of bringing it back together again into a functioning whole is increasingly important.

Chapter 3: Site Description & Methods

3.1 Site Description

This research focused on the work of Coordinating Centers (CCs) in multi-site cancer epidemiology projects. As discussed in the introduction, a CC is a central body tasked with coordination and operations management of a research project. Its main purpose is to facilitate collaboration among the diverse sites with the aim of completing all research objectives and producing impactful science. CCs form in a variety of ways, ranging from developing organically around a lead investigator of the project to being directly funded as a distinct Request for Application (RFA), which is how NIH advertises funding opportunities, from a funding agency (Rolland et al., 2011).

This research took place at the Fred Hutchinson Cancer Research Center (FHCRC), a National Cancer Institute (NCI)-designated cancer center in Seattle, Washington. This class of research institute is categorized by a high level of scientific excellence and patient-centered research. FHCRC has approximately 3,000 employees and is organized into five divisions. One of these divisions is the Public Health Sciences (PHS) Division, home to most of the organization's cancer epidemiologists. In 2011, FHCRC ranked 19th in total NIH funding, receiving in excess of $270 million, ahead of much larger institutions ("National Institutes of Health Research Portfolio Online Reporting Tools (RePORT)," 2012), and 3rd in funding from the NCI. I am an employee of FHCRC and work primarily on cancer epidemiology projects. I have not worked on either of the projects described here, though I work very closely with two of the PIs, Adam and Nigel.

FHCRC sits on a 14-acre campus overlooking Seattle's Lake Union, roughly halfway between downtown Seattle and the University of Washington campus. The FHCRC campus was built relatively recently, with the planning for the move from another area of Seattle to the current location beginning around 1992, according to an FHCRC brochure on the campus (FHCRC Campus Sustainability). Prior to that move, employees were scattered at multiple locations around Seattle. The campus development focused on both

creating spaces that really worked for modern science and creating sustainable buildings. In fact, the FHCRC campus has won more than 35 awards for its sustainability practices. The PHS Division is housed in the Arnold Building, one of about a dozen buildings on the FHCRC campus. The majority of this building is composed of offices, with several labs housed on the top floor of this 5-story building. The PHS director at the time of the building's construction insisted that virtually all work spaces be offices with few cubicles, to ensure quiet work spaces for all employees, regardless of rank. Offices surround a central, light-filled atrium with a spiral staircase. Dozens of conferences rooms of various sizes also can be found scattered throughout the Arnold Building.

FHCRC was selected as the research site for this project for several reasons. As an employee at FHCRC, I have access to a wide variety of CCs to study. Many of these CCs have been in existence for over 10 years, with strong records of success and scientific excellence as demonstrated by high-profile publications and repeated funding. Additionally, FHCRC itself is a leader in cancer epidemiology research, hosting many of the most successful multi-site projects in the country. Capturing this experience and expertise is crucial to understanding CCs. FHCRC's standing in the scientific community as a paragon of scientific excellence also gives greater credence to the results of my study, as the reputation of the Hutch positively affects my report.

I selected two CCs at FHCRC to follow intensively for this project. The first CC is part of the Screening Network (SN), a group focused on understanding the screening process for prostate, breast and cervical cancer. The second is the CC for the Biomarker Network (BN). The BN's mission is to identify biomarkers for use in the early detection of cancer and prognosis. Both of these CCs are housed within the Science Facilitation Team (SFT) program at FHCRC, so they have both staff and PIs in common. The SFT organization grew out of the group's experience coordinating a complicated chemoprevention trial that began in the early 1980s. Since that time, they have coordinated several multi-site projects.

While some CCs grow out of the lab or group of the lead PI, both the BN and SN CCs are funded as independent entities within the larger research program. That is, they were formed in response to a separate RFA and independently funded. As such, the CCs have their own specific aims and goals that must be met. The CCs are not simply tasked with coordinating the projects but also contribute to the science by developing novel statistical methods for analyzing the project data and by providing scientific leadership to cross-site projects and resources. These cross-site projects and resources include, for example, BN's lung team project which involves several BN member sites interested in lung cancer, or the SN data repository.

These two CCs were selected for several reasons. Given that they share some staff and PIs, they present an interesting opportunity to compare two different CCs while, in effect, partially controlling for one of the major variables in the study of science: namely, the people. Obviously, the CC is only one small part of the collaboration and a small portion of the people involved, but even given that, comparison between the two projects should yield interesting insight into how an experienced CC operates in different contexts. As such, I was able to investigate whether that expertise and experience translate from one project and context to another. The answer to this question has substantial ramifications for the field of scientific coordination. If a group with experienced staff is unable to translate their practices successfully to a new project, it calls into question whether such expertise and experience can really be used to develop resources for other groups. Further, the two projects are at very different stages in their lifecycles, allowing for interesting comparisons between the coordination of a nascent collaboration and one that's been working – and successful – for more than 10 years. Finally, my professional relationship with the two PIs allowed me easy access and a warm welcome. I have worked with both Adam and Nigel, two of the FN/SN PIs, for more than four years, and they have both been very supportive of my research.

3.1.1 The Biomarker Network

The Biomarker Network (BN) was first funded in 2000 and aims to "discover, develop, and validate biomarkers for early cancer detection, risk assessment and the molecular diagnosis and prognosis of early

cancer" [BN RFA]. The BN is funded via a U24 mechanism, a cooperative agreement similar to that of SN, below. This type of funding indicates a high level of involvement on the part of staff at the funding institute; in this case, NCI. What this means in practice varies widely by Program Officer, but generally means the PIs keep in regular contact with the PO throughout the year, as opposed to simple updates via annual progress reports. Currently, there are five Biomarker Reference Laboratories (BRLs), eight Clinical Validation Centers (CVCs), and 22 Biomarker Developmental Laboratories (BDLs) covering breast and ovarian, colorectal, lung, pancreas and prostate cancers. Each of these grantees may consist of multiple PIs at multiple institutions. Each has its own research objectives, in addition to its responsibilities to the network. There is also an Informatics Center housed at another institution. (See Figure 3.1: Organizational Structure of BN). The BN was recently refunded for an additional 5-year period in a competitive renewal process. The CC's first-year budget is just over $1 million (direct), while the budgets of the other components vary significantly. The BRL RFA allows for direct costs up to $300,000 per year, while the CVC and BDL RFAs allow for direct costs up to $600,000 per year. This is a very expensive research program with a large number of moving parts that need to be coordinated.

As with all NIH funding, BN grants are legally made to the lead PI's institution; however, in practice, grants are made to investigators and their labs, with one designated as the primary granting institution. For example, an application to the BRL program might include multiple PIs from multiple labs, but one investigator and his/her institution must be the primary institution, with other participating PIs receiving funds through subawards. Each of the PIs listed on the grant proposal is included as a Member of the BN and, thus, member of the Steering Committee, with one vote each. The NCI Program Coordinator and the PIs of the CC are also automatically part of the Steering Committee, which the RFA charges with the "scientific management and oversight of all Network activities, including monitoring the activities of the CC" (BN RFA). The BN Manual of Operations further defines the responsibilities of the SC, quoted verbatim:

- Develop guidelines for operating the BN

- Coordinate the research program within the BN
- Develop criteria for reviewing progress of the BN
- Establish and track milestones
- Develop and implement rules for sharing data and resources
- Disseminate information on the availability of resources (tissues, new technologies, and patients) within the BN
- Develop criteria for selecting an Associate Member of the BN
- Develop criteria on the use of the Core Funds
- Develop and approve protocols for clinical research through the BN
- Develop criteria on evaluating and reviewing data on potentially promising new biomarkers
- Prepare annual Progress Reports for submission to NCI at the end of each fiscal year

(BN Manual of Operations accessed online July 27, 2012).

From this list, we can see that most of the scientific management takes place at the Network level, under the watch of the SC. There is also a Network Consulting Team (NCT), which is an external advisory board whose members are selected by NCI. The NCT provides oversight of the entire BN program.

In contrast to the list, above, the areas of responsibility of the CC, as laid out in the RFA include:

- Network coordination, including such activities as logistical support for meetings and calls, website development and management of communications;
- Data management and protocol development, which encompasses responsibility for supporting the development, implementation and conduct of BN-collaborative research protocols; statistical analysis for clinical validation studies; and assisting with data collection for pooled data projects;
- Theoretical and applied statistical research, including conducting research on novel statistical methods and computation tools for the Network;
- Validation information systems and services, which includes the CC working closely with the informatics team at JPL, as mentioned above.

These responsibilities cover a wide swath of the scientific process, from administrative management to theoretical statistical methods development, allowing the CC to have input into virtually every major decision made in the project. This wide variety of responsibilities also implies a wide variety of skills of the CC staff members, ranging from project coordination to human subjects work, computer programming and statistical analyses. The BN RFA notes that the "CC must have expertise and capabilities in biostatistics, information technology, data management, protocol development, and logistical support. As the number and types of Network validation studies will vary during the five year funding period, the number and identity of CC personnel should change in response to the scientific opportunities" (BN RFA). As discussed in the literature review section, above, this requirement for such a

variety of talent is an enormous challenge for CCs as they must have access to such, often specialized, expertise on a sporadic – and often instantaneous – basis of indeterminate length while being unable to offer long-term support to those experts.

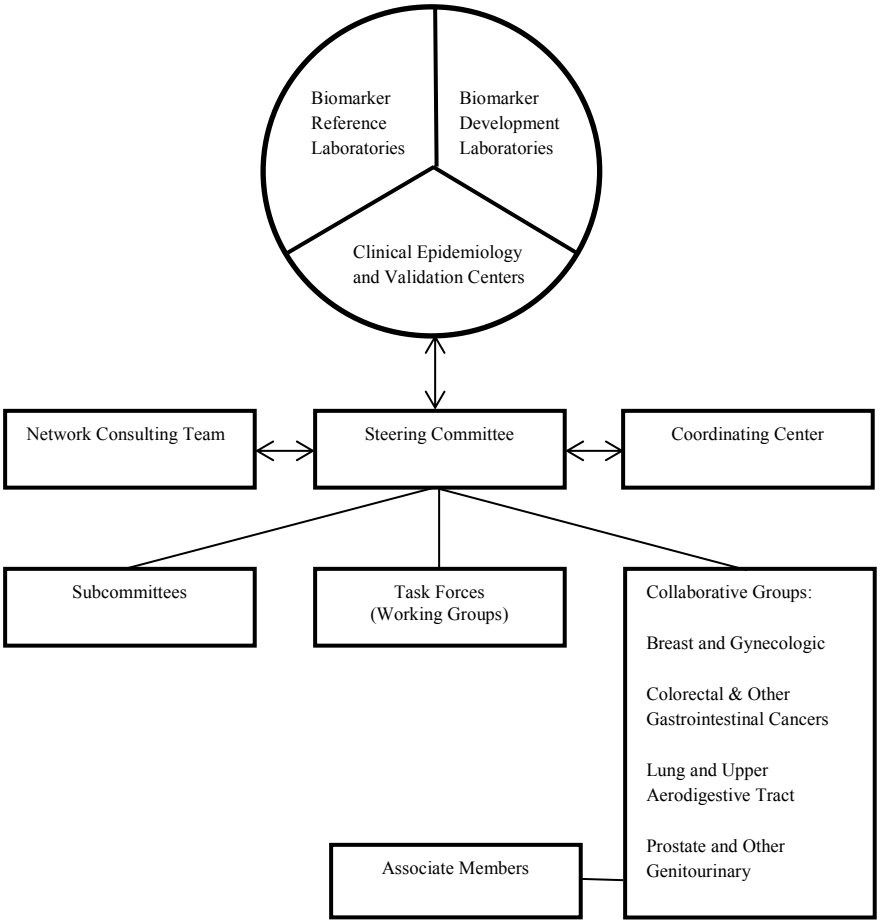

Figure 3.1: Organizational Structure of BN (adapted from BN website)

3.1.2 The Screening Network

Funded in Fall 2012 for a 5-year period, SN's goal is to document the screening processes for prostate, cervical and breast cancer, identifying issues in the processes that lead to negative outcomes such as undetected cancers or cancers detected too late to treat. There are seven SN Research Centers (RCs) around the country, including one focused on cervical cancer, three on prostate cancer and three on breast cancer. (See Figure 3.2 for the SN organizational structure.) One of the defining characteristics of SN is that participating sites must have access to medical records of defined populations, which means that sites

are either large HMOs or regional medical providers. As such, some of the investigators have considerable experience in research consortia, whereas others appear not to have. Several of the investigators in the breast group are involved in a similar consortium called the Breast Cancer Surveillance Consortium (BCSC), as is one of the CC PIs.

Figure 3.2: Map of Screening Network Research Centers, Coordinating Center and Funding Agency (adapted from SN website)

SN is funded via the NIH U01 funding mechanism, also called "Research Project Cooperative Agreement." As with the BN U24 award described above, this type of funding indicates a high level of involvement on the part of staff at the funding institute; in this case, NCI. The budget for this project is around $1 million for the first year (direct costs), indicating a substantial investment in the CC. The RFA notes that NCI anticipated funding 9-12 RCS, with proposed budgets expected to range from $0.75 to $1.5 million per year (total costs). However, as noted above, only seven RCs were funded.

Like the BN, discussed above, SN is governed by the SN Steering Committee (SC). The SC is composed of two representatives from each of the RCs, the CC and NCI. Each entity has one vote. The RFA lists the following responsibilities for the SC, quoted verbatim:

- Overseeing the overall organization of the SN initiative and for reviewing its research goals;
- Developing the appropriate structure of Working Groups to promote the exchange of experiences, protocols, and ideas across the SN;
- Establishing advisory committees and subcommittees, as necessary, to ensure the progress of the individual SN as well as of the overall SN initiative;
- Coordinating trans-Network projects such as the standardization of terminology
- Reviewing the potential of shared support infrastructure(s) at individual SN Research Centers to serve the needs of other SN Research Centers or the entire initiative;
- Developing (jointly with Working Groups) the SN-wide requests for trans-Network Projects, and conducting their evaluation and selection;
- Developing and recommending progress report formats for individual SN pilot and cross-SN projects.
- Ensuring that the SN Research Projects and the Coordinating Center take advantage of existing NCI and NIH resources and programs;
- Making recommendation for re-directing or termination of SN pilot projects that become unpromising or unproductive; and
- Scheduling semiannual PD/PI's meeting at which all SN investigators will present their scientific progress and future plans (SN RFA)

In addition, SN has a Scientific Consulting Committee, which serves as an external advisory board. This RFA is a bit less explicit about responsibility for the scientific direction of the Network than the RFA of the BN.

The SN's CC RFA lists three main components required for the CC:

- Leadership and administrative core, which encompasses both the scientific, statistical and organizational leadership, as well as the coordination tasks such as organizing meetings;
- Central data repository/data management, which focuses on the pooling of data at the CC; and
- Dissemination/outreach, serving as a resource for the scientific community at large (SN CC RFA).

Both the second and third components are focused around the development of the central data repository. The RFA notes that "[t]he primary function of the [CC] will be to establish and maintain a Central Data Repository for the entire SN initiative" (SN RFA). This resource will not only serve the SN researchers but may also become available eventually to outside researchers via a data request policy.

3.2 Research Methods

In this research, I investigated the question, How do Coordinating Centers facilitate the network-level work of a Coordinating Center Enabled Network? To answer this large question, I identified the

distinguishing characteristics of a CCEN, described the work each CCEN component performs and defined what facilitation means in a CCEN.

At the most basic level, the overall aim of this research on CCs is to develop guidelines and best practices that will help CCs better support the process of coordination and support for collaborative biomedical research. I aim to do so by developing a deeper understanding of CC work practices and the challenges they face, with the hope of shaping recommendations for improved systems, both technological and social. This type of deep, nuanced, rich understanding, as it is embedded in local context, is best captured through the use of qualitative research methods.

3.2.1 Qualitative Research & Ethnographic Methods

In order to answer this research question, I gathered detailed data about how CCs work. I sought to understand how CC staff members think about their roles and responsibilities within the collaborative effort and how their collaborators think about the CC and its staff. In order to develop improved work processes and practices, I needed to develop a thorough understanding of current processes and practices in order to identify those aspects that are working well and those that are lacking.

Qualitative research aims to produce rich data that don't simply quantify or describe a setting from the outsider's perspective but present an understanding of that setting from the point of view of those on the inside. One popular approach to collecting qualitative data is through the employ of ethnographic methods. Ethnographic field research is devoted to the study of the everyday lives of people and groups of people through two distinct activities: (1) entering into the social setting and getting to know the people; and (2) systematically observing and recording the details of that setting (Emerson, Fretz, & Shaw, 1995). Two key tools of ethnographic research are field observations and interviews.

Field observations are recorded in the form of ethnographic fieldnotes. Emerson et al. (1995) describe fieldnotes as an "account *describing* experiences and observations the researcher has made while participating in an intense and involved manner" (Emerson et al., 1995, emphasis in original). They also

stress there is no one correct account of an event, because writing fieldnotes involves perception and interpretation of what the researcher is observing and experiencing. All experiences are filtered through the observer's own experiences and worldview as they are recorded. Ethnographic interviewing is another tool used to elucidate the processes of a group. As noted by Weiss (1994), through interviewing, we can learn "about people's interior experiences" (Weiss, 1994). Fundamentally, I seek to understand how CC staff members, as well as their external collaborators, experience the coordination of multi-site projects. In order to do so, I need a deeper, more holistic look at their work practices, their view on how they fit into the project, and what they feel others expect of them. The answers to these questions are best obtained through interviews that allow participants to speak in their own voices about their experiences.

One critical limitation to the use of qualitative research, in general, is the potential bias of researchers in observing and analyzing their findings. Qualitative researchers cannot describe everything they see and must choose carefully what to pay attention to and what to record. In doing so, qualitative researchers bring their own biases and theoretical and social backgrounds to their work. Of course, this is also true of quantitative researchers, despite frequent claims of "objectivity." In this particular research, my own biases as the project manager of a CC for an international research collaboration may color my view of the meetings as I observed and interviewed participants. I have my own views on how a CC should be run and how the CC staff should respond to issues and opportunities. In order to counteract that, I kept this potential bias in the forefront of my mind as I listened and observed. An additional limitation is that, as a single researcher, I was constrained in the amount of data I could collect and the breadth of my viewpoint. There were inevitably things I missed and things that other researchers would have captured. Finally, because I collected data in a very specific situation – two CCs with the same staff and PIs – there are limits to the generalizations I can make from my results. Through conversations with colleagues, all well versed in the studies of collaborative research, I sought to alleviate these final concerns by comparing what I've learned with their previous findings in other situations, as well as the previous findings of other CSCW researchers working in the area of collaborative science.

It is also important to note here that the majority of participants in this research project, the CC PIs and staff, are my coworkers at FHCRC. This raises important ethical questions of bias, as well. While these groups do not provide any funding for my work in general, or this project specifically, it is still in my best professional interest to treat my report of their activities carefully. While this is true of all researchers, in that the goal should never be to destroy anyone with an overly savage report, it is especially true for me, given my relationship with my colleagues. Again, I have kept this bias in mind as I have analyzed my data, seeking to present the participants in a way that remains true to the data without damaging anyone's reputation.

Finally, my role as a participant-observer has implications for the analyses presented here. While I was not a participant in these CCs in the traditional sense, my status as a "member" of the general world of CC management and of FHCRC made me more of an insider than other research might have been. The full effect of this insider status is difficult to evaluate systematically, but it is likely that CC staff and PIs were more open and forthright with me than they might otherwise have been with another researcher. It is also possible that they talked about their work with me, a CC project manager, in a less superficial way than they might have with someone with no experience in CCs or collaborative research. The flip side of that trust and openness is that they may have occasionally shared thoughts with me that could be damaging to the project if shared publicly. As noted above, I have made every attempt in this dissertation to treat their trust in me with the greatest of respect and to strike a balance between protecting the privacy of my participants and reporting my research accurately.

3.3 Data Collection & Analysis

Data were collected at FHCRC beginning in January 2012 and ending in August 2012. Both field observations and interviews were conducted, as described in more detail below. In the end, I gathered around 95 hours of observations and conducted 17 interviews. These data were collected with approval from the FHCRC IRB.

3.3.1 Observations

Each of my observations was captured using field notes which detail the discussions taking place, as well as observations about the meeting. I attended weekly CC operations team meetings for both SN and BN, which take place in a conference room at FHCRC. During these meetings, the group addresses items on an agenda created by the project manager (PM), as well as any other issues on people's minds. Items raised may include such discussions as the allocation of shared personnel to various projects, meeting arrangements and uncooperative collaborators. Attendance at the meetings varies but generally includes the project manager and other coordination staff, the CC PIs, the informatics staff and the data management team. Meetings are fairly informal, last one hour, and are led by the PM and the PIs. Occasionally, other meetings were held that were open for me to attend, including the meeting between the SN CC and the NCI management team. This meeting took place every two weeks immediately preceding the weekly operations meeting. I attended a total of 50 operations meeting; 19 of these were the BN operations meetings, while the remaining 31 were the SN operations meetings, some of which included NCI representatives.

Additionally, I attended large group meetings for both SN and BN. The two SN meetings took place in mid-March 2012 at NCI headquarters in Bethesda, MD and in mid-September at FHCRC. The two-day meetings were attended by approximately 50 people, including the RC PIs and data managers of the 7 sites, the CC staff and NCI program officers. The BN meeting took place in Tempe, AZ, also in mid-March 2012. This two-and-a-half-day meeting was attended by the members of the Steering Committee (i.e., the PIs of all funded BN RCs) and focused on the scientific direction of the program. Fieldnotes of my observations were collected throughout both meetings.

3.3.2 Interviews

After several months of observations, interview instruments were designed to support interviews with CC staff and external collaborators in both BN and SN. I selected staff members who are most heavily involved with the projects, including some who are most active on BN, some who are most active on SN

and some who are equally active in both. CC staff were asked questions about their perceptions of what the CC's role and responsibilities are, as well as their own roles and responsibilities. Further, they were asked about challenges of coordinating multi-site research projects and how they handle those challenges. The complete interview instrument is attached as Appendix A.

I also identified external collaborators for both BN and SN who were invited to take part in interviews for this project. By interviewing external collaborators, I gathered more information on how others in the collaboration view the role and the work of the CC, as well as gaining a greater understanding of the overall collaborative environment in which the CC operates. For BN, I initially selected one PI from each of the four organ-site working groups, in addition to a representative from the funding agency. After several follow-up emails, only one of those initial contact proved fruitful, and I changed tactics and asked Adam, a BN/SN PI, for recommendations of PIs who might be willing to talk with me. Eventually, I interviewed three RC PIs, representing three of the four organ-site working groups, and one funding agency representative. These interviews were conducted via telephone, with the exception of David, who is an FHCRC employee. That interview was conducted in person in a conference room at FHCRC. For the SN, I had even less success with my initial contacts. I originally emailed three RC PIs, one from each of the organ-specific WGs, in addition to a funding agency representative. In the end, I interviewed one funding agency representative, two PIs from the colorectal working group, and one from the breast working group, either by phone or in person. These external collaborators were asked about the roles of the CC, funding agency and RCs, as well as challenges they have noticed in the CCEN's work.

After developing the interview instrument, I conducted two pilot interviews in order to "clarify the aims and frame of the study before interviewing its primary respondents" (Weiss, 1994). Based on the results of these pilot interviews, I further refined my instrument and moved forward with the rest of my interviews.

CC staff members were interviewed in person at FHCRC in a reserved conference room. These interviews lasted between 18 and 94 minutes, with an arithmetic mean of 52 minutes. In general, the interviews with CC staff were longer than the interviews with RC PIs. Before beginning the interviews, I ensured that the participant had signed the IRB-approved consent form and explained that their responses would not be shared with anyone although anonymized quotes may appear in published papers. Interviews were digitally recorded, transferred to a secure storage location, and transcribed by an outside firm.

3.3.3 Data Analysis

Both observation fieldnotes and interview transcripts were analyzed using a grounded theory approach. Grounded theory methods "consist of systematic, yet flexible guidelines for collecting and analyzing qualitative data to construct theories 'grounded' in the data themselves" (Charmaz, 2009). By hewing closely to the data, qualitative researchers are able to construct theoretical analyses that reflect the real lives of their participants. Additionally, grounded theorists engage in data analysis throughout the project, rather than just at the end, in order to allow for adjustments in the conduct of the research, as needed (Charmaz, 2009). This ensures that the researcher's questions and approach remain focused on the data.

The path from the data collected in this project to the analysis presented in this dissertation required many iterations to complete and included several rounds of data coding, memo writing, chapter drafts and conversations with my committee members. The research sites are complex and nested in ways that made it difficult to tease out connections; this complexity is compounded by the lack of literature on CCs. Based on the relevant literature, my research questions and my knowledge of the data, I created an initial set of codes to apply to my fieldnotes. After coding the first several fieldnotes, I realized that these initial codes were too broad and were unhelpfully covering large portions of the notes instead of pinpointing areas of interest (e.g., "organizational responsibilities, leadership or tasks aka articulation work"). I refined my codes to be more specific (e.g., "meeting leadership and arrangements") and applied them to my fieldnotes. On this initial pass, some additional themes emerged and were incorporated into my coding dictionary and applied to all fieldnotes. These themes included certain actions that occurred

frequently in the meeting, topics that warranted repeated discussion or interactions I noticed on more than one occasion. Once transcribed, interviews were closed-coded according to questions in a first pass through, then again for the themes identified in my fieldnotes. All codes are listed in Appendix B.

Once all data were coded, conceptual memos were written for most codes, representing the first attempt to take my analysis from descriptive to analytical. Some questions were combined for memos, as the codes were most useful that way. One example of this was the questions about project success, all of which were described in one memo. Throughout the process, my analysis was used to further refine my research questions, keeping my analyses grounded in the data. Memos were used to construct the initial drafts of the chapters that follow, combining themes from the different memos in ways that stayed true to the data while beginning to craft a view into the way the CCs under study worked. Such memos allowed me to see connections among the different codes and the analysis as seen here began to take shape. During this process, I returned to the data several times to review specific interviews or questions to support that analysis.

3.3.4 Citation Conventions

In order to protect the privacy of participants in this study, when I quote from or refer to the Request for Applications (RFA) for the BN and SN, I do not give the actual URL. Instead, I simply reference the BN RFA or SN RFA. When citing interviews, I use the participant's name and the line number of the interview (e.g., Nigel, 234). When citing field notes, I use the field note's file name (e.g., SN_FN_2-23-2012).

Chapter 4: Coordinating Center Enabled Networks

To date, there has been little research on either Coordinating Centers (CCs) or the form of virtual organization here called "Coordinating Center Enabled Networks" or CCENs. The CCEN is characterized as a research project being conducted within a networked structure that employs a CC as a central body tasked with facilitating the project's achievement of its scientific objectives. In this chapter, I define and describe what a CCEN is and does, based on a combination of my own primary data collection (i.e., field observations and interviews) and secondary materials such as the projects' own websites, published manuscripts, and the funding agency's Requests for Applications (RFAs, the published documents that detail the requirements of the project).

The findings presented below reflect my research on two specific CCENs located at the Fred Hutchinson Cancer Research Center (FHCRC), known here as the Biomarker Network (BN) and the Screening Network (SN). While we know that collaborative research has increased dramatically in recent years (Falk-Krzesinski et al., 2011; Wuchty et al., 2007), there are no comprehensive data available on how those projects are structured or how many utilize a CC. Although it is impossible to say how common this structure is or how frequently it is used, the structure's utilization by the National Cancer Institute (NCI) for the organization of both networks under study here, as well as the prominence of these two CCENs and their participants, provide reason to believe that other collaborative networks may also follow this structure. In fact, there are several CCs at FHCRC that are part of networks similar to the CCENs described here, including the Women's Health Initiative CC, which is funded separately from the research centers but via a contract mechanism instead of a grant (Presolicitation Notice for a Clinical Coordinating Center for the Women's Health Initiative 2010-2015 Extension Study)[2].

[2] The Women's Health Initiative network was not studied for this dissertation, and its structure will be the subject of future research.

4.1 Coordinating Center-Enabled Networks: A Definition

Coordinating Center Enabled Networks (CCENs) are research networks comprised of investigators from Research Centers (RCs), representatives of a Funding Agency (FA), and the staff and PIs of a Coordinating Center (CC), all of whom are focused on achieving the overarching scientific goals of a collaborative research project, goals that can only be accomplished within a network structure. Seminara et al. (2007) define networks in epidemiology as "groups of scientists from multiple institutions who cooperate in research efforts involving, but not limited to, the conduct, analysis, and synthesis of information from multiple population studies" (Seminara et al., 2007). Such networks can be built and/or funded in a variety of ways (Rolland et al., 2011); however, in a CCEN, the RCs and the CC are funded as individual components of the network via separate RFAs. The CC does not have an official pre-existing connection to any of the RCs.

As the name implies, the employment of a CC as a tool to facilitate the network's scientific objectives is a defining characteristic of a CCEN. Per the RFAs, the CC's primary responsibilities revolve around the operational and logistical coordination of the collaborative activities, and the data management and data analysis for collaborative projects. CC staff and PIs are expected to organize all network meetings, guide all the collaborative activities to ensure the production of high-quality data, create systems to manage the CCEN data and perform statistical analyses on those data (BN RFA; SN RFA). The CC also plays a role in generally helping the group of diverse sites work together as a network. However, as will be shown in the following chapters, that role is not always well defined or even agreed upon.

The RCs are the grantees charged with performing the scientific work they proposed in their grant applications. The precise nature of the work each RC does varies, from recruiting patients to extracting data from databases, but is all done in service of the CCEN's overarching scientific objectives as defined in the RFA. In addition to their scientific work, the RC PIs are expected to participate in the collaborative activities of the CCEN. These activities include attendance at meetings, contribution to discussions about the scientific direction of the CCEN, active involvement in relevant Working Groups making decisions

about scientific implementation, and participation in resource (e.g., biosample or data) sharing in compliance with CCEN policies (BN RFA; SN RFA). When asked in interviews about their role in the network, RC PIs from both projects noted that their job was to do their local science, acting as "data partners" as one RC PI put it (Beatrice, 105; David, 54; Evan, 27; Joan 43). Additionally, four of the six RC PIs told me their role was to participate in the network by serving on committees and working groups, as needed, to guide the scientific agenda of the network, sharing ideas and helping one another do the best science possible (David, 54; Evan, 29; Kevin, 107; Joan, 45).

The FA representatives in a CCEN, highly respected scientists in their own right, are there to represent the funding agency's interests in the project (Nigel, 389). The aim of this involvement is to ensure that the work proceeds as expected by the original proponents of the project, in hopes of achieving the project's scientific goals. FA representatives answer questions about the FA's expectations and policies, in addition to giving input on the scientific direction (Rebecca, 64). Like the RC PIs, the FA scientists are expected to attend all meetings and contribute to the discussions about how to achieve the project's scientific goals (SN RFA). They also participate in working groups, as appropriate. They work very closely with the CC to track the progress of the CCEN, generally through participation in frequent conference calls between NCI and the CC about the work being accomplished (Tamara, 290).

Both CCENs in this study are funded as cooperative agreements, a specific type of NIH funding structure in which the FA representatives have "significant scientific and administrative input" into the operations of the network (BN RFA). The FA representatives are not allowed to give direct instructions to the grantees, either the CC or RC PIs, on how to do their work, but are expected to give suggestions and guidance to ensure the project is meeting the FA expectations (Rebecca, 63).

The combination of these three elements of the CCEN definition – a *scientific objective* being achieved through a *network of scientists* including a *CC as a facilitator* – together set the CCEN form of research apart from other types of research structures. In the following section, I will illustrate the three

components of this definition by describing two specific CCENs: the Biomarker Network (BN) and the Screening Network (SN).

4.2 CCENs in Practice: Examining the Biomarker Network and Screening Network as CCEN Exemplars

This research project focused on two instantiations of the CCEN organizational form: the Biomarker Network (BN) and the Screening Network (SN). In this section, I will describe the three defining characteristics of a CCEN in more detail, using the BN and SN as exemplars of this organizational form. As per the definition, both the BN and SN were formed in the service of an overarching *scientific objective*, take the form of a *network of scientists* and include a *CC tasked with facilitating* the group's progress toward its goals.

4.2.1 The Biomarker Network

4.2.1.1 Scientific Objective of the Biomarker Network

The overarching scientific objective of the Biomarker Network (BN) is the discovery and validation of biomarkers for cancer diagnosis, prognosis and treatment response. Biomarkers are defined as "cellular, biochemical, and molecular (including genetic and epigenetic) characteristics by which normal and/or abnormal processes can be recognized and/or monitored. Biomarkers are measurable in biological materials, such as in tissues, cells, and/or bodily fluids" (BN RFA). In short, a biomarker is something measureable that can indicate the presence or status of cancerous cells, or a detectable cancer process, in a person's body. The driving hope behind biomarker work is the early detection of cancer or prognosis after a cancer develops. Prognosis is becoming increasingly important, as scientists come to understand that sometimes aggressive treatment of cancer is more damaging than the cancer itself (James, 215). For example, some cancers develop so slowly that the patient is likely to die from other causes before dying from the cancer. In that case, painful and costly treatments are a poor course of action. Currently, it is

difficult to tell this slow-developing type of cancer from a more aggressive, fast-moving variety – prostate cancer being the most obvious example.

The long-term aim of biomarker work is to validate the marker's efficacy in clinical trials such that it can be used in clinical practice (Thomas, 15; BN_FN_6-6-2012; BN_FN_7-11-2012). To achieve this, the BN engages in clinical validation studies designed to take a promising biomarker and validate its ability to detect cancer in a target population. In a clinical validation study, biosamples (e.g., blood, urine, tissue) and patient data are collected from enrolled patients and the proposed biomarker is tested in a larger population. Because the science of biomarkers is relatively new, it is still challenging to move biomarkers from initial discovery stage, where the biomarker shows promise in a small population, to FDA approval. In biomarker science, this is known as "Death Valley," the no-man's-land between discovery of a promising biomarker and its approval for clinical practice (Adam, 29).

One of the motivating factors behind the creation of the BN was not only the biomarker discovery and validation work itself, but also the development of biomarker science, in an effort to close the gap between discovery and validation and eliminate the Death Valley (Adam, 37). As such, one of the scientific objectives of the CC is the development of novel statistical methods for the validation of biomarkers (BN RFA). The BN CC has developed ways to approach biomarker work that give the biomarker under study the greatest possible chance of success by emphasizing the cleanest possible validation study, eliminating sources of bias and generating the highest-quality data possible (Pepe, Feng, Janes, Bossuyt, & Potter, 2008).

4.2.1.2 Network Structure of the Biomarker Network

The BN is comprised of a CC, representatives from the FA (in this case, NCI) and three types of Research Centers (RCs). The RCs are designated as either biomarker discovery labs, biomarker reference labs or clinical validation centers (BN RFA). Each of these types of RCs plays a different role in achieving the network's goal of discovery and validation of biomarkers. Each RC also is affiliated with a specific

organ-site working group, such as the "Prostate and Urologic Cancers" research group. All PIs, including those of the CC and the RCs, as well as the key FA representatives also serve on the BN Steering Committee, which is charged with setting the group's scientific agenda. The BN also has an Executive Committee, a subset of the Steering Committee, that makes final decisions on funded projects. Finally, the BN has a Steering Committee Chair, whose institution also serves as the BN's "home" institution for the grant period (Thomas, 79). The BN has a number of working groups (WGs) set up to deal with different aspects of the BN, such as publications policy or workshops (BN website).

The work of the BN lends itself well to a network structure for several reasons. First, the work of discovering and validating biomarkers, the main scientific objective of this project, requires the contribution of investigators from a range of disciplines, including clinicians, molecular biologists, epidemiologists and biostatisticians. By pooling their expertise, the group can tackle all the dimensions of such a difficult problem. For example, epidemiologists help select the proper target populations, while biostatisticians calculate how many patients need to be recruited and clinicians contribute their knowledge of the clinical outcomes for patients with the cancer of interest. Second, it takes time and money to recruit enough patients to validate a marker, especially if the cancer is rare. Recruitment across a large number of clinics can make the process go more quickly.

4.2.1.3 The Coordinating Center of the Biomarker Network

The BN CC coordinates 4-5 clinical validation studies at a given time, shepherding the study from proposal to data analysis through manuscript publication. According to Karen, a BN staff member, coordinating a clinical validation study includes "setting up protocols, ensuring that sites have IRB approvals for the protocols, capturing the data, ensuring the high quality of the data, receiving the analytic data from the labs, and then analyzing that data" (Karen, 54). A clinical validation study is a very complicated process that involves the steps described in this section. As each study is unique, the process presented is a high-level, idealized version, but includes the major tasks undertaken in a validation study.

The purpose of a clinical validation study is to validate a promising biomarker's viability as a predictor of disease in an appropriate and sizeable population. Once a biomarker is ready to validate, as determined by an RC's pre-validation studies, an RC PI or group of RC PIs writes a proposal for a validation study of the marker. This proposal is first vetted by the relevant BN organ-site group, then the Executive Committee and the NCI program staff. The CC is responsible for the coordinative tasks of this process, including assisting the RC PI with the writing, distributing the proposal to everyone who needs to review it and scheduling the EC vote. The statistical staff of the CC may be involved at this stage, advising the proposal PI on the preliminary study design, population sample size and other issues (Karen, 141). Because the CC has a high-level overview of the CCEN's work, they can also ensure that the proposed validation study fits within the BN's scientific objectives and follows established best practices for validation studies (David, 48).

Once the proposal has been approved, members of the CC continue to work closely with the RC PIs to turn the proposal into a protocol and then to a Manual of Operations that directs the work of recruitment, data entry and sample collection that is done at the clinical sites (Karen, 145). This protocol development requires a series of conversations between the CC and RC PIs about how best to design the study, as many details need to be worked out, including which types of patients and how many patients must be recruited, what types of samples to collect and at what volumes, what data points must be collected and how the IRB should be structured (Karen, 145).

After the protocol has been developed, the CC completes the study setup, which includes creating the data entry forms which will be used by the participating study sites in the Virtual Specimen Information Management System (VSIMS) (Karen, 145). This is a massive undertaking, as most studies require substantial amount of data collection on each patient. Using the protocol developed in the previous step, the CC staff turns the data collection protocol into computerized data entry forms that will be filled out for each participant. These forms not only contain entry blanks for all required data, they also contain programming logic to test participant eligibility and validate the data entered (Tina, 107).

Next, the study staff at each site must be trained in the use of VSIMS for data entry and in the proper collection of biospecimens. Again, consistency in both data and sample collection is essential to the success of the study; to ensure the greatest validity for the study, the protocol must be followed exactly (Tamara, 370). In this stage, too, all Institutional Review Board (IRB) approvals are secured, both for the work done at the CC and the work done at the participating sites. The CC must have IRB approval for receiving data and analyzing it, while the RCs must have their own IRB approvals covering their local work. The CC staff generally creates an IRB template, which then must be modified by each site to satisfy its own institution's IRB's rules. The CC project coordinator monitors all site IRB statuses, reminding them when they need to submit updated approvals and (gently) threatening to revoke access to VSIMS if updated approvals aren't submitted (Sarah, 129).

After this training, study site staff are given full access to VSIMS and can begin study recruitment and implementation (BN_FN_02-01-2012, 53). Once recruitment begins, the CC monitors enrollment numbers for the study via reports from VSIMS (BN_FN_02-01-2012, 53; BN_FN_02-22-2012, 80). Any questions that sites have about implementation or adherence to the protocol are answered by the CC staff (BN_FN_02-01-2012, 46). Study site staff consent each participant, collect any biospecimens required by the protocol and enter data into VSIMS. The samples are then sent for processing as per the approved protocol. At some point during the study implementation phase, a member of the CC staff does an official site visit, during which she assesses the site's adherence to the protocol for data and sample collection, seeking to correct any protocol breaches that may be detected (Sarah, 115).

After the samples have been collected, they are shipped either for storage or for analysis, based on the agreed-upon protocol. If the samples are being analyzed immediately, applicable assays, or tests, are run on the samples and the resulting data are sent to the CC for statistical analyses (Nigel, 41, 67, 213). The CC has a designated statistician for each organ group who has been involved in the study from the beginning and who continues to work with the PI(s) on the analysis of the data (Adam, 125; David, 48; Nigel, 79).

4.2.1.4 Biomarker Network Summary

The Biomarker Network, then, fits the definition of a CCEN presented in this dissertation. Its scientific objective is the discovery and validation of biomarkers for cancer diagnosis and prognosis. This complicated undertaking can only be accomplished within a network format, as it requires the expertise and resources of a wide variety of actors. Finally, the BN includes an independently funded CC charged with facilitating the work of the project, focusing specifically on the coordination of clinical validation studies. This coordination requires the CC to engage in activities that range across organizing conference calls, creating and enforcing compliance with protocols and statistical analyses.

4.2.2 The Screening Network

4.2.2.1 Scientific Objective of the Screening Network

The overarching scientific objective of the Screening Network (SN) is to develop a greater understanding of the current screening processes in the United States for three cancers – breast, colorectal and cervical – and then investigate how to optimize those screening processes for individuals through personalized recommendations (SN RFA). For example, a woman with a family history of aggressive breast cancer who displays risk factors of high Body Mass Index (BMI) and heavy alcohol use may need to be screened for breast cancer more frequently than a woman with no family history and no major risk factors. Yet exactly how much more frequently that screening should be done or how great the benefit of increased screening may be is currently not known. Furthermore, screening recommendations, procedures and outcomes can vary substantially across regions, and ethnic and socio-economic groups, as well as by insurance company or status.

The more tangible objective of the SN is the creation of a data repository that details the screening process for a large and varied population (SN RFA). To that end, the SN includes seven RCs, each of whom covers a defined population (e.g., HMO members or state residents) to whose medical records they have access. (Only the records of those patients who have consented to be included in the research may be

accessed for the SN study.) The SN seeks to document the way participants in those populations experience the screening process. The SN RFA defines a "screening process" thus:

> Screening process – A series of sequential steps individuals undergo to be screened: 1) being identified as someone at risk, 2) undergoing the screening test (detection), 3) undergoing the appropriate diagnostic procedures for those with positive results (diagnosis), and 4) being treated for pre-neoplastic [pre-cancerous] conditions and/or cancers that are found. The coordination of these steps is critical to its overall effectiveness so the actions that link the steps such as being offered the test (recruitment), being notified of test results, or being referred for further evaluation of a positive screening test are also fundamental to a complete and effective screening process. Performing the screening test is just one of the specific steps in the screening process. The test is used to identify people at increased risk for disease. The diagnostic evaluation includes additional testing (imaging, laboratory, and/or biopsy) and comprehensive medical assessment of the results to identify a cancer (SN RFA).

A "screening event" happens when a person experiences any portion of this screening process for breast, colorectal or cervical cancer. These screening events will be the focus of the SN data repository and could include the entire process or just one of the steps described in the definition, above. For example, if a woman receives a reminder to visit her physician for a mammogram, that is the beginning of a screening event. If she does nothing further, that screening event ends. If, however, she makes an appointment for a mammogram, attends the testing and gets a result, that is also considered part of the original screening event. This quickly becomes confusing for researchers, as they try to extrapolate these screening events from medical records. If the woman waited a year after the initial reminder, is that still part of the same screening event? If she makes an appointment and doesn't show up but later makes another appointment, is that one screening event or two? Agreeing on what exactly constitutes a screening event, especially when all the data are in existing medical records, is one of the massive challenges facing the SN. Further complicating this issue of defining a screening event is that they must be distinguished from diagnostic tests, those tests ordered in response to symptoms, such as a mammogram done after a woman finds a lump in her breast. Diagnostic tests should not be included in this data repository of screening events, but it is not always easy to tell if the test was screening or diagnostic. Given the differences in approaches to screening in the different RCs, as well as the data collected by different health care providers, it is no small task to agree upon the data points that should be collected.

The RFA for the SN RCs listed a number of different data types that each funded SNRC was required to send to this repository, but these were given at a very high level. They included:

1. size of the population potentially served by the network
2. the method of offering screening (outreach reminders, inreach reminders, other, neither)
3. the proportion of the target population that is actually screened
4. the proportion of patients with positive test results
5. the proportion of patients with positive test results who undergo diagnostic testing
6. the proportion of patients diagnosed with cancer
7. the proportion of patients that are treated for that cancer
8. individual indications for a test
9. the type(s) of test used
10. the results of the screening test and follow-up diagnostic testing
11. whether a given individual was seen for a diagnostic workup
12. whether the individual was diagnosed with cancer
13. whether treatment occurred, and, if so, of what type (surgical, radiation, chemo, and/or hormonal)
14. demographic characteristics such as age, race, and socioeconomic status
15. detailed cancer information from a geographically defined cancer registry (SN RFA)

These data are not particularly controversial, but they are collected in very different ways in each of the different SN RCs. In fact, some of them are not collected by most SN RCs but must be inferred from other data. The data point *indication* (data point 8, above) is a good example of this and is meant to describe why a given test was performed, either for screening or diagnosis. SN's focus is on the *screening* process, so it is looking for data on tests used for screening purposes, not diagnostic purposes. If a woman is given a mammogram because she has symptoms that lead her provider to suspect she has breast cancer, that is a diagnostic test, not screening, and that testing event should not be included in the SN database. However, given the different types of health plan data that exist across the spectrum of SN RCs, coming to an agreement on how to infer a screening event in the absence of an explicit data point is incredibly difficult (SN_SCMTG_09-12-2012). For example, a patient's medical record may not indicate symptoms in a way that the database picks up, leaving the data managers to create different algorithms for inclusion as a screening event. These different algorithms can lead to inclusion of very different screening events, making comparison across groups difficult and statistically challenging.

Because the data of each SN RC are so different, they must first be transformed, or harmonized, by each RC to common data elements (CDEs). While in the BN, the CDEs are used to detail which data to collect prospectively (in the future study), in the SN, CDEs are used to define variables to which existing data can be transformed retrospectively using already collected data points. One common example of a CDE might be Body Mass Index (BMI), which is generally calculated as the weight in kilograms divided by the square of the height in meters. If an RC has collected both height and weight, they can easily harmonize those data to the common definition of BMI through a simple calculation. Another example might be a variable for smoking pack-years, a measure of the total number of cigarettes smoked over time. To calculate pack-years requires both the average number of cigarettes smoked each day and the number of years the person has been a smoker. This can then be transformed into a standard format that allows statisticians to compare a person who smokes, for example, one pack (20 cigarettes) every day for one year to someone who smokes half a pack (10 cigarettes) every day for two years.

4.2.2.2 Network Structure of the Screening Network

The Screening Network SN is, like the BN, comprised of a CC, representatives from the FA (again, NCI) and three types of Research Centers (RCs). In the case of the SN, the RCs are divided strictly along organ lines, with three RCs focused on breast cancer, three on colorectal cancer and one on cervical cancer. Two PI representatives from the CC, the FA and each RC serve on the SN Steering Committee, which is responsible for setting the scientific direction of the project (Rebecca, 26).

The SN also has set up a number of working groups (WGs) focused on various aspects of the project that do not require the full attention of the entire group. For example, there is a Governance Documents WG charged with writing the SN's governance documents (i.e., a policies and procedures manual detailing roles and responsibilities, as well as policies such as those that determine publication procedures) (SN_SCMTG_03-08-2012, 126). Each of the organ groups has a WG in which they are deciding on the appropriate data points to collect for the data repository (SN_FN_03-20-2012).

The SN's goal of creating a large data repository can only be done within a network configuration due to the need for large amounts of data from many different data sources. To generate the type of data already available in medical records on such a large and varied population would take years and enormous amounts of money, so it makes sense to take advantage of existing organizational structures and data sources. Furthermore, medical records contain millions of data points, while the data repository requires only a small fraction of them. Thus, the involvement of investigators who are deeply knowledgeable about both local screening procedures and data sources is critical to the success of this project. Those investigators can help guide the discussions about which data are important and available to send to the repository.

4.2.2.3 The Coordinating Center of the Screening Network

NCI's main objective for the SN is the creation of a data repository detailing the screening process in the populations studied by each SN RC; thus, the CC's primary objective is the creation of this repository. What this means in practice is not only creating the database itself to house the data but working with the SN RCs to build consensus on precisely which common data elements to collect. Building this consensus has been challenging, drawing on both the scientific and organizational skills of the CC.

Nigel, a CC PI, described to me in an interview the CC's original plan for creating the CDEs, which involved identifying a large group of CDEs common across all three cancer types and a smaller group of organ-specific CDEs for each cancer (Nigel, 435). However, the CC quickly found this approach of focusing on CDEs across cancer types was not working. A screening event for breast cancer is quite different than a screening event for colorectal or cervical cancer, and few common CDEs could be identified. Thus, the decision was made to allow organ-specific WGs to create their own CDEs for the screening process, while the CC would develop a handful of cross-organ CDEs (demographic information, insurance status), data points collected on every patient. Each of these WGs contains at least one CC statistician and one CC staff member, while the one CDE specialist attends as many WG meetings as possible (SN_SCMTG_09-12-2012).

One of the major difficulties of creating CDEs across the seven vastly different healthcare delivery systems of the RCs is that each site collects and has access to very different data points. An HMO has access to very different data than a safety-net system (designed for patients without health insurance). As such, the CC had virtually no idea what data were available to be sent to the data repository (Nigel, 437). Their plan, which was also quickly scrapped, was to propose a set of "data concepts" to the SN RCs and ask them to tell the CC what data were available in each of those areas. For example, a data concept might be "Screening History," with proposed variables such as date of last screen, type of last screen, ever had a positive screen. The strategy was designed to avoid two main problems. First, the CC did not want to receive a data dump of the millions of data points available in a given health care system's electronic records, then be forced to wade through all those data points to find the screening-related data points needed for the project. On the other hand, the CC did not want to ask the SN RCs to come up with the CDEs out of thin air. They felt that by proposing the data concepts, they could use their years of experience with cancer-related CDEs on the BN to kickstart the CDE conversation while minimizing the work required of both the SN RCs and the CCs. Nigel explained it this way:

> "Since we didn't know what data may exist at centers that they would be extracting and sending to us, the process that we were planning on was we would say, "Here's a concept; here's an idea. What data do you have that is related to this? What sort of data elements do you have? And then once we have those, we could then start looking for commonalities and start creating, "Here is a specific question. Here's how we might get from various centers." I think they found that hard to understand. And one reason we went with that method rather than just say, you know, "Why don't you just send us your questionnaires and we'll do this," is, with the HMOs, their database is huge. I mean they've got millions of data items, and we can't have them do that. So, the other way of – you know, "For this particular concept, what type of data do you have?" We found that would be more efficient. I think they felt that it wasn't. And I also think that it turned out that, you know, we were trying to sort of do some things at a higher sort of all-center level as opposed to organ specific. And so, it was after that meeting that we went away from the CDE working group meeting to each of the organ working groups meeting to do CDEs instead" (Nigel, 437).

Once the initial CDEs have been decided upon, the CC will send a data request packet to the SN RCs with detailed instructions on how to prepare their data for submission to the CC. These instructions include data definitions and information on acceptable file formats. Once received, the CC statisticians will perform analyses for any papers (SN_FN_05-22-2012). The CC is also responsible for building the database that will house the SN data.

4.2.2.4 Screening Network Summary

The Screening Network, then, fits the definition of a CCEN presented in this dissertation. Its scientific objective is the creation of a data repository that describes the screening process for cervical, breast and colorectal cancers in the populations accessible to the seven RCs. The creation of this data repository requires the resources, both in terms of data and scientific expertise, of a wide variety of participants. The SN CC is an independently funded participant in the CCEN, whose responsibility is to facilitate the creation of the repository. They do so by engaging in activities ranging from leading conversations about data points for inclusion to programming the database itself.

4.3 Chapter Summary

In this chapter, I have presented a definition and discussion of the specific type of virtual organization here called a Coordinating Center Enabled Network, or CCEN. A CCEN is defined as a research network comprised of investigators from Research Centers (RCs), representatives of a Funding Agency (FA), and the staff and PIs of a Coordinating Center (CC), all of whom are focused on the overarching scientific goals of a collaborative research project. This definition was illustrated by the presentation of two instantiations of the CCEN format: the Biomarker Network (BN) and the Screening Network (SN).

The scientific objective of the BN is to discover and validate biomarkers for cancer diagnosis and prognosis, whereas the objective of the SN is to create a data repository describing the screening processes for cervical, breast and colorectal cancer in the populations studied by seven different RCs. In both cases, the objectives described require the application of the talents of all those involved, including the RCs, the CC and the FA representatives. The two CCENs have similar organizational structures, in that each has participation by the three types of entities, relies on a Steering Committee for decisions about the scientific direction of the project and utilizes committees and working groups to deal with many of the aspects of collaboration research such as developing governance documents or a publication policy.

As the name implies, one of the defining characteristics of the CCEN is its inclusion of a CC to facilitate the scientific objectives of the project. The CC is independently funded, with no connection to any of the RCs, and has its own scientific objectives to achieve. In other words, the CC is not simply an administrative structure charged with organizing the work of the CC, but, rather, is deeply involved in the scientific work of the CCEN. The BN CC coordinates the clinical validation studies of promising biomarkers, providing input on the protocol development, enforcing compliance with that protocol and performing the data analysis. The SN CC works with the RCs to decide which data points can and should be included in the data repository under construction, guiding the conversations by drawing upon their extensive experience with cancer-related data. Performing these tasks requires a CC to possess and utilize a vast array of organizational and scientific knowledge.

Although virtually all scientific research projects have a defined scientific objective, many projects are done using a network of scientists, and many projects include a CC, only CCENs are defined by the presence of all three of these characteristics. This unique combination of factors creates opportunities for both organizational challenges and leveraging the experience and expertise of the CC.

Chapter 5: A Typology of Work in a Coordinating Center Enabled Network

The work of a Coordinating Center Enabled Network (CCEN) is varied and complex, ranging from the organization of conference calls and meetings to recruiting patients for clinical studies to running complex molecular experiments. In order to understand how the Coordinating Center (CC) facilitates the work of the CCEN, the overall goal of this research study, we must first understand precisely what that work entails. In this chapter, I present a typology of work practices of the CCEN that helps us to make sense of this complex organization. This typology grew out of my study data and represents a way to describe and categorize the work performed by CCENs, providing a richer definition of what a CCEN is and does. This typology description will be followed by an illustration of each type of work practice using examples from interviews and observations of participants in both the Biomarker Network (BN) and Screening Network (SN). Finally, the typology will be further expanded to include definitions of "local" work, work performed by one entity alone or two entities working together, and "network-level" work, those types of work performed by all three entities working together at once. As will be shown, this network-level work is the heart of what the network does, as it is the work that requires a network structure, a defining characteristic of a CCEN.

5.1 Categories of CCEN Work

In their quest to achieve the CCEN's scientific goals, CCEN participants engage in five types of work practices. Such work practices were observed in both projects, the Biomarker Network (BN) and Screening Network (SN), in field observations and interviews. All CCEN participants – Research Centers (RCs), Coordinating Center (CC), and the Funding Agency (FA) – may engage in work in each of these types of work practices at some time during the project, either independently or with others. The one exception to this is the lack of local scientific work by the FA representatives. Table 5.1 gives examples of the work that each participant does in each category and are described in detail below.

In this section, I offer a definition of the five types of CCEN work practices, focusing on the breadth of activities in each category and across participants. The following section, Section 5.2, will illustrate each category in greater detail with examples of work being done in each type of work practice.

Type of Work Practice	Example CC Tasks	Example RC PI Tasks	Example FA Tasks
Structural (SWP)	• Negotiating revisions to scientific objectives in case of unexpected funding changes • Suggesting changes to RFAs for new funding cycles	• Submitting grant proposal • Agreeing to accept funds and follow RFA • Suggesting changes to RFAs for new funding cycles	• Writing the RFA • Making funding changes • Selecting grantees • Revising scientific objectives as needed
Collaboration (CWP)	• Deciding how to interpret RFA in practice (with FA) • Negotiating questions of roles and responsibilities • Prioritization of projects in view of limited resources	• Participating in committees and working groups • Fulfilling role in CCEN • Attending meetings	• Deciding how to interpret RFA in practice (with CC) • Evaluating grantee progress, reviewing progress reports • Ensuring CCEN moving toward scientific objectives
Operational (OWP)	• Organizing meetings and conference calls • Programming data entry system • Coordinating protocol development • Management of IRB approvals	• Sending questions to CC • IRB documentation submission to CC • Registering for meetings	• Creating agendas for meetings • Scheduling site visits
Data (DWP)	• Distilling scientific questions to data points for collection • Statistical analyses • Protocol development and study design • Database design	• Data extraction to produce CCEN-requested datasets • Running assays or experiments according to agreed-upon CCEN protocols • Research to answer information requests from CC about data • Reviewing analysis results • Writing protocol with CC staff • Data entry	• Representing interest of FA in protocol development • Reviewing analysis results
Local Scientific Work Practices (LSWP)	• Development of novel statistical methods	• Running assays or experiments to discover new biomarkers	

Table 5.1: Examples of Five Types of Network-Level Work Practices in a CCEN, by Participant

5.1.1 Structural Work Practices

Structural work practices are those activities that shape the rules of the project and dictate the organizational structure of the CCEN, once funded and instantiated. Most of the structural work is done by the FA in the development of the Request for Applications (RFA, the project's definition by the FA), which specifies the scientific objectives of the project, the governance structure (i.e., required committees, how decisions will be made, and how the scientific direction will be set), and what the overall responsibilities of the grantees will be. While this work is predominantly in the realm of the funder, other CCEN members may need to participate in structural work if changes take place during the funding cycle, as discussed below.

Much of this type of work takes place before the RFA is even published. This work includes defining the objectives and scope of work of the project, the total funding available to each project component, the organizational structure of the project, the criteria by which applications will be judged and who at the funding agency will be responsible for the project. All of this work takes place at the funding agency and must follow strict federal guidelines. After the RFAs are published, potential grantees, including the RCs and CC who will eventually be funded, respond with proposals applying for the available funding. These proposals are reviewed by a peer review committee and evaluated based upon standard review criteria defined by NIH and any additional criteria defined in the RFA. Successful applicants are notified of their application's funding and the funds are disbursed. This is the standard process for funding at NIH. While this RFA creation and scientific review process takes place outside the scope of this study, it has a major impact on the composition of the CCEN. First, the RFA is written by the FA representatives without any real idea of who the eventual grantees will be. This means the project cannot be explicitly structured to take advantage of the strengths of specific grantees and the CCEN cannot be explicitly formed in a way that ensures complementary skills and expertise. While the FA representatives may have specific scientists or groups in mind while writing the RFA, there is no guarantee either that that scientist or group will apply for the funding or that s/he will eventually be chosen by the review process. In the NIH peer-

review system, the peer-review committee, not the NIH program officers, has the final word on how proposals get scored, though the program officers are involved in funding decisions. Thus, the RFA must be written such that the review criteria will lead to the best possible network of scientists, not an easy task. Second, the FA representatives writing the RFA have great latitude in designating the organizational structure that the CCEN will eventually take, without any real empirical evidence for or against potential structures (personal correspondence, Dr. Robert Croyle, Director, Division of Cancer Control and Population Sciences, National Institutes of Health, January 11, 2013). As such, the organizational structure dictated in the RFA, combined with the specific mix of grantees eventually chosen, have the potential to create a situation in which a group of people have difficulty working together.

While the majority of the structural tasks are completed before the collaboration is even formally inaugurated, sometimes changes made by the funding agency in mid-cycle require the CCEN participants to engage in structural work, such as when funding is changed (e.g., funding cuts) or scientific objectives must be modified due to new advances in knowledge. In this case, negotiations between the grantee and the funding agency may need to take place in order to determine how to adjust the grantee's deliverables. Furthermore, if a project is successfully refunded in subsequent funding cycles, the existing grantees may be asked for input on how the project should be structured in the next grant period.

Because most of this work is performed by the FA and does not include interaction with other CCEN entities, the majority of the structural work of the CCEN is outside the scope of this study. Some instances of structural work that involved interaction between the FA and RCs or CC in response to budget cuts or new funding cycles were observed or mentioned in interviews and will be discussed in greater detail, below.

5.1.2 Collaboration Work Practices (CWPs)

The collaboration work practices are the work of negotiating and deciding how to work together as a network, as well as the work of participating in those negotiations and decisions, all within the

organizational structure set up by the structural work practices discussed above. A CCEN brings together researchers with differing experiences, skillsets and motivations and must work together to create a path toward achieving the project's scientific objectives. The CWPs include allocating resources when there are competing priorities, participating in committees that set the scientific direction or make decisions about how projects will get done, as well as communicating project priorities and attending meetings and conference calls. As members of the network, everyone involved in the CCEN has responsibility for some type of CWPs. These CWPs can take up a substantial amount of collaborators' time, especially if CCEN members have differing ideas of how the network should proceed toward its scientific goals.

What is not included in this category is the work of deciding specific scientific or data questions, such as creating study protocols or developing a list of requested data points. Also not included are the work of scheduling committee meetings or organizing the in-person meetings. These activities are categorized as data work practices and operational work practices, respectively. This is a fine line to draw, but is important to make this distinction because the CWPs require different skills, different participants and different time commitments than those needed for the operational work and data work practices. By separating them out, we are able to get a fuller picture of precisely how the CCEN functions.

5.1.3 Operational Work Practices (OWPs)

The operational work practices are the administrative and technological tasks done in support of the other types of work. Their aim is to help the group's diverse and varying tasks function together as a whole, as in when the CC organizes conference calls so the group can get together and discuss how to collect data for a study or building a database that will receive appropriate data from the RCs and be used by the CC's statisticians in their statistical analyses. These activities are primarily logistical or technical in nature and, in general, require little scientific knowledge to complete them. This is not to say that those engaged in these practices have or use no scientific knowledge while performing OWPs, but, rather, that scientific knowledge is not generally required to complete these activities. Additional OWPs include such logistical tasks like organizing meetings, taking minutes, emailing collaborators for information and managing

project tasks. Also included in this category are technical tasks such as building the project databases, and building and maintaining the project website. However, the design of the project database is considered a data work practice, as it requires a deep understanding of the project's data and the application of extensive relevant scientific knowledge. These operational work practices are also sometimes referred to as research administration in the world of scientific research. This work is primarily initiated by the CC; however, the CCEN members also must participate in this kind of work. For example, RCs must respond to CC requests for IRB documentation or requests for scheduling information.

5.1.4 Data Work Practices (DWPs)

Data Work Practices are those activities the focus of which is the production (i.e., the RCs generating data via lab work or extracting data from local databases) and consumption (i.e., the receipt of data for statistical analyses) of high-quality data. The DWPs require a degree of scientific knowledge and expertise. This data work begins with the group's efforts to agree upon protocols and common sets of data to collect and extends through the receipt of the data and performance of statistical analyses. Included here are such tasks as developing the project protocols and study design for a clinical validation study, statistical analysis and designing scientific databases. Also included here are any activities done by the RCs to generate data in compliance with the agreed upon protocols, such as recruiting the correct patients or extracting agreed-upon datasets from local databases. Not included here are the back-and-forth communications involved in managing the development of protocol and data set agreements, such as requests for comments or reminders to review the protocol, which are OWPs.

5.1.5 Local Scientific Work Practices (LSWPs)

The local scientific work of the CCEN encompasses those activities done at the local organizations (i.e., at the individual RCs or CC) the aim of which is to achieve the individual scientific objectives that a participant proposed in his or her grant application. The LSWPs are scientific activities that happen in the CCEN that do not require interaction with other CCEN entities and are not guided by the collaborative

protocols developed by the CCEN members. In other words, these activities are done independently by an RC or the CC; no LSWPs were observed among, or attributed to, the FA.

As the name implies, these work practices utilize participants' extensive scientific expertise. Examples of such activities are the CC's development of novel statistical methods for biomarker science as well as assays RCs might run to discover promising new biomarkers. Those LSWPs performed at the RCs are outside the scope of this study, as that work was not observed in my fieldwork or discussed in interviews. However, some instances of the LSWPs undertaken by the CC will be discussed in greater detail, below.

5.2 CCEN Work Practices Typology in Illustration

To reiterate, this research project focused on two instantiations of the CCEN organizational form: the Biomarker Network (BN) and the Screening Network (SN). In this section, I will illustrate the typology of CCEN work practices using brief examples from both CCENs. This section does not cover the full extent of work being done in each category but, rather, uses examples of each to illustrate the contents and boundaries of the categories.

5.2.1 Work Practices of the Biomarker Network

The overarching scientific objective of the BN is the discovery and validation of biomarkers for use in cancer diagnosis and prognosis. To accomplish this objective, the BN participants engage in all five types of work practices defined above. In this section, I illustrate each of these work practices with examples from the work of the BN.

5.2.1.1 Structural Work Practices (SWPs) of the BN

Again, structural work practices are those activities that shape the rules of the project and dictate the organizational structure of the CCEN once funded and instantiated. While the majority of SWPs take place before the project is even funded, circumstances sometimes dictate that CCEN participants must engage in negotiations in response to changes in funding such as budget cuts or a new funding period. This is the case for the BN, which is currently in the middle of its third funding cycle (i.e., the BN has

been funded for three consecutive five-year grant periods). At the end of each cycle, the NCI program staff asked the CCEN members for input on any changes they would make to the structure of the BN and its scientific objectives, thus engaging them in SWPs whose outcome influences the way the project would be structured in the next cycle.

One of the changes suggested by the CC was to add what they call "Team Projects" to the requirements for each organ group, thus proposing a change to how the BN is structured. Some in the CC felt that not enough collaboration was happening within the biomarker discovery labs, which was holding back the entire BN. Adam, a BN PI told me, "'[t]eam projects' is a concept we proposed after the [first] two cycles … because we saw the individual biomarker discovery lab, most of them just do not have ability or capacity to move the biomarker to validation. So we thought maybe they needed some help. And so if we have team projects, as a team they can pool resources together, pool expertise together, can recruit the sample quicker and they can identify [some] of the most important questions" (Adam, 275). These Team Projects are still getting off the ground, but already they have led to greater collaboration among the discovery labs, which the CC hopes will result in more biomarkers to validate (Adam, 345). By adding additional responsibilities to the project requirements, the CC and the FA engaged in structural work.

Two other examples of the SWPs in which the BN FA representatives engage are the development of an annual evaluation system for grantees and the funding mechanism by which grantees receive funds. The BN has enacted an assertive evaluation system built into the yearly NCI site visits, which are designed to assess how well the grantee has done over the year according to agreed upon metrics. Each year, grantees must fill out an evaluation form on which they list their scientific and collaborative activities over the year, including how many meetings they attended, the number of biosamples they shared, how many team projects they joined (Thomas, 311; James, 161). The result of this evaluation system is to enforce expectations of collaboration and to make the rules of engagement explicit for CCEN members.

In addition, the funding structure of the BN developed by the FA representatives requires collaboration, as a portion of each RC's funds can only be spent on collaborative projects. The BN also sets aside funds from the overall BN funding pool to support larger collaborative projects. When discussing how he and his colleagues structured the BN, Thomas noted:

> Yeah, so from day one I wanted to emphasize and in fact in some cases I was very blunt to tell investigators that this is not an R01 [individual investigator grant]. Here is the need to work together. If you feel that you cannot work with others and share your findings with others towards the goal of validating biomarkers, then it's not your place to be here. I made very, very clear at the very outset; therefore, I emphasized the word collaboration, collaboration, collaboration. And then I went and said that we have built the funding mechanisms within [BN] such that it not only supports collaboration but also rewards collaboration as well. So the [BN] funding structure has very unique components. For example, each [BN] investigator has their own grant but part of their grant, almost about 30% of their grant, is restricted and that restriction is lifted only when they propose a collaborative study with other members of [BN] or for that matter for outside, not with [BN]. So 30% of their funding is strictly to do that, so that's what we call support mechanism. Then there is a reward system. Their reward system is that [BN] has set aside funds that are decided with the [Steering Committee Chair]... So they have set aside funds amounting to almost about five million dollars, but it is not for [Steering Committee Chair's] use. It is for the use for rewarding large validation and collaborative studies" (Thomas, 71).

These structural work practices in the BN set the expectation for not only the data and local scientific work but also the collaborative work that must take place in the project. By creating a structure in which collaboration is encouraged and supported, even required, by both the evaluation and funding mechanisms, the BN FA representatives have given shape to their idea of what the BN should look like.

5.2.1.2 Collaboration Work Practices (CWPs) of the BN

The collaboration work practices are the work of negotiating and deciding how to work together as a network, as well as the work of participating in those negotiations and decisions, all within the organizational structure set up by the structural work practices. One of those CWPs performed by the BN is the creation and enforcement of a culture of collaboration. Note that the structural work done by the FA of putting into place rules that require and encourage collaboration is different than the work of the BN members to build a culture of collaboration and sharing. The structural work is enforced by the FA, the only entity with the authority to "punish" (i.e., take away funding from) those that don't comply with the

rules, while the collaborative work to build a culture of collaboration and sharing is enforced by social pressure from the CCEN participants.

This cultural pressure to collaborate is apparent as the BN begins each new funding cycle, when some RCs leave the project and new ones join. The new RC PIs often come in with ideas of how they will work within the BN that reflect the dominant culture of science, which is competition. It is the job of everyone – the CC, NCI and returning RC PIs – to help the new RC PIs understand how science works in the BN. James, a BN RC PI, noted that the BN's years of experience working together has eliminated many difficulties of collaborating, including communication difficulties, though each new group of collaborators that joins the BN still must be acculturated.

> Betsy: So, one of the classic complaints about the difficulties of trans-disciplinary research is the difficulty in people like clinicians and basic scientists talking to one another, using different language. Have you experienced that in [BN]?
>
> James: Nope. In fact, the really extraordinary thing about the [BN] is because we've been living together for the last 12, 13 years, we talk the same language. Now, as new people, with each round of grant applications there are some people who drop out and some people who drop in, funded individuals, there's a big of a learning curve for them. So we help them understand, but there's a bulk of institutions that will carry over and be re-funded. That's the group, the organization that provides that, if you will, legacy information on to the next group.

Another BN investigator, Kevin, also agreed, noting that the organ group's jargon "is pretty much standardized" (Kevin, 255), so even the communicating across epidemiologists, biostatisticians, clinicians and basic scientists, which must happen in the BN, was not a challenge.

In addition to the sheer amount of time spent working together, James attributed the BN's ease of collaboration to the NCI program staff's consistent message of the requirement to collaborate in the BN, as discussed in the previous chapter. He noted that only those who chose to collaborate as a way of doing science would be successful in this group.

> Beyond that, you'd like to think that perhaps there are challenges in terms of distance collaboration, so the clinical centers are working with multiple centers across the United States, there really aren't, and, in fact, I'm going to say good things about the group on a regular basis but these days… Team science was not something that people recognized years ago. It was all investigator-initiated grants, and it was you against the rest of the world. What you were doing is you were submitting your grants and doing your research. That doesn't happen in the [BN]. In fact, in the [BN], everybody has filed a non-disclosure agreement.

> When we go into our meetings, everybody lifts up or opens up their books and shows everybody everything, because the ethos of the group is that if one member of the group benefits, everybody benefits. We are judged not as much by our individual institution's accomplishments. We are judged more by the group's accomplishments. So, because that ethos was instilled at the very beginning, what has happened was that program staff has really selected the membership of the [BN] based on the collaboration. The more collaborative you were, the more likely that you would be funded. So, what has happened over the past two funding cycles is that this is now a group of individuals who it's like a family where you value your partners almost more than yourselves (James, 125).

In this series of quotes from BN RC PIs, we see again how interwoven the work practices are. The structure put into place by Thomas and his FA colleagues has resulted in a culture of collaboration, a culture that is maintained and developed by the RC PIs themselves. As James noted in the quote above, this collaboration work also has an impact on the data and scientific work of the BN, in that "everybody lifts up or opens up their books and shows everybody everything" (James, 125). This increased sharing changes how the science will proceed, as it changes the data and information that are available.

5.2.1.3 Operational Work Practices (OWPs) of the BN

The operational work practices are the administrative and technological tasks done in support of the other types of work. Their aim is to help the group's diverse and varying tasks function together as a whole. In the BN's clinical validation studies, a significant amount of operational work is required to support the completion of the study, which is composed of many tasks that must come together in order for the project to work. To begin with, creating the protocol for a clinical validation study requires significant communication among the RC PIs, FA and CC staff working on the study; the work of coordinating this communication, distributing different versions to various reviewers, scheduling the vote by the relevant committees and posting the protocol on the website for others to see, even formatting and spell-checking the final document, is all operational work (Sarah, 105; BN_FN_2-29-2012). While the CC contributes substantial scientific work and expertise to the protocol-development process, as discussed in the next section on data work below, it is also responsible for the majority of the operational work required to produce the study protocol.

Another type of OWP is the building of technical systems for data entry and specimen tracking, as the systems are built in support of the data work that will use the systems. In the early years of the BN, the CC spent a substantial amount of time building the systems that today make them efficient enough to be able to coordinate 4-5 validation studies at a time with minimal staff (Karen, 93). (These systems are also being used by the SN.) The Virtual Specimen Information Management System (VSIMS) was developed by the BN CC to handle the data entry and tracking that takes place in a validation study. Each of the individual biomarker-validation studies receives its own instance of the VSIMS, with its data entry and database customized to the needs identified and described in the protocol development detailed in the section above (Sarah, 315). Several programmers are on staff in the CC do this work.

Another OWP is the management of Institutional Review Board (IRB) approvals. While the CC must receive IRB approval for their own work at FHCRC, each site in a clinical validation trial must also have local approval of their own work. The CC uses the information in the approved study protocol to draft a consent-form template for the study, which is then distributed to each study site, which must modify it as necessary to suit its own IRB requirements. Once the site has local IRB approval, it must send verification of this approval to the CC. Access to the data-entry system will not be given until this approval has been received by the CC. In fact, data-entry privileges will be revoked if confirmation of yearly re-approval is not submitted in a timely fashion (Sarah, 459).

Here, again, we see the interdependent nature of the different work practices. The OWPs have an important impact on the work they support. As mentioned above, a clinical validation study involves the coordination of many types of work. If those tasks aren't coordinated such that they come together in a way that allows the project to function as a whole, the project has little chance of success. The technical systems, too, must be not only functional and accessible, but must support the data work that depends upon them. If a database is built that does not match the datasets being produced by the RCs, the data cannot be received by the CC and the statisticians cannot perform their statistical analyses.

5.2.1.4 Data Work Practices (DWPs) of the BN

Data Work Practices are those activities the focus of which is the production and consumption of high-quality data. For the BN, the DWPs begin with the project proposal by an RC PI to the relevant organ working group. Once that proposal has been approved by the working group, it must also be approved by the Executive Committee. During this process, the RC PIs work closely with the CC statisticians and staff to ensure the protocol follows the BN's conventions and takes advantage of existing resources such as common data elements, data points with precise definitions. The CC also helps the RC PIs define their eligibility criteria more explicitly and prepare the forms for data entry. Note that the work of deciding which data points to collect is defined as data work, while the actual programming of the data entry forms is considered operational work.

In interviews with both internal and external BN participants, this data work was noted as critical to the success of the project. When asked about the role of the CC, Karen, a CC staff member, noted first that it was to ensure high quality data for the validation studies (Karen, 38). Several members of the CC also stressed that high-quality data is the top priority for all the data work that they do. Edith, a CC staff member, noted that "from my perspective, the role of the Coordinating Center is to make sure that the data sets are being collected in ways that produce the best possible quality" (Edith, 61). This focus on quality stems not just from a desire to do their jobs well, but from an understanding that only through high-quality data can they truly achieve the group's scientific data. If the data are at all suspect, the BN loses the ability to make claims about the quality of a biomarker, as described by Tamara, a BN staff member:

> "I think it's a process of educating folks that if you're trying to figure out a usable biomarker, it's imperative that your samples are uniform and are of the highest quality. So it benefits you to follow these protocols and I think it's educating the people to think in a bigger picture. This is going to be better science if we all do it in a standardized way that is of quality. And then ultimately we will have better outcomes because you won't have some crazy data set. … And so then we'll know, gosh, that biomarker failed and I'm pretty comfortable that it failed because my samples were of quality or wow, that biomarker had awesome results and I'm really confident with my data because my samples were really good quality" (Tamara, 370).

Tamara notes that it is the CC's responsibility to make sure that validation study sites understand why compliance with the protocol is so crucial, underscoring the importance of the communications work being done by the CC.

In addition to the protocol development described above, another example of data work in the BN is the work being done by the clinical validation study sites to obtain the data for the CC's analyses. This includes any work done in accordance with the agreed-upon protocol, such as recruiting patients who meet the eligibility criteria, collecting data on the patients via the data entry forms and collecting and processing biosamples as directed by the protocol.

5.2.1.5 Local Scientific Work Practices (LSWPs) of the BN

The local scientific work of the CCEN encompasses those activities that are done at the local organizations (i.e., at the individual RCs or CC) the aim of which is to achieve the scientific objectives of the CCEN but that do not require interaction with other CCEN entities and are not guided by collaborative protocols. While the RCs each have their own individual scientific objectives to achieve, the work that does not involve the CC is outside the scope of this study. However, the CC itself has scientific objectives required by the RFA; namely, the CC is charged with developing novel statistical methods for biomarker science. As discussed in Section 4.2.1.1, when the BN was first formed, the science of biomarkers was still very young and experiencing difficulties in achieving its goal of discovery and validation of biomarkers. Over the course of the first two funding cycles of the BN, the CC has developed study-design procedures and statistical methods that increase the chances a promising biomarker will be validated in a clinical study. This work takes place exclusively at the CC.

The development of novel statistical methods for dealing with the complications of accurately validating biomarkers is one of the main charges of the BN CC. When the BN began, little was understood about how to ensure that biomarker validation studies were reliable. As James, a BN RC PI noted, "the science of biomarkers is complicated. … One would like to think that – say you have a blood test or a urine test

that you think finds a cancer early. One would like to think that there is a very simple design of a study that will confirm that. Actually, it is extraordinarily complex" (James, 35). The CC has made a major contribution to the field of biomarker science by creating study design and clinical validation criteria for biomarker discovery and development (Pepe et al., 2008).

Thomas, an NCI representative, described the importance of the work of the CC, noting the lasting impact the CC's statistical work had had not only on the BN but on the field of biomarker science overall.

> "For example [CC statistician] is so well-known in the area of cleaning and early detection for her statistical research. [Adam], again, very well-known in the field, so they come up with some creative ideas and one of the creative ideas that you can think about was their publication on five-phase criteria for biomarker discovery and development. That was followed by what we call probe design and probe design actually expands on the five-phase criteria in a way that it gets what is required for each phase to move from phase one to phase two, to phase three [of biomarker discovery and development]. What should drive the study design? So they talk about clinical endpoints, then what sort of specimens are needed for that, to meet their clinical goal. So the probe design expands on five-phase criteria to elaborate on the requirements of the biomarker validation, depending on the organ sites you deal with. So I think those are the unique contributions that CC has made to the research within their coordinating center and this has been partially because we have leaders in the field of statistical design at Fred Hutch. So those were something that they did for the larger community but they also conducted studies within their center and that are very useful for everything we do within BN, and they do it despite the fact that they have very small amount of funding available to do those things" (Thomas, 109).

Aside from the obvious scientific benefits of developing stronger and more reliable methods for ensuring that the BN validation studies delivered trustworthy results, the CC's work on statistical methods and study design have had the added benefit of a boost to the entire field of biomarker science. This not only will help in the discovery of more biomarkers, it also will potentially influence the BN's chances of being re-funded as it is evidence of the group's efficacy.

Thomas further described the substantial impact the CC data work has had on the BN, noting he wishes they had more funding for the CC to expand their services from network-level work to the local work of the RCs.

> "Well so they are honestly – I don't want to brag about it but [CC staff] are so well-appreciated by members of the [BN] that some of the members started asking whether [the CC] can advise individual members on their statistical study design. That was not possible because of the funding restrictions and also the funding limitation. Without restriction and limitation we could not fund enough to [CC]. But [CC] agreed that on a case by case basis they will help individual investigators if the study is likely to lead to a large validation study" (Thomas, 153).

The CC has developed such a stellar reputation as biostatisticians who elevate the quality of a study that Thomas wishes they could help all members of the BN with their statistical work, especially in the realm of designing stronger studies. Well designed studies result in stronger, more valid, conclusions; even when the results of the study are null, scientists have produced new knowledge that can be trusted. Unfortunately, the resources of the CC are limited so that they are able to coordinate only 4-5 validation studies at a given time.

In general, the local scientific work practices are performed independently from the rest of the CCEN, though they still are conducted within the structure, especially the scientific objectives, set by the FA representatives when developing the project. While the CC's development of novel statistical methods for biomarker science will eventually be implemented by the CCEN in its study design, the work of actually developing those methods is done solely by the CC.

5.2.2 Work Practices of the Screening Network

The scientific objective of the Screening Network is to create a data repository that describes the screening processes experienced by the populations under study. To build such a repository requires engagement by members of the SN in the five types of work practices of a CCEN. In this section, I present examples of the work that takes place in the SN in each of the five categories.

5.2.2.1 Structural Work Practices (SWPs) of the SN

To reiterate, structural work practices are those activities that shape the rules of the project and dictate the structure of the CCEN once funded and instantiated or those in which the CCEN engages in response to changes in funding such as cuts or a new funding period. The SWPs of the SN have been unusually apparent in the SN because of several budget cuts experienced by both the CC and the RCs. The first problem came about when the CC first received notice that they had been selected as the CC for the SN. The original RFA noted that the total available funds for the CC were $2 million for the first fiscal year. The CC was contacted by the FA representatives before funding began and told that that amount had been

a mistake and the actual allocation was just $1.5 million, a 25% cut. However, the scope of work remained the same, leaving the CC scrambling to adjust its staffing and strategy (Adam, informal meeting, 1/9/12).

The RCs, too, have received cuts to their budgets. Near the end of the project's year one, RCs were informed they would be receiving a 6% cut to their budgets, on top of an initial 17% cut at the beginning of the project, and their grant terms would be shortened by three months (Rebecca, 64; Beatrice, 149; Tamara, 115). Joan, a SN PI talked about the impact of these cuts on her RC's scope of work, noting they turned to the FA representatives for advice on how to deal with the cuts.

> So, when this first got started, you know, we had discussions with [FA representative] about what we would not do with the major cuts involved. I mean, it will definitely affect our ability to support novel things beyond the scope of the letter of the law here. I'm hoping we won't have trouble doing the letter of the law in terms of what we've said. But it does [have an effect], particularly the thing where they've shortened the grant by three months is tricky (Joan, 209).

Beatrice also mentioned talking with FA representatives about how to absorb the cuts to their award:

> Yeah, so you know we actually talked to our project officer about that. He just said, you know, the NCI really cares about the quality of the data for the repository but with budget cuts, the quality of the data for our projects, as well as the quality of the data for the repository sort of starts – that's where you start seeing it. You can't do the QC or you can't chase down every loose end if you don't have the resources to do it (Beatrice, 69).

When a grantee accepts the funds from the FA, they are obligated to complete the work they proposed. However, when large funding cuts are made by the FA, negotiations need to take place about which parts of that work will not get done, basically seeking to realign the structure of the project with the work expected, based on the funding received. The second set of cuts to the RCs was especially disruptive, leading the FA representatives to spend a significant amount of time working with the RCs to help them understand the impact of the cuts on their work (Rebecca, 64; SN_SCMTG_9-13-2012). The time spent having these conversations not only took time away from the other work the FA representatives and RC PIs needed to do but also impacted the data work and local scientific work the RCs were able to do.

5.2.2.2 Collaboration Work Practices (CWPs) of the SN

The collaboration work practices are the work of negotiating and deciding how to work together as a network, as well as the work of participating in those negotiations and decisions, all within the organizational structure set up by the structural work practices. As with the BN, members of the SN engage in CWPs by participating in the working groups and committees of the CCEN. The SN has a Steering Committee charged with setting the group's direction, as well as a subgroup of the Steering Committee responsible for writing the governance documents (SN_SCMTG_03-8-2012). There are CDE-specific working groups, as well as one focused on a specific electronic medical record database that many of the RCs use. The number of PIs in the SN is not very large, creating a burden for RC PIs, who must participate in multiple groups to ensure their interests are represented. For example, any discussion of CDEs must take into account the perspective of all three cancer types. However, there is only one cervical site, leaving that PI feeling spread thin, given that she must be a part of the SC, the governance documents committee, the organ-specific working group and the single CDE working group. I observed this frustration at the first in-person meeting, when this PI was required to attend yet another working group meeting (SN_SCMTG_3-8-13), this time on the governance documents being written. While she showed little interest in most of the discussions, there were points where it was clear she felt her interests were different than those of the other RCs, especially in the area of decision making. One proposed decision-making rule was that each RC had one vote about which data would be required for submission to the data repository, which would have allowed the other RCs to overrule any potential objections she might have. It was clear that she agreed she needed to be represented and couldn't just leave the decisions to others, yet she still felt overtaxed by the requirements of participation. In this case, the CCEN has relaxed the requirements for participation of the PI and allowed the cervical site to send junior PIs to participate in some of the meetings (Rebecca, 211).

A large amount of time has been spent in the SN on the collaborative work of negotiating roles and responsibilities. As will be discussed more fully in Chapter 6, there have been differences of opinions

about the role of the CC, both on the part of the FA representatives and of the RCs. At times, the FA representatives felt the CC was overstepping its bounds by taking on too great of a leadership role and asked them to back off and allow the RCs to take the lead, while at least one RC PI felt the CC wasn't taking enough of a leadership role and wanted them to step up and do more (Rebecca, 54). Resolving these conflicts – or preventing them from happening in the first place – is work that falls into the category of CWPs.

The funding cuts discussed in the structural work section may have exacerbated existing resentments on the part of the RC PI who felt the CC wasn't doing its part. Feeling pressured to do more with fewer resources, Beatrice wanted the CC to do more (Beatrice, 115). The collaborative work of participating in working groups and committees described here generally set the scientific direction of the CCEN, which means they have an impact on the data work that will be described later.

Throughout my interviews and observations, the SN appeared to spend substantially more time on collaborative work such as setting direction, clarifying the project objectives, defining roles and responsibilities and discussing how to work together than did the BN (e.g., SN_FN_4-5-12; SN_FN_7-10-12). There are many factors that could account for this difference, such as the age of the projects (the SN was in its first year, while the BN was in its 13th), the project structure or the simply the personalities of those involved; the evidence suggests that it stems from a combination of the BN's extensive experience collaborating and the SN's lesser focus on collaboration as a goal in and of itself.

5.2.2.3 Operational Work Practices (OWPs) of the SN

Much like the operational work practices of the BN discussed in section 5.2.1.3, the OWPs of the SN are the administrative and technological tasks done in support of the other types of work. The aim of these practices is to help the group's diverse and varying tasks function together as a whole. As with the BN, most OWPs are initiated by the CC, though all CCEN members must engage in some OWPs; for example, the FA and/or RC PIs must respond to frequent requests for information from the CC, including

information to use in scheduling calls and meetings, requests for IRB approval documentation and input on agendas for upcoming meetings. These interactions constitute operational work.

The CC is also responsible for creating the actual database in which the data will be stored and managing the process of data transfer (SN RFA). The CC has created a data management Working Group which consists of several members of the CC charged with preparing not only the database itself but also investigating security requirements and data transfer technologies. Once the RCs have prepared their data for submission and submitted their IRB approvals to the CC, the CC will direct them on how to upload their data to the CC database (Tina, 355).

The SN CC is able to take advantage not only of the OWPs established by the BN CC but also its systems. Once the data are ready to submit to the CC, the SN will utilize the data-transfer systems already in place for the BN. Additionally, like the BN CC, the SN CC hosts several email lists and is responsible for managing the flow of communications among the network participants (BN Manual of Operations, 16; Kieran, 207). By utilizing these existing systems, the newer SN CC is able to take advantage of the collective experience of the BN and other SFT CCs that have come before it.

As with the BN, the operational work is done in support of the other work of the CC. The database that is being created will be built to hold the specific data points the SN has decided to include. Those decisions, made as part of the data work done by the BN, will dictate the structure of the database itself. The management of the IRB forms allows the work of receiving and analyzing the data at the CC. Finally, the work that the CC does to coordinate the various pieces of the work to create the data repository are all being done within the structure set up by NIH in the RFA and honed in the discussions described in section 5.2.2.2 on collaborative work practices.

5.2.2.4 Data Work Practices (DWPs) of the SN

Data Work Practices are those activities the focus of which is the production and consumption of high-quality data. The DWPs of the SN center around the distillation of scientific questions into data points for

inclusion in the screening database. As mentioned above, the objective of the SN is to create a data repository to describe the screening process; however, deciding precisely which data points go into the database, out of the millions of data points available to the RCs, is a complex and difficult process. (This process was described in Section 4.2.2.3.) In short, the CC, FA and RCs must agree on the way to describe the screening processes that encompasses the experiences of all the populations addressed by the different RCs and allows the CC to perform meaningful and informative statistical analyses. What this means in practice is the creation by the CC staff and PIs of proposals around which data to collect, discussion about the data points for each organ group (breast, cervical and colorectal), then a new proposal, and the cycle is repeated. Eventually, the groups will need to settle on a complete list of data elements for inclusion in the repository. As of this writing, only a handful of common data elements had been identified that cut across all organ groups, including data points like demographics (ethnicity, insurance status, and sex) (SN_SCMTG_9-12-2012; SN_FN_8-28-2012).

Once these common data elements have been agreed upon, the RCs will need to do the data work of extracting the requested data elements from their local organization's database. This is not as simple as it may sound. Because many of the data points are from electronic medical records with very diverse data, some data points may need to be derived, or calculated based on other data. The simple example of BMI given earlier illustrates how one might derive a variable, calculating it from the patient's height and weight. Some sites may simply not have certain data (Joan, 149). Some data will need to be extracted from electronic medical records using Natural Language Processing, which uses algorithms to search free text fields for structured data. This is an approach that is still somewhat unreliable and presents issues for statistical analyses, in that the RCs are unwilling to share their algorithms, leaving the statisticians unsure of how reliable the data are (SN_SCMTG_9-13-2012).

This unwillingness to share the algorithms points to a lack of openness and sharing within the SN, especially when compared to the BN, where, as James noted, "everybody lifts up or opens up their books and shows everybody everything, because the ethos of the group is that if one member of the group

benefits, everybody benefits" (James, 125). It remains to be seen if the SN will develop that level of trust and openness, but it is worth noting that BN members recounted NCI representatives stressing the requirement to collaborate from day one of the BN, while NCI representatives from the SN have been reluctant to push the RC PIs to share anything they didn't want to (Rebecca, 158; Charlie, 85; Adam, 871). While a culture of collaboration can take time to develop, the lack of structural requirements to collaborate (i.e., funding mechanisms and penalties and evaluation programs) is clearly not helping that development along, as in the BN.

5.2.2.5 Local Scientific Work Practices (LSWPs) of the SN

The local scientific work of the CCEN encompasses those activities done at the local organizations (i.e., at the individual RCs or CC) whose aim is to achieve the scientific objectives of the CCEN but that do not require interaction with other CCEN entities and are not guided by collaborative protocols. Each RC in the SN has its own scientific objectives to achieve, independent of their work on the data repository (SN RFA). The RFA for the SN Research Centers requested that, in addition to their contributions to the data repository, each RC propose three projects that would be done by that RC. The RFA defines the scope of these projects, with two being fairly open to definition by the RCs, while the third must focus on one of two specific topics.

> Each research program must consist of 3 connected Individual Research Projects. Two of these required research projects are to be fully defined by applicants. The third project, however, must address (within the context of the scientific theme of the Center): (i) comparative effectiveness of established and emerging screening processes in community practice; and (ii) the benefits and harms of screening across recognized cancer risk profiles" (SN RFA).

Beatrice, an RC PI, suggested that this aspect of the RFA was to make the effort to build the data resources for the data repository more interesting for the RCs, as well as extracting more value from those efforts.

> [M]y understanding was the [RFA] was written so that we could propose science that we were interested in and that could be supported by the infrastructure that was being developed… And I think the flavor of the [RFA] was such that … the [RCs] didn't feel like a service center… like our only role wasn't to collect

> these data and then hand them over to the [CC]. Our investment was then we'd have this infrastructure to execute these scientific projects and future ones, building on that infrastructure (Beatrice, 93).

What precisely those three projects were for each RC was unknown for the first year of the project, because the RCs were not required to share their research proposals with either the CC or with one another. The CC hoped to gain a greater understanding of the projects in order to develop projects in which the entire CCEN could participate, but those trans-SN projects have been put on hold indefinitely. While the RCs eventually relented and sent their research plans to the CC, their reluctance again underscores the lack of trust and collaborative spirit within the SN.

5.2.3 Summary: CCEN Work Practices Typology in Illustration

From these short examples and a comparison of the practices between the Biomarker Network and Screening Network, a few things become clear. First, the basic collaboration work practices like participating in working groups and committees and the operational work practices like organizing meetings and managing IRB approvals are quite similar between the two CCENs. This is not particularly surprising, given that the two CCs share many staff. Also, scheduling conference calls, building data tracking systems and participating in working groups are generally the same regardless of the scientific questions being addressed by the project. On the other hand, the data work practices and local scientific work practices are dissimilar between the BN and SN. Though the overall aim of these data work practices is still to achieve high-quality data, the specific scientific objectives of the projects dictate the types of data that must be generated, which, in turn, suggests the data work practices in which the CCEN participants must engage. The science of discovering and validating biomarkers is quite different than the science of building a data repository on screening processes.

Second, the SN has had to engage in more extensive collaboration work practices and post-funding structural work practices. While some of this work can quite probably be attributed simply to the fact that the SN is a relatively young collaboration and still doing some work to organize itself, the amount of time and resources being spent on those activities seems high when compared to the recollection of BN

participants about its beginnings. Evidence suggests that the lack of cultural focus on, and structural requirements for, collaboration in the SN have also contributed to the SN's struggle to work together effectively.

Finally, none of the work practices described in this chapter exists in a vacuum, but, rather, are interdependent, influencing and impacting one another. The local scientific work, which is the most autonomous in this typology, still relies on the structural work performed by the FA in defining the overarching objectives of the CCEN. The operational work is done in service of the other types of work, coordinating the varied and complex tasks done by the CCEN participants into a functioning whole. Without this operational work, the data work could not be accomplished, as it requires CCEN participants to interact and exchange information in order to be successfully completed, and the collaborative work to smooth the way for those participants to work together. Finally, all this is done within the organizational structure set by (and possibly renegotiated with) the FA representatives.

The interdependent nature of the five types of work practices of the CCEN means that strengths or weaknesses in one area can exaggerate strengths or exacerbate weaknesses in other areas. This is apparent in the BN, where a strong structure has created great strength in the collaborative work, which, in turn, creates a highly functional set of data work and local scientific work practices. The science of biomarkers is difficult and littered with failed projects, yet the BN has made great strides in moving that science forward in just 12 years. On the other hand, the weaker structure in the SN appears to be making the data work of deciding on data points for the data repository more difficult than it needs to be by: a) forcing the SN to spend significant amounts of time organizing themselves; and b) taking away resources (i.e., funding) in the midst of the project. The end result is that some in the group are disappointed in the amount of progress that has made thus far (Evan 119, Charlie 31).

5.3 Network-Level Work of the CCEN

As discussed in the definitions of the work practices in Section 5.1, some types of work practices involve just one or two participants working independently or together, while other types of work practices involve all three types of participants (the CC, RCs and the FA) working together at one time. The work practices that involve the engagement of all three types of participants at one time can be considered "Network-Level Work" in the CCEN; that is, work that involves all parts of the network. The collaborative work practices, operational work practices and data work practices are all network-level work, as the CC, RCs and FA all engage in these types of work practices together. The remaining two types of work practices (structural work practices and local scientific work practices) are considered "Local Work" of the CCEN, as they are primarily engaged in either individually or with one other type of entity. Table 5.2 highlights this distinction.

This distinction between network-level work and local work is important for this dissertation, as the network-level work is, really, the heart of the CCEN. One of the defining characteristics of a CCEN is that it is formed in service of a scientific objective *that can only be completed within a network structure*. The reason for forming a CCEN is to take advantage of the parties working together to build something that is, as heard in several interviews, a whole that is greater than the sum of its parts (Thomas, 231; Nigel, 311; Rebecca, 199; Joan, 267). As such, the network-level work, defined as the work done by all three entities working together, is the main work of the CCEN as a whole. Furthermore, the network-level work is the area where the CC does the most work and has the greatest impact on the work of the CCEN. Again, as the main question of this dissertation is how a CC facilitates the work of the CCEN, it is important to focus in on that work which involves the entire network and, especially, the CC.

	Type of Work Practice	Example CC Tasks	Example RC PI Tasks	Example FA Tasks
Local Work	Structural (SWP)	• Negotiating revisions to scientific objectives in case of unexpected funding changes • Suggesting changes to RFAs for new funding cycles	• Submitting grant proposal • Agreeing to accept funds and follow RFA • Suggesting changes to RFAs for new funding cycles	• Writing the RFA • Making funding changes • Selecting grantees • Revising scientific objectives as needed
Network-Level Work	Collaboration (CWP)	• Deciding how to interpret RFA in practice (with FA) • Negotiating questions of roles and responsibilities • Prioritization of projects in view of limited resources	• Participating in committees and working groups • Fulfilling role in CCEN • Attending meetings	• Deciding how to interpret RFA in practice (with CC) • Evaluating grantee progress, reviewing progress reports • Ensuring CCEN moving toward scientific objectives
	Operational (OWP)	• Organizing meetings and conference calls • Programming data entry system • Coordinating protocol development • Management of IRB approvals	• Sending questions to CC • IRB documentation submission to CC • Registering for meetings	• Creating agendas for meetings • Scheduling site visits
	Data (DWP)	• Distilling scientific questions to data points for collection • Statistical analyses • Protocol development and study design • Database design	• Research to answer information requests from CC about data • Reviewing analysis results • Writing protocol with CC staff • Data entry	• Representing interest of FA in protocol development • Reviewing analysis results
Local Work	Local Scientific Work Practices (LSWP)	• Development of novel statistical methods	• Running assays or experiments to discover new biomarkers	

Table 5.2: Network-Level Work of a CCEN

5.4 Chapter Summary

In this chapter, I have presented a typology of the work that takes place in a CCEN. This typology allows us to take a closer look at what the CCEN is and does, expanding our knowledge of this previously undescribed type of organization. I have illustrated the five types of work with examples from both the Biomarker Network and Screening Network, noting the similarities and differences between the work

practices of the two networks. I have shown that the collaboration work practices and operational work practices vary little between the BN and SN, though the data work practices and local scientific work practices are quite different between the two. The scientific objectives of the projects dictate the types of data being generated, which, in turn, suggests the data work practices performed in the two CCENs. Additionally, during this study, I observed that the SN spent substantially more time organizing itself through collaboration work practices and even on the structural work required after substantial cuts were made to the RC funding, resulting in a less productive first year than many had hoped. This was in contrast to the BN, which seems to spent little time engaged in either of these practices, freeing up time to spend on the data and scientific work practices in service of the BN's scientific objectives. The five types of work practices were also shown to be highly interdependent, with weaknesses or strengths in one work category having a significant impact on the ability of the CCEN to engage in work in other categories.

One of the implications for scientists from these findings is how important the structural work of the CCEN is. In an organizational structure where the types of work are as interwoven and interdependent as they are in a CCEN, the ramifications of these early decisions can be felt for the entire life of the project. If the structural work is done appropriately, the work proceeds as well as can be expected. However, when the structural work goes awry, as seen in the SN, the ripple effect of the decisions made while designing the RFA can be crippling to the project's scientific progress.

Furthermore, I have divided the types of work into *local work*, which involves only one or two CCEN entity, and *network-level work*, which involves all three types of entities working together. As one of the defining characteristics of the CCEN is that achievement of its scientific objectives required a network of scientists working together. As such, the work done by the CCEN participants working all together is at the heart of the CCEN's work. The distinction between local and network-level work also provides an analytical lens through which we can focus more precisely on the work of the CC, as the work practices involved in network-level work are those through which the CC has the greatest impact on the work of the CCEN. Precisely how the CC does this is the subject of the next chapter.

Chapter 6: Facilitating the Network-Level Work of the Coordinating Center Enabled Network

The central question of this dissertation is: How do Coordinating Centers (CCs) facilitate the network-level work of the Coordinating Center Enabled Network (CCEN)? In the previous chapter, I presented a typology that described and illustrated the work of the CCEN, zeroing in on what I consider the "network-level" work, defined as the work done in the CCEN that involves all parts of the network – Research Centers (RCs), the CC, and Funding Agency (FA) representatives – working together. This is in contrast to the "local" work of members of the CCEN, in which one or possibly two entities work together. This network-level work is at the center of what a networked science project like a CCEN is designed to do; namely, to take advantage of the skills and resources of a variety of participants to achieve a scientific objective that could not be achieved by a single investigator working alone.

The CC has a distinct role to play in that network-level work. While that role has been described using a variety of words, as illustrated in the sections that follow, the term "facilitation" is commonly used when discussing the work that the CC does in support of the collaboration (Rebecca, 22; SN RFA; Tina, 463; David, 154; Edith, 305). However, what exactly that facilitation entails in terms of day-to-day tasks is rarely spelled out in great detail in the RFA, leaving the CC to interpret its mandate based on its previous experience facilitating other collaborative projects. In fact, it is precisely this application of expertise and experience to the challenges faced by the CCEN that allows the CC to make an impact on the project's progress toward its scientific goals. In this chapter, I will present examples from both interviews and observations of situations in which the CC draws upon its collective and individual expertise and experience to facilitate the work of the CCEN. By doing so, I will begin to create a more precise definition of what it means to facilitate collaborative work in a CCEN.

6.1 The Coordinating Center's Role in Network-Level Work

The terms "facilitate," "support," "lead," and "coordinate" are used in the BN and SN RFAs to describe the responsibilities of the CC in the network (BN RFA, SN RFA). Yet what precisely those terms mean is

not always clear and must be interpreted and defined by the CCEN participants. In this section, I present several examples of the work the CC does in the BN and SN in order to illustrate how the CC has interpreted and reified their mandate. Table 6.1 highlights some of the example tasks of the CC in the three types of network-level work.

Type of Work Practice	Example CC Tasks	Example RC PI Tasks	Example FA Tasks
Collaboration (CWP)	• Deciding how to interpret RFA in practice (with FA) • Negotiating questions of roles and responsibilities • Prioritization of projects in view of limited resources	• Participating in committees and working groups • Fulfilling role in CCEN • Attending meetings	• Deciding how to interpret RFA in practice (with CC) • Evaluating grantee progress, reviewing progress reports • Ensuring CCEN moving toward scientific objectives
Operational (OWP)	• Organizing meetings and conference calls • Programming data entry system • Coordinating protocol development • Management of IRB approvals	• Sending questions to CC • IRB documentation submission to CC • Registering for meetings	• Creating agendas for meetings • Scheduling site visits
Data (DWP)	• Distilling scientific questions to data points for collection • Statistical analyses • Protocol development and study design • Database design	• Research to answer information requests from CC about data • Reviewing analysis results • Writing protocol with CC staff • Data entry	• Representing interest of FA in protocol development • Reviewing analysis results

Table 6.1: Examples of Three Types of Network-Level Work Practices in a CCEN, by Participant, with CC highlighted

6.1.1 The Biomarker Network Coordinating Center and its Network-Level Work

According to the RFA, the BN CC "supports statistical and computational analysis, data management and protocol development, and informatics infrastructure and coordinates network-wide meetings and conferences" (BN RFA). Here, we see the words "supports" and "coordinates" but the RFA contains few details on what precisely those words mean or how the CC should accomplish those goals. The BN CC has been working as part of the BN for more than 12 years and has developed systems and processes in

response to challenges they have faced over the years. These systems and processes represent the collective and individual experiences and expertise of the BN CC and their application to the BN's work positively impacts the CCEN's progress toward its scientific objectives. Here, I describe examples of how the BN CC approaches their work in the BN.

6.1.1.1 Collaboration Work Practices of the Biomarker Network

Collaboration Work Practices encompass the work that members of the CCEN do in order to organize themselves as a Network. Included here are the activities that affect how the work of the CCEN gets done, such as negotiating resources, roles and responsibilities; enforcing compliance with both cultural and regulatory requirements; and participation in subgroups or working groups. In the BN CC RFA, there is no mention of such a role and none of my participants specifically mentioned this work of the CC as a part of its responsibilities. Yet I witnessed many instances of collaborative work practices in my observations and oblique references to these CWPs in interviews, as discussed below. The effect of the collaborative work practices described here on the BN is that the project stays focused on its scientific objectives without getting sidetracked.

The prioritization of work falls into the category of collaboration work practices, as it determines which activities will receive funding and be allowed to proceed with support from the CC. As discussed in section 5.2.1.2 on the CWPs of the BN, the BN experiences few problems working together collaboratively and sharing their work openly. Thus, the CC spends little of its time focused on CWPs concerned with the negotiation of roles and responsibilities or developing a culture of sharing. However, the BN is still a large network of prominent scientists working in a rapidly changing and developing field; this results in a plethora of potential projects and directions the BN could possibly undertake. With limited resources, the BN does not have the funding that would allow them to follow every promising scientific lead, so some projects just won't get done. Though there is nothing in the RFA that specifically assigns this task of project prioritization to the CC, the BN CC recognized that they have a unique perspective on the overall work of the BN, a perspective that allows them to stay focused on the

overarching goals and evaluate which projects will move the BN closer to achieving those goals (Adam, 109; Karen, 109). As such, the CC has devised several processes to deal with this issue of project prioritization over the years. One of these is a system of evaluating proposed projects.

During the first grant period, the BN CC realized that they were unable to coordinate all of the studies being proposed by RC PIs; they simply did not have the resources. So, the CC PIs went through each proposed project and rated it based on criteria such as scientific impact and required resources. They then ranked the projects according to these criteria. At first, FA representatives and the Executive Committee were very resistant to this approach, thinking the CC has overstepped its bounds, and rejected the idea. However, the CC presented their rationale and their methodology to the NCI and EC at their next site visit, and the visitors were quickly convinced that this was the right approach.

> "So our proposal to NCI is we help them to identify because after we had so many team projects and the NCI thought we should coordinate all. And we said no, no, that's not possible. So we offered and we said we're going to read those team projects and identify which one we think is the good one. And good in the sense that their prospective collection, in other words they do not have bias and it's more likely to be very useful by the end. And so we will rank them as higher priority and we propose that we coordinate those. So at first they were not happy. They wanted us to do all so actually they had a site visit, I think in year one, and that was one important question. So the two people, the NCI project director and the two chairs of BN [visited us for our site visit]. We presented our thinking why, and we tell them here's our rankings. And so after that - and after our presentation, they had closed discussion. And so after that then it's yeah, we'll do it the way you guys say it. And so then they never raised that issue again. … Because our criteria is clear. If it is approved, the study design principle is a prospective collected, and those are high quality ones that we rank them high. And then each organ group, they cannot do many. So he says they only can do one, because it's expensive to do well" (Adam, 371).

The CC's experience with study coordination and scientific expertise allowed them to make a rational, evidence-based case for which studies should receive access to the CC's limited resources. Furthermore, they did so using criteria that were objective and drawn from the scientific objectives of the BN. The effect of this action on the part of the CC was twofold. First, by creating an objective system of scoring based on scientific merit, the CC eliminated some of the political issues of evaluating the projects; scientists are not immune to the pressures of supporting a project because it is proposed by a powerful colleague. Second, by providing leadership in the area of project prioritization, the CC saved the BN from spending a substantial amount of time: had the CC not done this, all the work of devising criteria, scoring

each project on those criteria, then ranking them would have eaten up a significant amount of time in future Steering Committee meetings.

The CC itself also must often prioritize its own work, which is limited to the coordination of 4-5 validation studies at one time. In fact, a significant amount of each CC operations team weekly meeting is spent updating one another on progress on given tasks, information the project manager then uses to decide how those tasks should be scheduled and prioritized (e.g., BN_FN_2-22-12, BN_FN_4-4-12, BN_FN_4-22-12). Even with this limitation, the work often grows beyond their capacity. For example, one current validation study began with just a few sites, but then quickly expanded to more than 20 when it became apparent that the original sites would be unable to meet the study's recruitment goals. Each additional site means additional training, additional VSIMS users, additional questions for the CC staff to field, all without a corresponding increase in funds. Given this often overwhelming amount of work for which the entire CCEN depends on the CC, Karen, the project director, noted that she must do a substantial amount of prioritization of work for herself, as well as for her staff (Karen, 241). She said that she rarely has questions about what should be prioritized, but when she does, or when the prioritization might be controversial, she confers with the BN CC PIs (Karen, 113).

6.1.1.2 Operational Work Practices of the Biomarker Network

The operational work practices are the administrative and technological tasks done in support of the BN's work. Their aim is to help the group's diverse and varying tasks function together as a whole. In the BN's coordination of clinical validation studies, a large amount of operational work is required to support the completion of the study. The CC has been coordinating these clinical validation studies for more than a decade, gradually honing their practices over time to develop systems and processes the aim of which is to keep the studies moving as smoothly and efficiently as possible.

The CC's tasks that fall into the category of operational work practices include the coordination of the work around developing the protocol, distributing the eligibility flow charts, scheduling site visits and

monitoring protocol compliance. What this means, more specifically, is interacting with the involved parties to ensure the work gets completed on time and according to the approved study protocol. While the CC contributes substantial scientific work and expertise to protocol development process, as discussed in the next section on data work practices, the CC is also responsible for tracking the conversations about the protocol, checking to make sure the protocol follows the best-practice guidelines they have developed, scheduling approval votes by the EC, distributing the protocol to the EC for review, posting the protocol on the website for others to see, even formatting and spell-checking the final document (Sarah, 105; BN_FN_2-29-2012).

The CC also builds the technological systems that are used to coordinate the collection of the data. In the early years of the BN, the CC spent a substantial amount of time building the systems that today make them efficient enough to be able to coordinate 4-5 validation studies at a time with minimal staff. (These systems are also being used by the SN.) The Virtual Specimen Information Management System (VSIMS) was developed by the BN CC to handle the data entry and tracking that takes place in a validation study (Karen, 93). Each of the individual biomarker validation studies receives its own instance of the VSIMS, with its data entry and database customized to the needs identified and described in the protocol development (Karen, 151). Several programmers are on staff in the SFT team to do this work. The VSIMS system is in continuous use and is evaluated regularly by the IT staff to see how it can better support the BN's validation systems (BN_FN_2-22-2012). Because VSIMS has been in use continuously for over a decade, it encodes lessons learned by the CC from previous studies, benefiting future studies through incremental improvements made by the IT staff.

The aim of these operational work practices is to keep the work of the BN moving toward its scientific goals by ensuring the work of the various participants function together as a whole. To do so, the CC utilizes its experience managing previous studies to ensure that the moving pieces of each project come together as seamlessly as possible. The result of this operational work on the part of the CC is that many of the administrative tasks that might otherwise fall to the RCs are, instead, performed by the CC. A study

site participating in a clinical validation study no longer has to create its own data-entry system, instead relying on the one provided by the CC. An RC PI who wants to propose a new project can do so without having to spend time figuring out the required format or distributing copies to interested parties; all of this is done by the CC. Additionally, these operational tasks are done in a way that takes advantage of the lessons learned by the CC over their 12 years of working on the BN.

6.1.1.3 Data Work Practices of the Biomarker Network

Data Work Practices are those activities the focus of which is the production and consumption of high-quality data. The main objective of the BN CC in the area of data work is providing support for clinical validation projects to ensure the data produced by the study sites and received by the CC will result in the strongest, more informative statistical analyses possible (Karen, 38, 50). Focusing on high quality data ensures the results of the validation study will provide valid, believable evidence for or against the biomarker. It is in this area of data work that the CC's experience and expertise play out most explicitly, with the greatest impact on the BN's progress toward its scientific objectives. The CC learns important lessons from each validation study they coordinate, lessons which are then incorporated into improved processes for subsequent studies. Two specific examples of data work processes that have developed and improved in response to previous experience of the CC can be found in the CC's work on: a) the development of common data elements (CDEs) and data-entry forms, and b) on the creation of eligibility flowcharts.

6.1.1.3.1 Common Data Elements and Data-Entry Forms

While the RC PIs writing the protocol and leading the study are responsible for defining the aims of the validation study, they rely heavily on the expertise of the CC in both leading the conversations required to discern precisely which data points should be collected and how to represent those data points in the data entry system that will be used by the study sites.

The actual data needs of a validation study will vary based on the proposed clinical purpose for the marker. When possible and appropriate, the CC tries to standardize the data collected from each study into common data elements (CDEs). To reiterate, a CDE is a data point with a standardized definition. For example, the CC might use a CDE to standardize the way smoking data are collected from all BN studies, requesting "Cigarettes Per Day" and "Years Smoked" from each study participant for studies collecting smoking data. Because these CDEs have been used in the past, in both BN and other SFT studies, they have been vetted and shown to be well described and useful (Kieran, 185). The CC has compiled a list of standardized CDEs, which allows for the more rapid development of protocols, in that a PI can go through the list and select those most applicable to the study (Kieran, 181). If new CDEs are necessary for a new study, they are created and may be incorporated into future projects.

The information provided in the protocol is used by the CC to develop the CDEs, but conversations are still required to ensure that the right data points are being collected. When asked about the process of developing the data-collection protocol, Edith described a meandering, iterative process of working with the study PIs to nail down precisely which data they want to collect on the various participants. She noted that the process is time-consuming because of the need to have so many conversations and the necessary effort it takes to ensure that everyone is talking about the same thing. She noted that it can be difficult to get answers from busy PIs about the minutiae of the data points.

> "I could walk you through [the process] but it's really more like wandering around in the forest. It's an iterative process. I actually have a slide about this with arrows going everywhere. … So when someone proposes a project they usually say, 'We want to collect these kinds of information about the patients that are supplying these samples or the patients that are being analyzed in some way.' And they can be fairly general. And so we will talk to them and say, 'Okay, let's try to come up with a specific list of all the data points that will collect this information that you want.' Sometimes we use data lists from other projects and adapt them. And we'll send them just either an Excel sheet or a Word document that is more precise. Then they'll say, 'Oh yeah, well we really didn't mean that, that that. We meant this, this, this. And this is what this other study did and the way they collected it, but that's not the way we think about it so we want it phrased differently'" (Edith, 174).

As we can see from this quote, the PIs can find it difficult to express precisely which data points they want for their research. This is understandable, as they are not necessarily data experts. One of the ways in which the CC adds value to the data work of the CCEN is by leading this process of choosing the data

points in a way that ensures the data collection will lead to a valid analysis. If the data that are collected by the study sites are not actually what the researchers needed in order to confirm the validity of the biomarker, the entire study was a waste of time and money. To avoid this, the CC draws upon its previous experience creating and using CDEs to ensure the appropriate and necessary data are collected.

When asked for a specific example of a time when she experienced that disconnect between what an RC PI *thought* s/he wanted and what s/he *actually* wanted, Edith described this incident:

> A project that we're working on currently, one of the forms is collecting information on nodules, lung nodules. And we have never collected that kind of information before. So, we're collecting information from either CT scans or MRIs. And there are a lot of technical data points that have to do with running a CT machine or an MRI machine that we don't necessarily know what they really mean. But it's obvious to the clinicians who do it all the time. And so we've had a lot of back and forth about how best to organize that information and exactly what information is needed. And we finally realized that what we really wanted was not information for every CT scan. But we wanted information on every nodule, whether it was a CT scan or an MRI. And then we'd follow that nodule and follow up, and that was a huge difference. And so just working that out took a lot. For instance, they sent us a long list of data elements from their database but it assumes you know what they're doing. So, you know, you start with what they give you and you try to figure it out but then you have to go back to them and say, "Well, I think this means this and it would look this way in our system. Tell us what needs to change" (Edith, 271).

Here, we see how Edith and her team, through a series of interactions with the study PIs, discovered that the data the project required centered not around a CT scan or MRI, but around each nodule and what was done to and known about it. These are two fundamentally different data structures, a fact that would only be apparent to experienced data collectors. In the example above, it is particularly interesting to note that Edith and the CC team weren't experts in data around CT scans and MRIs, but, rather, experts in the work needed to collect the right data. The CC cannot possibly have expertise in every area of medicine and existing CDE in every area of cancer. While the scientific knowledge they do have proves very useful, their skill in leading conversations to *appropriate* data, as evidenced by both these examples, may be even more important to the outcome of a validation study.

After being identified and thoroughly defined, the CDEs are then developed into the data-entry forms used by the study sites to enter data into the Virtual Specimen Information Management System (VSIMS). The CC staff turns the data-collection protocol into computerized data entry forms that will be

filled out for each participant and builds the database to store them (Karen, 205). These forms not only contain entry blanks for all required data, they also contain programming logic to test participant eligibility and validate the data entered. As appropriate, checks are built into the various data-entry fields, such as validating ranges for weight and height. This is another point where institutional knowledge can be used for developing checks. A previous bad data-entry experience can lead to a new informal rule to always, for example, check that a certain lab measurement is entered as an integer and not a range, as the CC learned on one project. The benefit of the time and effort that goes into the creation of such precise and validated data-entry forms is the resulting quality of the data. “What we can do is, every field that’s entered we can check that against another forms field. Or we can go into our specimens system and check that against something entered in there. That gets very customized per study. ... Every one of those studies needs special customization. But what that special customization brings is quality data. So whoever is entering the data, it helps them ensure that the quality that they’re entering, that they’re making fewer mistakes" (Tina, 107) Again, this relentless focus on data quality on the part of the CC staff adds value to the scientific work of the RCs.

6.1.1.3.2 Eligibility Criteria Flowcharts

A second example of the CC’s experience and expertise being used to improve systems and processes that result in high-quality data is the work they do developing eligibility criteria flowcharts. When designing a clinical validation study, it is crucial to have precisely defined and scientifically appropriate criteria to determine who is eligible to be enrolled as a participant. After early BN studies where the eligibility criteria encoded in the protocol and, subsequently, in the data entry forms, turned out to be not quite right – either eligible participants not being enrolled or ineligible participants being enrolled – the CC developed an intricate process for ensuring all parties involved were on the same page regarding eligibility and that this understanding of the eligibility was precisely encoded in the protocol (Edith, 249). This process was designed to ensure that both: (a) the PIs themselves were clear on the implications of the eligibility criteria they had proposed; and (b) there were no misunderstandings in terminology or intention

as the CC interpreted what the PI had proposed (Edith, 184). Edith described her goal in developing the eligibility criteria flowcharts as creating an explicit documentation of who will be included and excluded in the study in such a way that the logic contained in the flowchart can be easily programmed into the data entry system, all with the goal of making sure the proper participants are recruited.

> "My goal in an eligibility flowchart is to combine in one document all the online phrasing of each data point that's required to determine exclusion and inclusion. And also the place in the database where the programmer can find where that data point will be stored. And also the – so you found this data point, it's got this value for this person, what do you do with that? And so the idea is for each, to create a point where you start off with a data point. You describe everything about it, and you have arrows that point to the options depending on the value of the data point" (Edith, 222).

In essence, Edith's work on the eligibility flowcharts acts as a bridge between the data work of identifying the eligibility criteria and the operational work of building the instance of VSIMS.

To accomplish the development of the flowcharts requires iterative conversations with the PI of the protocol, Edith, other CC staff, and the project statisticians at the CC, who are called in to evaluate the eligibility criteria and calculate the number of participants likely to be recruited under the proposed rules at the proposed sites. From these conversations, the CDE specialist creates a flow chart that makes explicit which data points determine eligibility. For example, the first data point used to determine eligibility might be age, removing any patients under 60. Next, a check on the patient's previous diagnosis of cancer might remove more patients from eligibility, possibly including only those with no previous cancer diagnosis. Such detailed attention to eligibility decisions allows the PI to adjust the recruitment strategy or the study sites before the study begins, rather than after, saving time and money. There may still be adjustments once the project gets underway, of course, but they are likely to be less dramatic thanks to the forethought involved in this process. In the end, this work should result in fewer ineligible participants being enrolled and fewer eligible participants being missed. This helps the study reach its recruitment goals more quickly and with fewer headaches.

The aim of the CC data work practices described here is to ensure that the study sites generate high-quality data that can then be sent to the CC for analysis. By focusing their efforts on using their

experience with previous validation studies to improve future data collection in clinical validation studies, the CC takes advantage of the skills and knowledge, both individual and collective, that they have developed over the more than a decade of coordinating such studies. The result of this focus is studies that operate more smoothly because the routine challenges of designing a study, like deciding which data to collect and which patients are eligible for inclusion in the study, have been addressed through the development and use of evidence-based systems and processes that reflect the CC's experience.

6.1.2 The Screening Network Coordinating Center and its Network-Level Work

The SN RFA describes the following role for the SN CC

> Specifically, the [CC] will serve as a repository for pooled data (to be gathered by the Research Centers) on the screened populations, the screening processes, and ultimate clinical outcomes. The [CC] must also provide necessary expertise and lead for standardizing pooled data and their analyses. In addition, the [CC] is expected to facilitate other trans-[SN] activities and collaborative research (SN RFA).

The mandate for creating the data repository is clear, but the remainder of the RFA text gives the CC little guidance on how to "lead" the standardization of the pooled data or how to "facilitate" the trans-SN activities and collaborative research. While the SN CC is still quite new – it was just finishing its first year of operations at the end of my observations – many of its members also work on the BN CC and have for years. As such, they are able to take advantage of the experience and knowledge they have acquired over the past decade of coordinating clinical validation studies, even when the precise tasks are very different than what they do in the BN. However, as will be shown in the sections that follow, the CC's attempts to apply their experience and expertise did not always go as planned, which has had a substantial impact on the scientific outcomes of the project.

6.1.2.1 Collaboration Work Practices of the Screening Network

To reiterate, the collaboration work practices of a CCEN are the work of negotiating and deciding how to work together as a network. These CWPs are one area in which the SN has had difficulty translating the skills and knowledge it developed as the CC of the BN to a new project. Specifically, the SN struggled with the negotiation of roles and responsibilities, especially the division of labor between the FA

representatives and the CC. Before the CC received their funds, but after they had received word they had been selected as the CC, the NCI branch chief responsible for the SN visited the CC at FHCRC and emphasized how important it was that they take a leadership role in "governing" the SN (Adam, 707).

Early on in SN, as it became clear that the vision of the CC and NCI representatives were not in synch, Rebecca (an NCI representative) asked the CC to write a list of their roles and responsibilities for their work in the SN. Tamara, a CC staff member, described how Charlie, the CC PI, used the word "leadership" in several areas of their responsibilities description. They were promptly, and in no uncertain terms, told by Rebecca to remove that word from the entire document.

> To give a little bit of background we were asked, pretty close to the beginning, we were asked by NCI to create roles and responsibilities, a list of roles and responsibilities for our self, NCI, the steering committee, and then the research centers. And so [Charlie] wrote those up and he used the term leadership in a lot of what [CC] was responsible for. You know be a leader in the organ groups, be a leader in getting the data and [the FA representatives] really balked at that. And they really didn't think that – they thought that we were overstepping our bounds in that the actual term leadership was kind of a poor choice of words, in their opinion. And so [Charlie] kind of fought it for a little bit and then he's like okay, I'll let it go and he rephrased it.
>
> And I think all along we felt that we were, we were to be the leadership of data analysis and coordinating the data, but I think NCI felt that we should actually be more of a team player… You know, yes, we were the team player but we were also the folks that were responsible for the bigger picture and for, again, kind of pushing the others forward. So at first NCI didn't see us in the leadership role; they felt it should be more of a collaboration. So we then started operating more as a collaborative part and NCI actually came back and said you know what, you should be doing being the leadership and taking more of a lead in the steering committee, taking more of a lead in the working groups and the [RCs]. And we said well, you know, look that's what we had intended and then you said no. And they said oh, well, I think maybe that was a mistake. And so they do now, I think finally, I think we're finally more on the same page where they feel that we are kind of leadership in the realm of the data, getting the CDE's, getting the data, you know making the decisions around that (Tamara, 71).

Other CC staff also expressed cautious optimism that they were going to be allowed to play what they viewed as their proper role in the SN going forward. They felt they had the knowledge, expertise and position to lead the SN forward by coordinating and facilitating the work of the SN network.

This conflict over roles and responsibilities between the FA representatives and CC had the unintended consequence of sparking a conflict between the CC and the RCs over their respective roles. When the CC was told to back off and let the RCs come forward to take more of a leadership role, this placed greater

demands on the time of the RCs, who had not planned for this work. As the RC PIs were forced to spend more time on working groups and deciding on CDEs, they have had less time to spend on their individual-level projects. Combined with the funding cuts, these extra demands left the RC PIs frustrated (Beatrice, 105). This issue will be discussed in greater detail in relation to the data work practices of the SN in Section 6.1.2.3, below.

Each collaborator only has so much time that can be spent on participation in the SN. When that time is spent negotiating how to best work together, those negotiations take time and energy away from the time available to spend making scientific progress. In a new collaboration, some negotiation may be necessary or desirable, as two CC interviewees told me (Adam, 737; Tina. 455), but the levels of frustration expressed by interviewees from the CC, the RCs and the FA (Charlie, 73; Adam, 637; Tamara, 71; Rebecca, 57; Beatrice, 105) indicate that the energy put into these negotiations left the participants unsatisfied with the distribution of responsibility.

I also observed many instances of frustration that manifested as arguments about how the work should be moving forward. One of these instances involved deciding the agenda for the second in-person meeting. The CC PIs wanted to feature a colleague with a reputation as a big thinker in the area of screening to lead the large group discussions about the future of the SN. Rebecca, an NCI representative, resisted this idea, worried that the RC PIs would feel like this person was an outsider, brought in by the CC to tell them how to do their work. Some in the CC felt it was their role to set the agenda and were visibly frustrated at Rebecca's rejection of their idea for a speaker. In one CC-NCI meeting in which this was argued over, a CC PI angrily told Rebecca that she needed to trust them. Rebecca, equally angrily, replied that she was just doing her job and looking out for the SN's interests. Eventually Rebecca agreed to the speaker, but the hard feelings over who was controlling the agenda seemed to linger (SN_FN_NCI_6-12-12, SN_FN_NCI_6-26-12).

What is especially noticeable about these conflicts within the SN over roles and responsibilities is how absent such conflicts are in the BN. As discussed in both sections 5.2.1.2 and 6.1.1.1 on the collaboration practices of the BN and the BN CC, respectively, the BN experiences few such conflicts over how to work together. In fact, interviewees described exceptionally strong working relationships among all involved in the BN. So why, exactly, has the SN struggled to work together. Evidence presented in this dissertation points to a combination of several issues. First, the SN experienced large funding cuts imposed on the RCs and the CC, leaving everyone involved struggling to complete the work that they had proposed in their grant applications, paving the way for collaborators to feel overworked and resentful of being asked yet again to do more with less. Second, the insistence by the FA representatives that the CC take a back seat to the RCs in terms of leading the work of the SN kept the CC from doing the work they felt they had been asked to do; namely, leading the work of building the data repository. This not only left the CC staff and PIs confused and unhappy, but kept them from using their experience and expertise to make the work go more smoothly, the way they have in the BN. While it is impossible to say that any one of these factors, alone, caused the struggles the SN has experienced in its collaborative work practices, it is fair to suggest that they have all contributed.

6.1.2.2 Operational Work Practices of the Screening Network

The operational work practices in which the SN CC engages are very similar to those of the BN, and the SN CC has been more successful in translating their skills and knowledge in this area to the work of the SN, taking advantage not only of the OWPs established by the BN CC but also its systems. In addition to the operational duties mentioned above in the section on BN OWPs, like organizing calls and meetings, the SN CC's OWPs in support of the CCEN are focused primarily around coordinating the discussions about which data points to submit to the proposed database. This work involves collecting information from the SN RCs about which data they are willing and able to share, as well as answering questions about IRB issues (SN_FN_05-22-2012). Like the BN CC, the SN CC hosts several email lists and is responsible for managing the flow of communications among the network participants (Kieran, 215).

The CC is also responsible for creating the actual database in which the data will be stored. The CC has created a data management Working Group which consists of several members of the CC charged with preparing not only the database itself but also investigating security requirements and data transfer technologies (Tina, 355). In effect, this group is the translational link between the scientific work of designing the database and the operational work of the programmers creating the database. Once the data are ready to submit to the CC, the SN will utilize the data transfer systems put in place for the BN. In that way, the newer SN CC is able to take advantage of the collective experience of the BN and other SFT CCs that have come before it.

6.1.2.3 Data Work Practices of the Screening Network

Data work practices are those activities the focus of which is the production and consumption of high-quality data. The DWPs of the SN center around the distillation of scientific questions into data points for inclusion in the screening database. As a reminder, the main goal of the SN is to create and populate a database that describes the screening process for three cancers: breast, colorectal and cervical. During the period of this research, the majority of data-related work being done by the SN CC was focused on securing agreement from the SN RCs about which data elements to send to the data repository and in what form. As they work toward that objective, the CC is utilizing its extensive knowledge of cancer-related data elements to steer the group toward choosing the data points that will result in the best analyses. The CC's experience in data collection for the BN gives them a deep understanding of the potential pitfalls in collecting and harmonizing such data. However, the CC has experienced difficulty getting the RC PIs to agree on which data points to collect and struggled to take full advantage of their experience and knowledge in data collection procedures (Edith, 315; Nigel, 509).

The difficulties experienced by the SN in agreeing upon data points to collect was described in detail in section 4.2.2.3, but in summary, the plan of the CC to build consensus among the RCs on which data points should be deposited in the data repository was to first propose high-level "data concepts" to which the RCs could respond with more specifics on the data available to them. The CC also planned for

primarily cross-organ data elements, data points that could be collected on all three organs. When that did not work, the CC shifted tactics and asked the RCs to work in organ-specific groups to propose data points that made sense for each cancer – cervical, breast or colorectal.

It is important to note, also, that even within the CC, coming to an agreement required time and effort. While at SN meetings the CC's CDE proposals were presented as a cohesive strategy, the CC itself had engaged in weeks of discussion about which CDEs to request, when and how to request them, what the permissible values could be and how the CDEs would potentially be extracted from the RC databases (e.g., SN_FN_3-20-12, SN_FN-4-10-12). Some of these discussions were about the science behind the CDEs but often discussions focused on the social and organizational issues about requesting data, including willingness and cooperation on the part of the RCs. I did not observe any instances where these CC disagreements were communicated with the larger CCEN audience.

One of the SN RC PIs interviewed for this study felt that the main issue with securing agreement on data points was that the CC simply didn't have experience with or understand health plan data. The PI felt that the CC's lack of knowledge about the data gathered in healthcare delivery systems was putting more of the burden on the RC PIs, taking the effort they needed to devote to their local work. The end result, the PI believed, was that the RCs and, therefore, the SN, were less productive than they would have otherwise been.

> Beatrice: I think that data from healthcare delivery systems is very different [than what] would [be] collect[ed] in a large observational study like a cohort study data [or] that you would collect from a randomized trial. And I think our role in the collaboration is to sort of advise and inform folks who are not familiar with health plan data, to the extent that that will facilitate the work in coming up with the common data elements and what's feasible. I think that over the past year the [RCs] have led more of that effort because I think that the [CC] hasn't been familiar with the kind of data that we have and how we access it. But I really think that the [CC] should have been leading that effort you know rather than the [RCs]. But so, in terms of collaboration, you know we're scientific partners, we're data partners. I think that there's been a little bit of confusion as to who's leading the common data, you know who's leading the work to create the data repository.
>
> Betsy: And where do you think that confusion stems from?

Beatrice: I think that confusion stem from folks at the [CC] not understanding our data. It's hard to lead something when you don't have a knowledge of the systems that you're relying on to extract the data.

Betsy: And what has been the effect of that confusion?

Beatrice: I think we've been less productive. You know, I think that the first meeting was not very productive, the first in person meeting. This [second] meeting went better, but I think it went better because people from the [RCs] were asked to lead the organ specific breakout groups whereas I think the [CC] has been more like relying on the [RCs] to provide information and then they go back and sort of try to synthesize information for the next step.

Betsy: And so you think that the [CC] should be leading that process –

Beatrice: Yeah.

Betsy: More than they're doing?

Beatrice: Yeah um hmm, for sure.

Betsy: What effect would that have if the [CC] were leading versus the [RCs]?

Beatrice: Well it think it's an effort issue right? It's the coordinating center. I mean that's within the scope of the work of the coordinating center. That's what they're funded to do. You know I think that the [RCs], it's all effort right? And so if you're spending effort on this, it's less effort that you have to spend on you know whatever else is needed (Beatrice, 105).

As discussed above in section 6.1.2.1 on roles and responsibilities, the CC's ceding leadership to the RCs was demanded by the NCI representatives. This request from NCI was not transparent to the RCs, so they saw the CC's move of placing more responsibility on the RCs as an indication of lack of knowledge about the data and desire to do less work.

Adding to the difficulties of negotiating data for submission to the repository was the CC's acknowledged lack of expertise in cervical cancer. Adam, the SN CC PI noted that his knowledge of cervical cancer was not as advanced as Charlie's knowledge of breast cancer.

> I think, as a biostatistician you see, the more you know about the subject area the better because you can communicate it to them. You can get their respect. It's as if okay, you know what they are thinking. But biostatisticians have an opportunity to work on any cancer. So then it's a difficulty to know a lot in one specific cancer unless you are doing it for a long time. So for me, that's one of my challenges on cervical cancer, because I never worked on cervical cancer. And [Charlie], you see he has worked on the breast cancer for 20 years (Adam, 795).

As of the end of my observations for this study, the breast working group had made substantially more progress than the cervical and colorectal groups. This progress was due, in part, to Charlie's knowledge of breast cancer, but this group also had another major advantage in that two of the three RCs also participate in another consortium on breast cancer screening. Charlie was involved with that consortium for many years, as were some of the RC PIs in the breast working group. As such, those participants were accustomed to the demands for data sharing in a screening-related consortium and, in fact, had already developed many common data elements for that project (Edith, 317).

Drawing on their experience as the BN CC, the SN CC came to the project with expertise not only in cancer data but in how to lead conversations with RC PIs that lead to decisions on data points required for specific projects, which, in turn, leads to high-quality data for reliable and trustworthy statistical analyses. However, the SN CC struggled to translate that data work expertise into scientific progress for the SN in its first year. Despite their attempts to lead the RC PIs to decisions on common data elements, the CC was stymied by their struggles with the FA representatives and RC PIs about the CC's role in the SN. These conflicts exacerbated apparent gaps in the CC's knowledge of both health plan data and cervical cancer, neither of which appear to have been particularly glaring, despite Beatrice's assessment. Given the diversity of data across the RCs, which ranged from university-based research hospitals to rural clinics in the southwest, it is unreasonable to expect the CC to be an expert in the data available in all those settings. And, while Adam's experience in cervical cancer was limited, he still brought more than 25 years of experience in biostatistical analyses on cancer data to the table.

In this example, we see how the different types of work affect one another. The funding cuts (structural work) affected how the RC PIs interpreted their role in the SN, leaving Beatrice feeling as though her RC was being asked to do more than they had committed to, with fewer funds. The conflicts over leadership and roles of the CC, FA and RCs (collaborative work) kept the CC from fully applying its expertise in leading the development of common data elements (data work) by forcing them to reduce their role in the SN leadership. The increased time the RCs were expected to spend on the data work of the data

repository, combined with the funding cuts, also reduced the amount of time they could spend on their local scientific work.

6.2 Chapter Summary

The descriptions of the CC activities in the three areas of network-level work presented in this chapter begin to illustrate what facilitation of collaborative research entails. In the BN, focus on supporting the clinical validation studies has led to the development of systems and processes iteratively improved over their years of experience in coordinating such studies. The end result of the CC's work in the areas of collaboration work, operational work and data work is that the RC PIs spend less time organizing themselves and their work and more time immersed in the scientific work that results in the high-quality data required to validate biomarkers successfully. On the other hand, the SN has struggled to take advantage of those same systems and processes to support the SN's progress toward the development of a data repository. This inability to translate its experience to a new project setting has resulted in slower progress than most SN participants had hoped for.

The contrast between these two CCENs highlights the contribution of the CC to the project's achievement of its scientific goals. When the BN is allowed to utilize its skills and knowledge, as when they ranked and prioritized the BN projects or when they lead discussions of common data elements, the project moves more smoothly toward valid scientific outcomes and allows the RC PIs to expend more effort on their science and less on administrative duties. When the SN was not allowed to utilize its skills and knowledge, when the FA representatives specifically asked them not to apply their experience and expertise to the scientific and organizational challenges at hand, the entire CCEN suffered. Not only did the RC PIs have to spend more time on discussions about how to organize themselves and administrative tasks, but the scientific progress is stifled.

Several of the examples of CC work practices given above also provide evidence of the interwoven nature of the work done by the CC. In the discussion in section 6.1.1.1 on BN collaboration work practices, we

see how the structural work done by the FA representatives to create a structure in which collaboration is both required and encouraged by NCI has created a network in which collaboration is also expected by the RC PIs of one another. In creating such a structure, the RC PIs appear to expend less effort thinking about how to work together, as evidenced by the lack of discussion in observations or interviews about doing collaborative work, freeing up time to spend on their data and local scientific work. In turn, the success of the data and scientific work reinforces the message that collaboration leads to scientific progress.

In the SN, Beatrice's statement about the CC's lack of understanding of health plan data reflects the intersection of the CC's data work practices and SN collaborative work practices. The FA representative's request for the CC to allow the RCs to take a greater role in the discussions about data for the repository impacted the work of selecting those data points, leaving Beatrice resentful of the additional effort required from her group and blaming the CC's ignorance for this additional effort. Additionally, the RCs have less time available to them to focus on their local scientific work.

In order to facilitate the work of the CCEN, CC staff and PIs apply the knowledge and experience they have developed over their collective decades of experience managing and supporting collaborative, multi-institutional research projects. They do so using the systems and processes they have developed iteratively over time to address the challenges of networked science. The point of this facilitation is to make the work of the CCEN go more smoothly and to generate high-quality data. When the CC is allowed to play this distinct role as a facilitator, as in the BN, the network-level work of the CCEN moves toward the achievement of its scientific goals with little resistance. However, when the role of the CC is limited, as in the SN, weaknesses and conflicts in one area of work spill over into other areas and the SN is not in a position to counteract this negative cascade. This effect is discussed in greater detail in the following chapter.

Chapter 7: Conclusion and Implications

This research project encompassed two main goals: exploring the functions of a CC in a CCEN, and understanding how the work of the CC facilitates the CCEN's progress toward its scientific goals. These goals have been addressed in this dissertation by answering the following research questions:

- How do Coordinating Centers facilitate the network-level work of Coordinating Center-Enabled Networks (CCENs)?
- What are the distinguishing characteristics of a CCEN?
- What work do each of the CCEN components perform?
- What does facilitation mean in a CCEN?

In this chapter, I summarize the findings that have been presented in answer to those questions, discuss implications of these findings for both CCEN practitioners and the field of Computer Supported Cooperative Work (CSCW), then present directions for future research.

7.1 Overview of Findings

Coordinating Center Enabled Networks (CCENs) are research networks comprised of investigators from Research Centers (RCs), representatives of a Funding Agency (FA), and the staff and PIs of a Coordinating Center (CC), all of whom are focused on achieving the overarching scientific goals of a collaborative research project, goals that can only be accomplished within a network structure. One of the defining characteristics of the CCEN is its inclusion of a CC to facilitate the group's achievement of its scientific objectives. These three characteristics – a *scientific objective* that can only be achieved using a *network of scientists* and a *CC tasked with facilitating the group's progress* – when combined together for the unique CCEN organizational structure.

The work of a CCEN is complex and requires a wide variety of work practices to complete its objectives. A CCEN engages in five types of work in pursuit of its scientific goals:

1. *Structural Work Practices (SWPs)*
   Structural work practices are those activities that shape the rules of the project and dictate the organizational structure the CCEN should take once funded and instantiated. Because most of this

work is performed by the FA and does not include interaction with other CCEN entities, the majority of the structural work of the CCEN is outside the scope of this study. Some instances of structural work involved interaction between the FA and RCs or CC in response to budget cuts or new funding cycles.

2. *Collaboration Work Practices (CWPs)*
   The collaboration work practices are the work of negotiating and deciding how to work together as a network, as well as the work of participating in those negotiations and decisions, all within the organizational structure set up by the structural work practices discussed above.

3. *Operational Work Practices (OWPs)*
   The operational work practices are the administrative and technological tasks done in support of the other types of work. Their aim is to help the group's diverse and varying tasks function together as a whole, as in when the CC organizes conference calls so the group can get together and discuss how to collect data for a study. These activities are primarily logistical or technical in nature and, in general, require little scientific knowledge to complete them. This work is primarily initiated by the CC; however, the CCEN members also must participate in this kind of work. For example, RCs must respond to CC requests for IRB documentation or requests for scheduling information.

4. *Data Work Practices (DWPs)*
   Data Work Practices are those activities whose focus is the production (i.e., the RCs generating data via lab work or extracting data from local databases) and consumption (i.e., the receipt of data for statistical analyses) of the highest possible quality data. The DWPs require a degree of scientific knowledge and expertise.

5. *Local Scientific Work Practices (LSWPs)*
   The local scientific work of the CCEN encompasses those activities done at the local organizations (i.e., at the individual RCs or CC) whose aim is to achieve the individual scientific objectives a participant proposed in his or her grant application. The LSWPs are scientific activities that happen in the CCEN that do not require interaction with other CCEN entities and are not guided by the collaborative protocols developed by the CCEN members. In other words, these activities are done independently by an RC or the CC; no LSWPs were observed or attributed to the FA.

All three types of CCEN entities – the CC, RCs and FA – engage in some instances of each of these types of work at some point in the CCEN, with the exception of the FA's lack of local scientific work. In practice, these types of work are interwoven and interdependent; none exists in a vacuum. The structural work sets the rules and lays out the scientific objectives within which the rest of the work is performed. The collaborative work culture that grows out of that structure largely determines how the work of the CCEN will get done by determining which work gets prioritized and what roles participants play. The

operational work practices support the data work, which is driven by the scientific objectives laid out in the structural work, as is the local scientific work. When one of these areas of work is not done in a way that is appropriate for the group's needs, that weakness can spill over into other areas, exacerbating or creating weaknesses there. On the other hand, an area of strength also can spill over, bringing stability to places in the project that are less stable.

Three of these types of work – the collaboration work practices, the operational work practices and the data work practices – involve all three of the types of entities working together at the same time and are called "network-level" work. This network-level work stands in opposition to the "local" work, in which an entity may engage individually or with one other type of entity. The definition of a CCEN notes that its aim is a scientific objective that can only be achieved within a network structure. Thus, this network-level work that engages the group working together is the real heart of the CCEN's work. This focus on the network-level work provides an analytical lens through which to view the CC's work more clearly and assess its impact on the work of the CCEN.

The central question that has guided this dissertation research is: How does the CC facilitate the network-level work of the CCEN? We find the answer to this question in the CC's application of their experience and expertise to the challenges of collaborative research. The CC plays a distinct role in the CCEN, facilitating the network-level work of the project, with the aim of making the work of the CCEN go more smoothly and generating high-quality data. This facilitation involves the application of the CC's collective and individual knowledge and experience, amassed over years of experience managing and supporting collaborative, multi-institutional research projects. In service of this goal, the CC has developed systems and processes to address the challenges of networked science. When the CC is allowed to play this distinct role as a facilitator, as in the BN, the network-level work of the CCEN moves toward the achievement of its scientific goals with little resistance. However, when the role of the CC is limited, as in the SN, weaknesses and conflicts in one area of work spill over into other areas and the SN is not in a position to counteract these negative forces.

7.2 A Model of Collaborative Work

The five categories of CCEN work described above approximate the boundaries between types of work. By dividing the work done in a CCEN into these five categories, it is possible to get a clearer picture of a very complex work environment and see why it is not easy to develop systems to support such interdependent work practices, where tasks affect one another and do not always have clear boundaries between them. Such a model of collaborative work begins to tease apart the types of work and can help us to anticipate and plan for these interdependencies. While I have divided the work into neat categories, not all the work done falls neatly into just one of these boxes. In fact, the major CCEN activities – the creation of a data repository in the SN and the coordination of clinical validation studies in the BN – explicitly involve at least two types of work and implicitly rely on the work from other categories. This is true of the two activities described in the previous chapter – the creation of the data repository for the SN and the coordination of clinical validation studies for the BN.

In the creation of the SN data repository, the ultimate scientific goal is the creation of a data repository that describes the entire screening process. In order to achieve that goal, the CCEN members engage in work that determines what kinds of data to send to the repository at the CC and do the work required to extract that data and transfer it to the CC. The CC must keep a high-level perspective of these discussions to ensure consistent, high-quality data, while also engaging in coordinative work that keeps the discussions flowing and moving forward. Everyone in the CCEN must feel like their participation is important to the eventual success of the CCEN and engage fully in the work, including participating in meetings and serving on the Working Groups making the decisions. All of this work takes place within the structure set forth by NCI. When this structure is not adequately matched to the demands of the work that must take place in the CCEN, the work begins to break down. Ordinary challenges of collaborative work, such as creating an agenda for an investigator meeting or deciding on data points for inclusion in a data repository, are exacerbated and turn into problems that take time and effort to solve, time and effort that could be better spent on advancing the science.

For the BN, coordinating a clinical validation study likewise requires the application of all five types of work practices. The ultimate scientific goal is to receive FDA approval for clinical use of the biomarker under study, demonstrating its clinical impact. To achieve that goal, the science behind the biomarker must be rigorous, and the data that are collected during the validation study must be clean and of the highest quality. The number of moving parts that come together to create a working multi-site study are large, and include a strong, clear protocol with appropriate eligibility criteria spelled out precisely, easy-to-use data entry forms that collect the correct data, biosample collection instructions that are easy to understand and proper transmission of data. Participating sites must understand the importance of their contribution to the study and the importance of following the protocol precisely. Member PIs must participate in the groups leading these projects. And NCI must create incentive structures that encourage and reward collaboration as opposed to individualism. When this structure is in place, participants understand what is expected of them and what will happen if they don't comply with the requirements. This allows investigators to spend more time on their science and less trying to figure out their role.

From these stories of two CCENs, we can begin to see how these types of work are not only intertwined but interdependent. In Figure 7.1, I present a model of collaborative work based on the five types of work discussed in this dissertation. The arrows moving from the bottom types of work to the upper types indicate the influence of that layer's objectives. The scientific objectives, most explicitly instantiated in the local scientific and data work practices of the CCEN, influence all of work that takes place in the layers above. Conversely, the work that takes place in the upper layers provides the structure in which the work of the lower layers takes place. In other words, as we proceed through the layers, the work on the bottom layers could not take place without the work of the top layers. We can also venture that if the work on the top layers is not done well – meaning, the structure is not sound – the work that takes place further down the stack will not go well or will go more slowly. It is important to note that the influence and support do not necessarily proceed linearly through the various layers, but, rather, move in the general direction of the arrows. In other words, the data work could influence the collaboration work

directly, without needing to have an intermediate influence on the operational work. Likewise, the structural work can support the local scientific work directly.

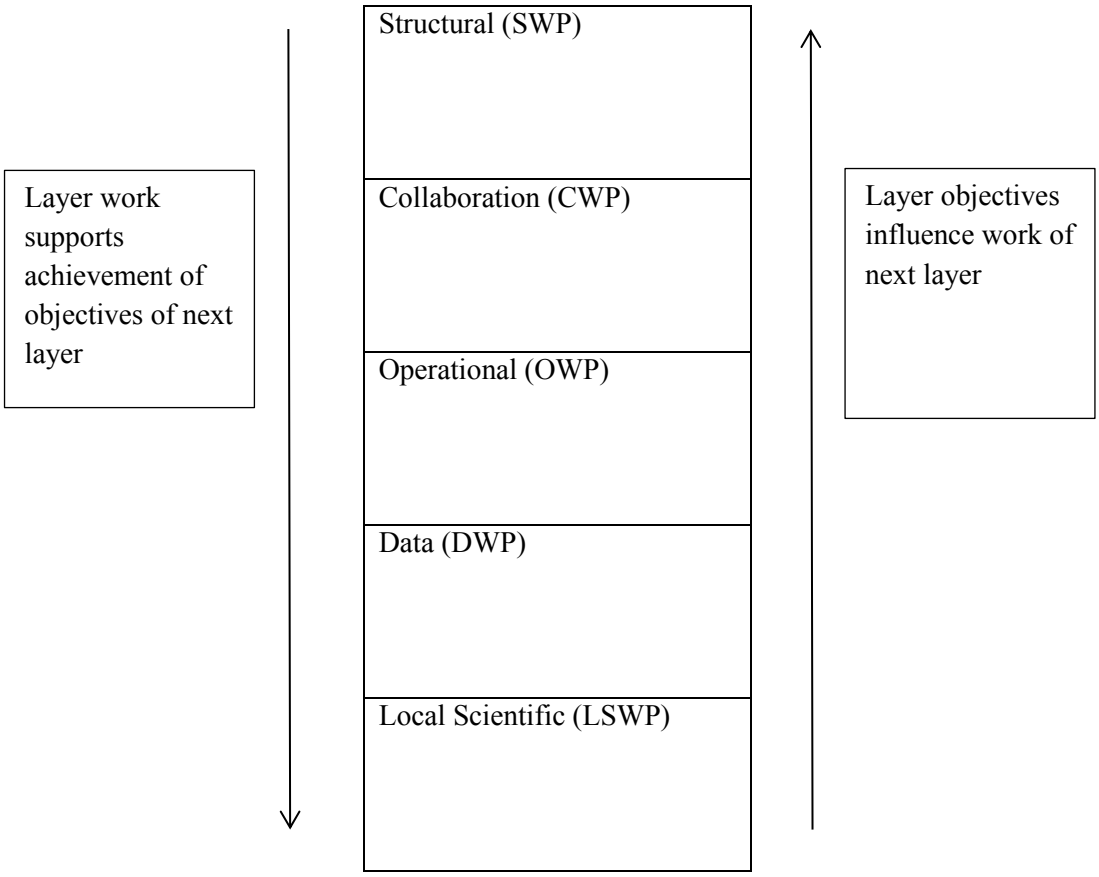

Figure 7.1: Work Practices of a Coordinating Center Enabled Network

To illustrate this interdependency more clearly, Figures 7.2 and 7.3 demonstrate the ways in which it played out in the BN and SN. In the SN, clarity of the roles in the RFA, the funding mechanism and the evaluation criteria created a culture that focused on collaboration. This led to open sharing of samples and data, which resulted in more time spent on the science and greater scientific progress. This, in turn, fed back into the culture of collaboration. However, in the BN, the combination of a lack of clarity of the role of the CC in the RFA and evaluation criteria combined with budget cuts to the RCs led to misunderstandings among the CCs, FA and RCs. This, in turn, led to difficulties in sharing data and resources, the outcome of which was less time being spent on the science and disappointing scientific progress.

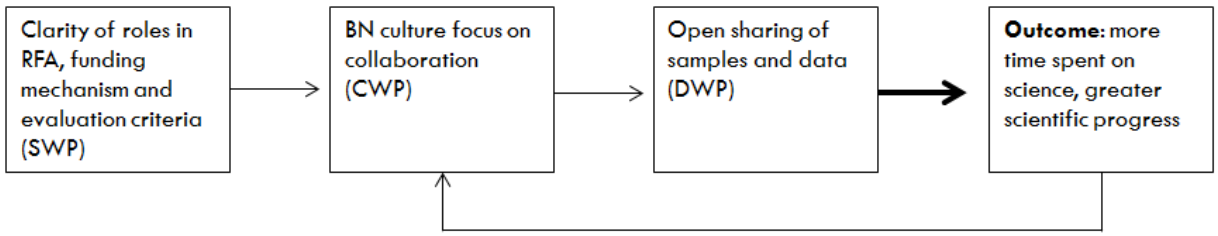

Figure 7.2: Interdependency of work practices in the Biomarker Network

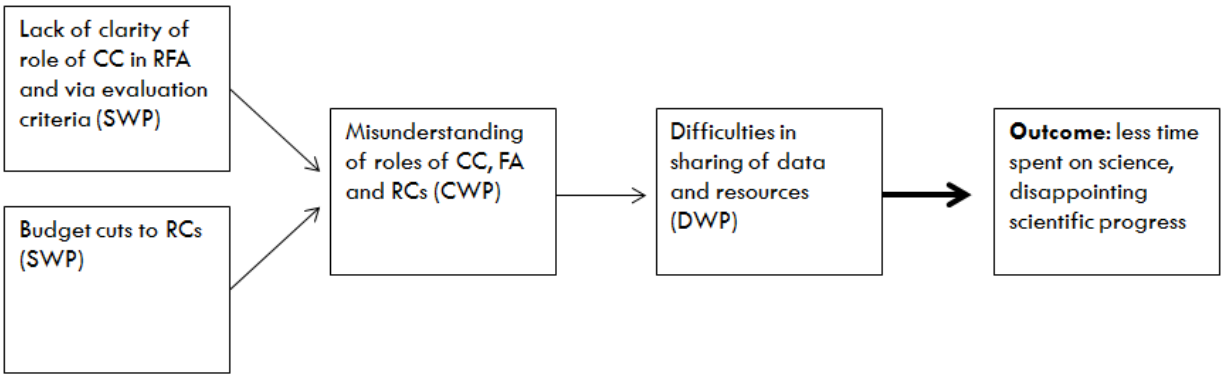

Figure 7.3: Interdependency of work practices in the Screening Network

7.3 Implications for CCEN Practitioners

The findings presented here present opportunities for CCEN practitioners, including CC, FA and RC leaders, to improve the performance not just of CCENs, but of all networked science projects. The facilitation work described in this dissertation is not a new type of work; in fact, it is work that all networked science projects must do in order to accomplish their goals. Studying the CCEN allows us to see this facilitation work made explicit by its assignment to an independent entity, the CC. As such, any multi-site project should be able to take advantage of the lessons learned here.

With that in mind, I offer the following recommendations for improving the professional practice of collaborative, networked science:

7.3.1 Improved definition of facilitation and collaborative activities

While this dissertation advances an initial definition of what it means to facilitate collaborative science, it is but a start. The field of networked science is currently missing a generally accepted definition of

facilitation and necessary collaborative activities in a networked project, despite the fact that most projects call for some type of facilitation. Rectifying this deficiency could make scientific progress easier for multi-institutional collaborative projects by making the role of the CC clear for all involved. This need for clarity extends to all participants in the project, especially around the question of what types of collaboration are expected. The BN's funding and evaluation structure didn't just encourage collaboration, it required it. By laying out clear expectations for the collaborative RC PIs needed to do, the BN leadership simplified the task of participating in the BN, removing one more obstacle to the accomplishment of their scientific goals. These clear expectations also ensured less friction between participants, making the collaboration work smoother.

7.3.2 Increased use of the CCEN structure, with improved tools and procedures

It is clear that both CCs described in this dissertation had a positive impact on the science of the CCEN of which they were a part. Even with the SN's struggles to define everyone's role in the collaboration, the group was still making use of the CC's expertise in data and operational work to its advantage. While CCs are not uncommon in the world of collaborative cancer epidemiology, it is my belief that they should become even more widely utilized in projects involving more than a handful of collaborating institutions. But the CCEN structure would greatly benefit from the development of certain tools and procedures by the Funding Agencies. Data presented in this dissertation show that CCs spend substantial amounts of time figuring out how to work together, often repeating certain processes for each new collaborative project. Standardizing some of those processes could save CCENs time and effort as they begin working together. For example, standard governance documents could be made available to the CCEN, lessening the amount of time it takes them to decide upon their operating procedures. Project start dates could include a required face-to-face meeting, ensuring that the network begins to form an identity and trust. Standard CCEN publication policies could be developed. In short, the FA could begin to experiment with ways to lessen the amount of time and effort required by a CCEN to get up and running.

7.3.3 Increased support for CCs

The trick to achieving increased use of the CCEN structure, of course, is to identify and develop CCs with the type of experience held by the SFT group at FHCRC. As shown in this dissertation, it was the CC's application of their expertise to the challenges of collaborative research that allowed them to have such a strong impact on the work of the CCEN. However, that kind of experience and expertise take time to develop. Funding agencies should find a way to invest in such groups that allows organizations like the SFT to attract and retain talent and institutional knowledge. Such groups require stable funding, including incentives to document their policies and procedures and requirements to evaluate their efforts to facilitate the work of the CCEN.

7.4 Implications for Computer Supported Cooperative Work (CSCW)

This research also has implications for CSCW, a field concerned with how groups of people work together using technology, specifically in the areas of what solutions to the social-technical gap might look like and what facilitation really means on a day-to-day basis.

7.4.1 Moving from the Social-Technical Gap to the Social-Social Gap

Ackerman (2000) describes the social-technical gap as the space between what we know we must support socially but are unable to support technically (Ackerman, 2000). In Chapter 2, I posited the existence, too, of a social-social gap, the space between the social support workers need in order to meet the challenges of collaborative science and the actual social systems available to provide that support. In other words, perhaps the persistent social-technical gap indicates that it is not enough simply to design technologies to support social processes, but that we must also design social processes and organizational structures *around* those technologies.

The CCs described in this dissertation provide one example of a group that has managed to bridge both the social-technical and social-social gaps inherent in supporting collaborative science. The BN and SN CCs developed technologies to support such common scientific tasks as data entry at remote clinical sites

and uploading data to a common database. However, these technologies did not function in a vacuum, but, rather, required the development of social processes, norms and structures *around* the technologies in order to be successful. The conversations that the CC staff led were crucial to the operations of those technologies, both in their development and in their implantation and use. While CSCW has traditionally focused on technical solutions (the "computer-" in computer-supported cooperative work), this dissertation provides evidence that the identification of social solutions around the technology is equally important to solving the problems of collaborative work.

7.4.2 Organizational Concerns

While previous research has identified issues in building cyberinfrastructures (CIs), few of the concerns mentioned, such as tensions between short-term and long-term objectives (Ribes & Finholt, 2007) and challenges in identification of relevant team members (Lee et al., 2006; Bietz et al., 2010), were apparent in my observations and interviews. I hypothesize that this difference may be attributable to two potential causes. First, the project goals of the groups detailed in those articles were very different than the project goals of the CCs. CIs were generally focused on building large infrastructures with specific aims of building novel, long-term, massive databases and tools, while the BN and SN were focused on achieving the short-term objectives of their five-year projects. This is not to say that the BN and SN weren't concerned with being funded again at the end of those five years or developing their science, but, rather, that those outcomes would be a result of completing the project's scientific objectives instead of a goal in and of themselves. Second, the existence of a central coordination team may have simply eliminated the organizational issues noted by Lee et al. (2006) and Bietz et al. (2010). Individual team members did not have to be concerned with figuring out their counterparts or where they stood in a project, because the CC managed those issues when the work crossed organizational boundaries. In either case, this is clearly an area for future study. As CSCW researchers seek to build tools to support distributed collaboration, understanding how organizational concerns differ, as well as where they converge, between relatively short-term projects and long-term infrastructure projects is crucial.

7.4.3 Expanding Articulation Work

In CSCW literature, one theory for understanding coordination in distributed work is Articulation Work (AW), which focuses on the work of making work go well. However, AW alone is not able to explain fully the types of work being done in the BN and SN in support of the CCEN's scientific work. Specifically, AW requires more development in the areas of (1) the effect of structural and organizational forces, (2) how AW fits into the high-level objectives of the work being done on the project, and (3) the relationship between the AW and non-AW work being done.

First, the key articles on AW are limited in their discussion of structural and organizational forces and their effect on the AW being done. How exactly these forces affect the work is not discussed in detail, yet this is a critical point we need to understand. In general, workers cannot simply make up rules or structures in which to operate and engage in negotiations about work. As a first step toward expanding this area of AW, this dissertation has described how the structural work done in a CCEN affects the other types of work being done and, really, the outcomes of the project. To recap briefly, that structural work creates the rules under which the collaboration takes place, defining roles and responsibilities for participants and laying out the scientific objectives of the project. Most of that work is done prior to the instantiation of the CCEN and appears fully-formed with the RFA. However, if that structural work is not done well (i.e., roles are not described thoroughly enough or funding is not sufficient for the work being requested) or if changes are required during the project that negatively impact how collaborators do their work, that structural work becomes very visible and can lead to problems as the work of the CCEN progresses.

Second, AW focuses little on the overarching, high-level objectives of the work, which in large part drive the articulation work taking place in a CCEN. While some AW may occur in most distributed scientific collaborations, such as scheduling conference calls or distributing meeting minutes, much of the AW taking place in a CCEN is a result of the project's scientific objectives. In this dissertation, I have

described how the CC staff have developed processes to deal with the specific challenges of their projects, such as developing work flows to craft eligibility flow charts for clinical validation studies or leading conversations about indication of tests (screening vs. diagnostic) in the development of the SN data repository. These are different kinds of AW done in service of achieving different scientific objectives, not general AW done in service of supporting scientific collaboration. These details can matter, and it is impossible to understand the full scope of the AW that takes place in a CCEN without understanding the nature of the project in which it is being done. As more CSCW research is done in this area, it would be interesting to explore whether this objective-specific AW represents a separate category of AW, much like Gerson's division of AW into local AW and metawork.

Third, the relationship between AW and the non-AW work in a collaborative project is unclear in the existing AW theory. What non-AW takes place in collaborative work and how does it interact with the AW being done? Again using our CCENs as an example, how do the data work practices affect or interact with the operational or domain work practices? In this dissertation (Section 7.2), I have put forth one theory of the effects of these interactions, but it is unclear if this model will apply to other types of collaborative work. Understanding these interactions is absolutely critical to our work in CSCW, because developing systems to support one kind of work within a project must take into account how that work affects other work being done on the project or within the organization.

7.4.4 Articulation Work vs. Typology of CCEN Work Practices

In Chapter 2, I explored Gerson's distinction between metawork and local articulation work. To reiterate, metawork is the work of bringing together different types of work, while local articulation work is concerned with bringing together what is needed in a "particular time and place" or "local situation" (Gerson, 2008). In a virtual organization, the term "local" can be misleading, as it doesn't necessarily just refer to geographic place but could also refer to a part of a project. In our CCENs, scheduling a conference call can be considered local under Gerson's definition, despite the fact that the work crosses

many locations, because it deals with a particular situation. Given these definitions, the collaboration work practices described in the typology of CCEN work practices look similar to metawork, while the operational work practices correspond to local articulation work. Table 7.1 compares these two theories.

However, this leaves out two remaining types of work practices discussed in the dissertation which also fall into the category of "the work of making work go well;" namely, the structural and data work practices. As discussed above, the structural conditions that affect the collaborative work are not fully explored in the AW theory, making it difficult to compare them to the category of structural work practices, beyond agreeing that they are crucial to the rest of the work being done. It is this structural work that sets the rules and boundaries for how work can proceed in the CCEN. Without understanding this structural work and its effects, it becomes difficult to understand the full extent of the articulation work that takes place. The limits and possibilities set by the structural work have a lasting impact on the work of the project.

The data work does not appear to fit neatly into any of the categories of the AW theory but, rather, seems to be a combination of local articulation work, metawork and domain work focused on the end goal of high quality data. While it would be possible to distribute the tasks currently categorized as data work into those three AW categories, separating them out deprives them of their context and restricts our ability to see the interwoven nature of the work being done around data. This full understanding of the data work in a CCEN is critical for the practice of CSCW, as a system that does not support all aspects of the data work will fail to support the CC's highly developed work practices. On the other hand, viewing the data work practices as one category of work gives us a more holistic view of the CC's work and hews more closely to the view the CC has of its own work.

Finally, the local science practices in the CCEN work practice typology are also left out of the AW framework, making comparisons difficult there, as well. Yet it is clear from the work presented in this dissertation that the addition of these three types of work practices (structural, data and local scientific)

gives us a much fuller, richer understanding of the totality of work being done in a collaborative project. The relationships between them have been described in this dissertation, but this work is just a first step toward understanding how collaboration happens.

CCEN Typology of Work Category	**Articulation Work Concept**
Structural Work	Structural and organizational conditions
Collaboration Work	Metawork
Operational Work	Local articulation work
Data Work	Combination of local articulation work, metawork and domain work
Local Scientific Work	N/A

Table 7.1: Comparison of CCEN Typology of Work categories with Articulation Work concepts

7.4.5 Coordinating Centers

Section 2.3 describes the limited amount of previous work done around coordinating centers, the work they do and its impact on the scientific objectives of their projects. While the majority of those studies focused on coordinating centers for clinical trials, as opposed to observational studies, many of their assertions about their work are reflected in the experiences of the BN and SN CCs. For example, the emphasis that several of the CC articles put on the CC's role of assuring the validity of the scientific outcomes of the projects was clearly seen in my data, as was the importance of the CC's responsibility for monitoring implementation and adherence to the protocol. The categories of work discussed in section 2.3.1 included (1) statistical coordination and management, (2) study coordination, and (3) administrative and secretarial support. These three categories of work differ somewhat from the typology of CCEN work practices. While the tasks included in statistical coordination and management all fall into the typology category of data work practices, those included in study coordination are scattered among the three types of network-level work practices (collaboration, operation and data work practices). To reiterate, study coordination included protocol and form development, monitoring adherence to the protocol or performance monitoring, computer system development, training of staff, documentation and archiving of study information, communications, adherence to institutional policies, reporting, allocation of CC resources and manuscript preparation. These tasks fall into the category of collaboration work (e.g.,

allocation of resources), operational work (e.g., communications, reporting) or data work (e.g., protocol and form development). There was little mention in the CC literature about the local scientific work being done in these collaborative projects, or of the structural work being done. Yet, a full picture of the work done at the CC must include those types of work, as well, in order to fully understand how the work of a collaborative project gets done.

The Coordinating Center Models Project (CCMP), which took place in the mid-1970s, failed to identify a common set of activities common across all the CCs that it studied. In Section 2.3.1, I posited that perhaps this was simply a reflection of the novelty of the CC as an organizational structure rather than a true impossibility of developing a model of CC activities. While the initial CCMP project did not find common activities, this dissertation has found both common activities in two CCs (the BN and SN CCs), as well as common activities between these two CCs and CCs described in literature published after the CCMP project. This may represent the development of *de facto* standards of coordination in distributed scientific collaboration, though substantially more research is needed to understand if this is true across types of coordinating centers and types of scientific projects.

7.5 Future Research

This research project is, to the best of my knowledge, the first to investigate the use of CCs in collaborative cancer epidemiology projects, and the first to define and describe the type of virtual organization here called Coordinating Center Enabled Networks, or CCENs. As such, I have just begun to scratch the surface of this complex type of organization. I have developed the following questions for future research:

7.5.1 What other types of Coordinating Centers and networks with CCs exist?

This research looked at only two CCs at one institution, but there are more that exist outside of FHCRC and outside of cancer epidemiology. Future research will work to identify as many CCs as possible at other institutions and in other disciplines. Discovering both the breadth and depth of CCs will help

develop the definition of what a CC is and does, as well as increase our understanding of the impact of both CCs themselves and coordination in general. It will also allow the description of the different kinds of networks that include CCs, also expanding the field's understanding of how collaborative science works.

7.5.2 What do the Funding Agency representatives and the Research Center PIs experience in a CCEN?

One of the limitations of this research was its viewpoint of the CC and limited documentation of the experience of the other two types of participants. While one FA representative and three RC PIs from each project were interviewed, these hardly represent the diversity of viewpoints of the large numbers of participants in each CCEN. Additionally, I had virtually no view into the local work of the FA or RCs, limiting my understanding of the work they do there as they interact with the CC and other CCEN participants.

7.5.3 Do CCENs exist in other disciplines, outside of cancer epidemiology? If so, how do they operate?

Another limitation to this research was its focus on two CCENs operating in the realm of cancer epidemiology. We need to understand if this organizational form exists outside of this discipline and, if so, compare how those CCENs work to those documented here.

7.5.4 What is the long-term effect of involvement with a CC of data experts?

As discussed in this dissertation, RC PIs who had worked with those in the CC who are experts in data and the processes of data work, including both the statisticians and other CC staff with expertise in developing processes that led to high-quality data, felt that these interactions had improved their own local practice of science by exposing them to new ways of doing science. First, is there a way to replicate that experience for those not involved in CCENs? And, second, what is the long-term effect of being

involved with the CC's experts? Do the RC PIs continue to utilize the tactics they learned from the CC or do they revert back to their local ways?

7.6 Summary

As science tackles the most pressing questions of our time, including improving global health and developing energy independence, interdisciplinary team science will continue to increase as the method of choice. Coordinating Centers have the potential to alleviate many of the difficulties of collaborative team science by transferring the administrative burden of collaboration from the PIs to a group of individuals with experience facilitating collaborative research. As has been shown in this dissertation, allowing a CC to facilitate this research by applying its extensive experience and expertise to the challenges of collaborative research, PIs are able to spend more time on their science and make greater progress toward the achievement of the group's scientific objectives.

References

Ackerman, M. S. (2000). The intellectual challenge of CSCW: The gap between social requirements and technical feasibility. *Human–Computer Interaction, 15*(2), 179-203.

Bangdiwala, S. I., de Paula, C. S., Ramiro, L. S., & Muñoz, S. R. (2003). Coordination of international multicenter studies: governance and administrative structure. *Salud pública de México, 45*(1).

Berge, K. C. (1980). Perceptions of the coordinating center: as viewed by an advisory board. *Controlled clinical trials, 1*(2), 143-146.

Bietz, M., Baumer, E., & Lee, C. (2010). Synergizing in Cyberinfrastructure Development. *Computer Supported Cooperative Work (CSCW), 19*(3), 245-281. doi: 10.1007/s10606-010-9114-y

Blumenstein, B. A., James, K. E., Lind, B. K., & Mitchell, H. E. (1995). Functions and organization of coordinating centers for multicenter studies. *Controlled clinical trials, 16*(2).

Charmaz, K. (2009). *Constructing grounded theory : a practical guide through qualitative analysis*. Los Angeles; London: SAGE.

Collins, J. F., Martin, S., Kent, E., Liuni, C., Garg, R., Egan, D., & on behalf of the, D. I. G. I. (2003). The use of regional coordinating centers in large clinical trials: the DIG trial. *Controlled clinical trials, 24*(6).

Corbin, J. M., & Strauss, A. L. (1993). The Articulation of Work through Interaction. *The Sociological Quarterly, 34*(1), 71-83.

Cummings, J., & Kiesler, S. (2007). Coordination costs and project outcomes in multi-university collaborations. *Research Policy, 36*, 1620-1634. doi: 10.1016/j.respol.2007.09.001

Cummings, J., & US National Science Foundation. (2008). *Beyond being there: A blueprint for advancing the design, development, and evaluation of virtual organizations*: NSF.

Cummings, J. N., & Kiesler, S. (2005). Collaborative Research across Disciplinary and Organizational Boundaries. *Social Studies of Science, 35*(5), 703-722.

Curb, J. D., Ford, C., Hawkins, C. M., Smith, E. O., Zimbaldi, N., Carter, B., & Cooper, C. (1983). A coordinating center in a clinical trial: the Hypertension Detection and Followup Program. *Controlled clinical trials, 4*(3), 171-186.

Emerson, R. M., Fretz, R. I., & Shaw, L. L. (1995). *Writing ethnographic fieldnotes*. Chicago: University of Chicago Press.

Falk-Krzesinski, H. J., Contractor, N., Fiore, S. M., Hall, K. L., Kane, C., Keyton, J., . . . Trochim, W. (2011). Mapping a research agenda for the science of team science. *Res. Eval. Research Evaluation, 20*(2), 145-158.

Fred Hutchinson Cancer Research Center, "Campus Sustainability" Brochure. fhcrc.org. Accessed March 18, 2013. https://www.fhcrc.org/content/dam/public/About%20Us/campus_sustainability.pdf.

Gerson, E. M. (2008). Reach, bracket, and the limits of rationalized coordination: Some challenges for CSCW. In C. H. Mark S. Ackerman, Tomas Erickson, and Wendy A. Kellogg (Ed.), *Resources, Co-evolution, and Artifacts: Theory in CSCW* (pp. 193-220). New York: Springer-Verlag.

Greene, S. M., Hart, G., & Wagner, E. H. (2005). Measuring and improving performance in multicenter research consortia. *Journal of the National Cancer Institute. Monographs*(35), 26-32.

Lachin, J. M. (1980). Perceptions of the coordinating center: foreword. *Controlled clinical trials, 1*(2), 125-126.

Lawrence, K. (2006). Walking the Tightrope: The Balancing Acts of a Large e-Research Project. *COMPUTER SUPPORTED COOPERATIVE WORK, 15*(4), 385-411.

Lee, C. P., Dourish, P., & Mark, G. (2006). The Human Infrastructure of Cyberinfrastructure. In Proc. CSCW 2006, ACM Press (2006), 483 - 492.

Margitic, S. E., Morgan, T. M., Sager, M. A., & Furberg, C. D. (1995). Lessons learned from a prospective meta-analysis. *Journal of the American Geriatrics Society, 43*(4), 435-439.

Meinert, C., Heinz, E., & Forman, S. (1983). Role and methods of the coordinating center. *Controlled Clinical Trials Controlled Clinical Trials, 4*(4), 355-375.

Meinert, C. L., & Coordinating Center Models Project. (1977). *Proceedings of the fourth annual meeting of personnel involved in coordinating collaborative clinical trials, Chapel Hill, N.C., 5/19-20/77*, Baltimore.

Mowery, R. L., & Williams, O. D. (1979). Aspects of clinic monitoring in large-scale multiclinic trials. *Clinical pharmacology and therapeutics, 25*(5), 717-719.

National Institutes of Health Research Portfolio Online Reporting Tools (RePORT). (2012). Retrieved May 29, 2012, from http://report.nih.gov/award/index.cfm?ot=&fy=2011&state=&ic=&fm=&orgid=#tab2

Overton, H. H. (1980). Perceptions of the coordinating center: as viewed by a clinic coordinator. *Controlled clinical trials, 1*(2), 133-136.

Pepe, M. S., Feng, Z., Janes, H., Bossuyt, P. M., & Potter, J. D. (2008). Pivotal Evaluation of the Accuracy of a Biomarker Used for Classification or Prediction: Standards for Study Design. *Journal of the National Cancer Institute, 100*(20), 1432-1438. doi: 10.1093/jnci/djn326

Prout, T. E. (1980). Perceptions of the coordinating center: as viewed by the principal investigators. *Controlled clinical trials, 1*(2), 127-131.

Ribes, D., & Finholt, T. A. (2007). *Tensions across the scales: planning infrastructure for the long-term.* Paper presented at the Proceedings of the 2007 international ACM conference on Supporting group work.

Rifkind, B. M. (1980). Perceptions of the coordinating center: as viewed by a project officer. *Controlled clinical trials, 1*(2), 137-141.

Rolland, B., Smith, B. R., & Potter, J. D. (2011). Coordinating centers in cancer epidemiology research: the Asia cohort consortium coordinating center. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology, 20*(10), 2115-2119.

Sedransk, N. (1980). Perceptions of the coordinating center: as viewed by a site visitor. *Controlled clinical trials, 1*(2), 147-152.

Seminara, D., Khoury, M. J., O'Brien, T. R., Manolio, T., Gwinn, M. L., Little, J., . . . Network of Investigator, N. (2007). The emergence of networks in human genome epidemiology: challenges and opportunities. *Epidemiology (Cambridge, Mass.), 18*(1), 1-8.

Simon, H. A. (1996). *The sciences of the artificial*. Cambridge, Mass.: MIT Press.

Strauss, A. (1988). The Articulation of Project Work: An Organizational Process. *The Sociological Quarterly, 29*(2), 163-178.

Symposium on Coordinating Clinical Trials. (1978). *Proceedings of the fifth annual Symposium on Coordinating Clinical Trials.* Paper presented at the Coordinating Center Models Project, Baltimore; Springfield, Va.

Weiss, R. S. (1994). *Learning from strangers : the art and method of qualitative interview studies*. New York; Toronto; New York: Free Press ; Maxwell Macmillan Canada ; Maxwell Macmillan International.

Winget, M., Kincaid, H., Lin, P., Li, L., Kelly, S., & Thornquist, M. (2005). A web-based system for managing and co-ordinating multiple multisite studies. *Clinical Trials, 2*(1), 42-49.

Wuchty, S., Jones, B. F., & Uzzi, B. (2007). The increasing dominance of teams in production of knowledge. *Science (New York, N.Y.), 316*(5827), 1036-1039.

Appendix A: Final Interview Protocol

Interview Questions for CC staff:

1. Tell me, at a high level, what the goals of the project [SN or BN] are and how the collaboration is structured. [Ask them to draw the org chart of the collaboration.]
2. What is the role of the CC within this collaboration?
3. What is your role within the CC? How much effort do you devote to this project?
4. What do you think NCI thinks the CC's role is? What do the [RCs] think the CC's role is?
5. Tell me about the cross-site projects the CC is involved in coordinating. What are they?
6. BN: 6(a) Walk me through the process of coordinating a new team project.
   6(b) Walk me through the process of coordinating a new reference set.
   6(c) Walk me through the process of coordinating a clinical validation study.
   SN: 6(d) Tell me about the process of coordinating the CDE creation.
   6(e) Tell me about the process of creating the SN governance documents.
7. Can you give me two examples of challenges you encountered when working on [BN team projects or reference sets or clinical validation studies/SN CDEs or governance documents].
8. How were these issues resolved? How did you respond to these challenges?
9. How have the sites responded to the CC's efforts to coordinate these cross-site projects?
10. How successful do you think the CC has been in coordinating these cross-site projects?
11. How should the CC define success for itself? How do you define success for [BN/SN]? Are they different?

Interview Questions for external collaborators:

1. Tell me, at a high level, what the goals of the project [SN or BN] are and how the collaboration is structured. [Ask them to draw the org chart of the collaboration.]
2. What is your role within the collaboration? How much effort do you devote to this project?
3. What is the role of the CC within this collaboration?
4. What do you think NCI thinks the CC's role is? What does the CC think your role is? What is the role of RCs?
5. Tell me about the cross-site projects the CC is involved in coordinating. What are they?
6. Tell me about your involvement in one of the following
   a. [BN] team project
   b. [BN] reference set
   c. [BN] clinical validation study
   d. [SN] CDE creation
   e. [SN] governance documents creation
7. Can you give me two examples of challenges you encountered when working on [BN team projects or reference sets or clinical validation studies/SN CDEs or governance documents].
8. How were these issues resolved? How did the CC respond to these challenges?
9. How successful do you think the CC has been in coordinating these cross-site projects?
10. How should the CC define success for itself? How do you define success for [SN/BN]? Are they different?

Appendix B: Codes Used in Data Analysis

1. *Juicy
2. barriers to collaboration
3. [BN] sample reference sets
4. [BN] study coordination (recruitment, etc) for biomarker validation studies and clinical trials
5. [BN] team projects
6. CDEs and data
7. collaborative work instance
8. confusion and misunderstandings
9. cross-CC knowledge sharing
10. expectations of roles in collaboration for CC, NCI or collaborators
11. funding or budget discussions or issues
12. governance and decision making issues
13. IRB
14. meeting leadership and arrangements
15. personnel issues
16. Q1:Background
17. Q12:Success Questions
18. Q2:Project Goals
19. Q2_2:Structure of Collaboration
20. Q3:Role of CC
21. Q4:P Role in CC and FTE
22. Q5:NCI on Role of CC
23. Q5_2:Members on Role of CC
24. Q6:Cross-Site Projects Being Coordinated
25. Q7:Coordination or Participation in Process for Cross-Site Projects
26. Q8:Examples of Challenges
27. [SN] mapping screening process
28. [SN] organ-specific groups
29. [SN] trans initiatives
30. statistical analyses
31. success
32. tools development and informatics
33. zorganizational responsibilities, leadership or tasks aka articulation work
34. zscientific responsibilities, leadership or tasks

Appendix C: Acronyms

BN:	Biomarker Network
CC:	Coordinating Center
CCEN:	Coordinating Center Enabled Network
CDE:	Common Data Element
CSCW:	Computer Supported Cooperative Work
FA:	Funding Agency
FHCRC:	Fred Hutchinson Cancer Research Center
NCI:	National Cancer Institute
NIH:	National Institutes of Health
PI:	Principal Investigator
RC:	Research Center
RFA:	Request for Applications
SN:	Screening Network

Curriculum Vitae

Betsy Rolland

Research Statement

As a researcher, my area of interest is the coordination and support of collaborative biomedical research, specifically in the realm of cancer epidemiology. The central question of my dissertation is how coordinating centers facilitate the network-level work of cancer epidemiology networks and to what effect. My related research has focused on the reuse of data in cancer epidemiology datasets by post-doctoral fellows, investigating the process post-docs go through in order to understand and use datasets collected by others. I am also interested in the use of information in collaborative research and have studied the changing role of librarians in biomedical research.

Education

2013 **PhD Candidate, Human Centered Design & Engineering** (College of Engineering)
University of Washington, Seattle
Adviser: Professors Charlotte P. Lee
Dissertation: "Greater than the Sum of its Parts:" Coordinating Centers as Facilitators of Network-Level Work in Cancer Epidemiology Coordinating Center Enabled Networks

2007 **Master of Library and Information Science**
University of Washington, Information School, Seattle, WA
Coursework included information behavior, programming for information systems, organization of resources, project management, biomedical and health informatics, usability testing and instructional and training strategies.

1993 **BA, Slavic Languages and Literature**
Northwestern University, Evanston, IL

Employment

5/07-present **Fred Hutchinson Cancer Research Center,** Seattle, WA
Principal Investigator (9/11-8/13)
Senior Project Manager, Asia Cohort Consortium Coordinating Center (2/08-present)
Information Analyst and Information Services Team Project Manager, Alliance Management Office (5/07-1/08)

9/05-6/07 **University of Washington,** Seattle, WA
Graduate Research Assistant to Director of Web Services, College of Arts and Sciences, Dean's Office (9/05-6/07)
Information Architect, Department of Medicinal Chemistry, Daggett Lab (10/06-3/07)

8/02-8/05 **Corel (formerly Jasc Software, Inc.),** Eden Prairie, MN
Automation Engineer (6/03-8/05)
Quality Assurance Engineer (8/02-6/03)

6/99-6/02 **Macromedia**, San Francisco, CA and Minneapolis, MN

Quality Assurance Engineer, White Box Testing Lead (12/01-6/02)
Build & Release/Installation Engineer, LikeMinds (6/99-12/01)

5/98-3/99* **Plageman, Lund & Miller LLP**, Oakland, CA
Office Manager

4/96-1/98* **Kaiser, Inc.,** Oakland, CA
Administrative Assistant

9/95-3/96* **Montgomery Securities**, San Francisco, CA
Research Assistant

1/94-2/95* **Starlite Investments**, Moscow, Russia
Commercial Assistant

9/93-12/93* **Moscow School District**, Moscow, Russia
English Teacher

*Dates are approximate

Refereed Archival Conference Publications

2013 Rolland B and Lee CP. (2013). "Beyond Trust and Reliability: Reusing Data in Collaborative Cancer Epidemiology Research." Proceedings of the ACM Conference on Computer Supported Cooperative Work (CSCW). San Antonio, Texas, USA, ACM: 435-444.

Refereed Archival Journal Publications

2012 Lin Y; Fu R; Grant, E; Chen Y; Eun Lee J; Gupta PC; Ramadas K; Inoue M; Tsugane S; Gao YT; Tamakoshi A; Shu XO; Ozasa K; Tsuji I; Kakizaki M; Tanaka H; Chen CJ; Yoo KY; Ahn YO; Ahsan H; Pednekar MS; Sauvaget C; Sasazuki S; Yang G; Xiang YB; Ohishi W; Watanabe T; Nishino Y; Matsuo K; You SL; Park SK; Kim DH; Parvez F; Rolland B; McLerran D; Sinha R; Boffetta P; Zheng W; Thornquist M; Feng Z; Kang D; Potter JD. (2012). "Association of body mass index and risk of death from pancreas cancer in Asians: findings from the Asia Cohort Consortium." European Journal of Cancer Prevention. Published Online October 10, 2012.

Song M, Rolland B, Potter JD, Kang D. (2012). "Asia Cohort Consortium: Challenges for Collaborative Research." *Journal of Epidemiology*. Published Online May 10, 2012.

2011 Boffetta P, Hazelton WD, Chen Y, Sinha R, Inoue M, Gao YT, Koh WP, Shu XO, Grant EJ, Tsuji I, Nishino I, You SL, Yoo K, Yuan JM, Kim J, Tsugane S, Yang G, Wang R, Xiang YB, Ozasa K, Nagai M, Kakizaki M, Chen CJ, Park SK, Shin A, Ahsan H, Qu CX, Lee JE, Thornquist M, Rolland B, Feng Z, Zheng W, and Potter JD. (2011). "Body mass, tobacco smoking, alcohol drinking and risk of cancer of the small intestine—a pooled analysis of over 500,000 subjects in the Asia Cohort Consortium." *Annals of Oncology* 2011 (Advance Access, published online December 5, 2011.)

Rolland B, Smith BR, Potter JD. (2011). "Coordinating Centers in Cancer-Epidemiology Research: The Asia Cohort Consortium Coordinating Center." *Cancer Epidemiology, Biomarkers and Prevention*. Published OnlineFirst July 29, 2011.

Boffetta P, McLerran D, Chen Y, Inoue M, Sinha R, He J, Gupta PC, Tsugane S, Irie F, Tamakoshi A, Gao YT, Shu XO, Wang R, Tsuji I, Kuriyama S, Matsuo K, Satoh H, Chen CJ, Yuan JM, Yoo KY, Ahsan H, Pan WH, Gu D, Pednekar MS, Sasazuki S, Sairenchi T, Yang G, Xiang YB, Nagai M, Tanaka H, Nishino Y, You SL, Koh WP, Park SK, Shen CY, Thornquist M, Kang D, Rolland B, Feng Z, Zheng W, Potter JD. (2011). "Body mass index and diabetes in Asia: a cross-sectional pooled analysis of 900,000 individuals in the Asia cohort consortium." P*LoS One*. 6(6):e19930. Epub 2011 Jun 22.

Zheng W, McLerran DF, Rolland B, Zhang X, Inoue M, Matsuo K, He J, Gupta PC, Ramadas K, Tsugane S, Irie F, Tamakoshi A, Gao YT, Wang R, Shu XO, Tsuji I, Kuriyama S, Tanaka H, Satoh H, Chen CJ, Yuan JM, Yoo KY, Ahsan H, Pan WH, Gu D, Pednekar MS, Sauvaget C, Sasazuki S, Sairenchi T, Yang G, Xiang YB, Nagai M, Suzuki T, Nishino Y, You SL, Koh WP, Park SK, Chen Y, Shen CY, Thornquist M, Feng Z, Kang D, Boffetta P, Potter JD. (2011). "Association between Body-Mass Index and Risk of Death in More Than 1 Million Asians." *New England Journal of Medicine* 364(8): 719-729.

Workshop Papers

2012 Post-Doctoral Researchers' Use of Preexisting Data in Cancer Epidemiology Research. Rolland B and Lee CP. Position paper for conference workshop on Data-Intensive Science. ACM Conference on Computer Supported Cooperative Work, February 2012, Bellevue, WA.

Magazine Articles and Research Reports

2010 Glenn E. and Rolland B (2010). "Librarians in biomedical research: new roles and opportunities (SLA Research Grant Findings)." *Information Outlook* 14(7): 26(24).

Rolland B and Glenn E. (2010). "Experimenting Outside the Information Center: Non-Traditional Roles for Information Professionals in Biomedical Research." Final grant report, published on SLA Website at http://www.sla.org/pdfs/2008SLAResGrant.pdf.

Posters

2012 Coordinating Centers in Cancer Epidemiology. Rolland B and Lee CP. Poster presentation at Science of Team Science conference, April 2012, Chicago, IL.

2011 Coordinating Centers in Cancer Epidemiology. Rolland B and Lee CP. Poster presentation at Prentice Symposium: Emerging Methodological Issues in Population-Based Chronic Disease Research, October 2011, Seattle, WA.

2009 Showing the Way in SharePoint: What Every Librarian Should Know. Rolland B and Glenn E. Poster presentation at SLA (Special Libraries Association) conference, June 2009, Washington, DC.

2008 A Process for Developing Collaborative Portals for International Biomedical Research Collaborations. Rolland B, Glenn E. and Kim JY. Poster presentation at Global Partners in Public Health, PHI08 conference, September 2008, Seattle, WA.

Conference Presentations

2012 Expanding Teams: Information Professionals and Biomedical Research. Rolland B and Glenn E. Conference presentation at Medical Libraries Association (MLA) Annual Conference, May 2012, Seattle, WA.

2011 Recontexting Disease: Interpreting Shared Data in Cancer Epidemiology. Rolland B and Lee CP. Conference presentation at Society for Social Studies of Science (4S) Annual Conference, November 2011, Cleveland, OH.

2010 National Network of Libraries of Medicine (NNLM) Rendezvous: Experimenting Outside the Information Center. Rolland B and Glenn E. Webinar presentation, July 2010, Seattle, WA.

Experimenting Outside the Information Center: Non-traditional Roles for Information Professionals In Biomedical Research. Rolland B and Glenn E. Conference presentation at SLA (Special Libraries Association) conference, June 2010, New Orleans, LA.

A Panel Discussion on Innovation and Alignment within Research Environments. Washington Medical Librarians Association conference, March 2010, Bothell, WA.

2009 Out of the Library: Integrating Information Professionals into Collaborative Research. Rolland B and Glenn E. Conference presentation at Society of Research Administrators annual conference, October 2009, Seattle, WA.

Grants

2011 The Role of Coordinating Centers in Collaborative Cancer-Epidemiology Studies
PI: Rolland (lead), Lee (University of Washington)
National Cancer Institute R03-CA150036; $185,866

2008 Experimenting Outside the Information Center: Non-Traditional Roles for Information Professionals in Biomedical Research.
PI: Rolland, Glenn (Seattle Biomedical Research Institute)
Special Libraries Association 2008 SLA Research Grant; $20,000

Guest Lectures

2010	University of Washington, Department of Human Centered Design & Engineering: HCDE Project Management, Dr. Charlotte P. Lee
2008	University of Washington Information School: LIS 584 Knowledge Management, Dr. Jeffrey Kim
	University of Washington Information School: LIS 588 Special Libraries, Nancy Gershenfeld