Toward the Design and Testing of a Model-Sharing Collaboratory

by

Mika A. Tomczak

B.S. Mechanical Engineering
Massachusetts Institute of Technology, 2006

SUBMITTED TO THE DEPARTMENT OF MECHANICAL ENGINEERING
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE IN MECHANICAL ENGINEERING
AT THE
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
JUNE 2008

© 2008 Massachusetts Institute of Technology
All rights reserved

The author hereby grants to MIT permission to reproduce
and to distribute publicly paper and electronic
copies of this thesis document in whole or in part
in any medium now known or hereafter created.

Signature of Author: ____________________________

Department of Mechanical Engineering
May 14, 2008

Certified by: ________________________________
David R. Wallace
Associate Professor of Mechanical Engineering
Thesis Supervisor

Accepted by: ________________________________
Lallit Anand
Professor of Mechanical Engineering
Graduate Officer, Department of Mechanical Engineering
Toward the Design and Testing of a Model-Sharing Collaboratory

by

Mika A. Tomczak

Submitted to the Department of Mechanical Engineering
on May 14, 2008 in Partial Fulfillment of the
Requirements for the Degree of Master of Science in
Mechanical Engineering

ABSTRACT

The frequency and importance of collaboration in scientific research continue to increase, and technologies to facilitate these collaborative efforts are being developed. Collaboratories, or Internet-based virtual laboratories, are one such example of these distance technologies. This thesis seeks to increase the efficiency and effectiveness of collaborative endeavors through the creation of a new type of collaboratory.

First, a new collaboratory is proposed and described. This model-sharing collaboratory permits scientists and researchers to publish their computer models and simulations in an interactive format on the Web, allowing other scientists and researchers to use and experiment with their models first-hand. The collaboratory also allows users to create new models, by adding new features to others' models, or by combining more than one existing model. Facilitating scientists' direct interaction with their colleagues' work will minimize the repetition of work and increase the common knowledge shared by the scientists.

Then, a specific target user community for this collaboratory is examined, because the user group ultimately defines the success or failure of a collaboratory. The community based around the Journal of Industrial Ecology, a quarterly academic publication in the field of industrial ecology, is analyzed. Its members' interests, professional goals, computer use habits, and collaboration patterns are all examined; it is concluded that the community has sufficient collaboration readiness and sufficient technology readiness to accept the new collaboratory.

Finally, future pilot tests are described, and critical questions that remain unanswered are proposed. How the answers to these unknowns will help refine the collaboratory is discussed, and how the collaboratory should ultimately be deployed as a free, stand-alone software package is explained.

Thesis Supervisor: David R. Wallace
Title: Associate Professor of Mechanical Engineering
Acknowledgements

There are many people, in many places, without whom the past few years would have been far less interesting and enjoyable.

Firstly, I thank my advisor, David Wallace. David, without your guidance and helpful suggestions, there would be no thesis. Your passion for teaching exceeds that of all others.

I also thank my fellow members of the CADlab. Specifically, I tip my hat to Maureen Lynch for all of her invaluable help. Maureen, you really do keep the lab running.

To my parents, thank you for your perpetual support and encouragement. Mum and Dad, thanks for (generally) keeping me in the loop.

And thanks to the rest of my family, for your open doors, hospitality, and a guaranteed place to stay at Christmas time.

Finally, I cannot express enough gratitude to my former lab mate and better half, " gm " Sittha Sukkasi. You know how much I appreciated your advice and direction in the lab, and now your support from afar is invaluable. I will see you soon!
# Table of Contents

## List of Figures

- Figure 1: Example Figure Description

## List of Tables

- Table 1: Example Table Description

## Chapter 1 Introduction

1.1 Motivation

1.2 Overview

## Chapter 2 Background

2.1 Collaboration in Research

2.1.1 The Evolution of Collaboration and its Meaning Today

2.1.2 The Motivation Behind Collaboration

2.1.3 Who Collaborates

2.1.4 The Benefits of Collaboration

2.1.5 The Costs of Collaboration

2.2 Collaboratories

2.2.1 The Evolution and Meaning of the Collaboratory

2.2.2 Relevant Past Work on Collaboratories

2.2.3 Future Collaboratory Design Considerations

## Chapter 3 The Proposed Collaboratory

3.1 The Theory and Motivation Behind the Collaboratory

3.1.1 A Paradigm Shift

3.1.2 Existing Influences

3.2 The Technology

3.2.1 An Existing Model

3.2.2 DOME

3.3 The Implementation

3.3.1 The Collaboratory Definition, Goals, and Key Concepts

3.3.2 Description of Features
List of Figures

Figure 1: The collaboratory's homepage ................................................................. 46
Figure 2: The issue page in "article view" ............................................................. 47
Figure 3: The article-specific Web page ............................................................... 48
Figure 4: The Web interface of a model ............................................................... 49
Figure 5: Additional features related to the model ................................................. 49
Figure 6: The model-specific discussion forum page ........................................... 50
Figure 7: The model view of the collaboratory ..................................................... 51
Figure 8: The model search page ........................................................................ 52
Figure 9: A list of all models and simulations available in the database ............... 53
Figure 10: The integration option for synthesizing new models ......................... 54
Figure 11: The breadcrumb navigation trail leading to a model page ................. 54
Figure 12: The percentage of co-authored articles per publication, over time ... 72
Figure 13: The inter-group collaboration factor for inter-group collaborations ... 74
List of Tables

Table 1: Beaver’s reasons for which people collaborate (Beaver, 2001) ............20
Table 2: Katz and Martin’s model of who collaborates with whom (Katz & Martin, 1997)..................................................................................................................................................................................23
Table 3: Bozeman and Corley’s Six Collaboration Strategy Types..................23
Table 4: The Seven Types of Collaboratories, as Defined by the SOC (Bos et al., 2007)..................................................................................................................................................................................31
Table 5: Dominant Resource and Activity of Each Collaboratory Type (Bos et al., 2007)..................................................................................................................................................................................33
Table 6: Modified version Katz and Martin’s model (Katz & Martin, 1997)........73
Chapter 1 Introduction

1.1 Motivation

The process of conducting scientific research is no longer a solitary affair. In fact, research has been an increasingly collaborative activity since the mid-1900s, when the concept of collective work effort first surfaced in the scientific domain (Beaver, 2001). Since that time, collaboration has become not only increasingly prevalent in research-based pursuits, but also increasingly broad in scope and breadth.

In recent years, scientific research has been reshaped by the introduction of new technologies; one of the most notable of these technologies is the now-ubiquitous Internet. The increased presence of the Internet and Web technologies has changed the way research takes place, increasing the ease with which scientists can communicate with one another (Finholt, 2002).

The use of these Internet technologies to work with and collaborate with ones remotely-located fellow scientists and researchers has given birth to a new concept: the collaboratory. The collaboratory is essentially a “lab without walls,” or a virtual lab through which scientists and researchers distributed throughout the world can collaboratively work and communicate. The collaboratory is a relatively young concept, but it is ripe for development and innovation.

This thesis seeks to build upon the notion of the collaboratory and the drive to increase the ease with which scientists and researchers can collaborate. As such, a new collaboratory is proposed, with several fundamental goals. First, the collaboratory seeks to improve the efficacy of collaboration (both collocated and remote), and expand the horizon of potential for collaborative research. It also tries to introduce novel functions that would offer new collaborative capabilities and ultimately make collaboration more productive, time-efficient, and beneficial to all parties involved. Finally, the collaboratory aims to reduce the perceived costs of collaboration to the scientists involved in the effort.

This collaboratory is a Web application that is in the early stages of development. This thesis aims to increase the scope of collaboratories as they are known today, by introducing new functionalities that have never before been utilized in a collaboratory environment.
1.2 Overview

This thesis is made up of three main parts: an overview of the present role of collaboration in research, and the increasing presence of the collaboratory in scientific collaboration; a proposal of a new collaboratory structure that should improve the efficacy of collaborations; and a detailed analysis of the target user group for this new collaboratory. Additionally, a discussion of how the collaboratory should be tested and suggestions for its ultimate deployment are provided. The remaining portion of this chapter will outline the structure of this thesis in more detail.

Chapter 2 - Background

To establish the building blocks for the rest of this thesis, Chapter 2 first provides the reader with a summary of the important work that has been done to date on collaboration in research. A variety of matters that relate to collaboration in research are discussed.

First, the evolution of collaboration in the scientific domain is addressed, so that the current connotation of "collaboration" can be appreciated. After establishing its meaning, the different factors that motivate scientists and researchers to collaborate are discussed; understanding why scientists choose to collaborate with others helps when trying to facilitate the collaborative process. Then, the question of who collaborates is explored; following the lead of various authors in the field, different approaches are taken to examine the types of people and organizations who participate in collaborative efforts. Finally, the benefits and costs, as perceived by the collaborators, are discussed.

After obtaining an understanding of collaboration as it stands today, the overview is taken one step further. The term "collaboratory" is introduced and defined. Briefly, a collaboratory is like a research laboratory without walls, or an organizational entity that supports collaborative efforts that span distance. It permits interaction between scientists and researchers who share nothing more than common interests, and who may not even be known to each other. The collaboratory is the fundamental construct around which the remainder of this thesis is built.

Then, the main research projects to date that have studied collaboratories are discussed; of note, the Science of Collaboratories (SOC) project is examined in detail, because the taxonomy of collaboratories that they provide is later expanded upon. Finally, considerations that should be borne in mind when designing new collaboratory technologies are discussed; these considerations will be revisited in later chapters.

Chapter 3 - The Proposed Collaboratory

Chapter 3 describes in detail the new collaboratory that is proposed in this thesis. This new model-sharing collaboratory, which is intended for use in conjunction with academic publications, allows users to publish on the Internet the computer models and simulations related to their research. Then, readers of the journal can
both interact with and use the models they are reading about in the publication, and build upon those models to synthesize new ones.

First, the theory and motivation behind the collaboratory are discussed. This new model-sharing collaboratory represents a paradigm shift in the organization of science; “big science” and the distribution of work among many researchers is increasingly common.

Then, the essence of the technology behind the collaboratory is discussed. Two existing projects from the author’s research lab provide the technical basis for the collaboratory. First, an existing model of Commons-Oriented Information Syntheses, which provides the theoretical basis for the collaboratory, is explained. Second, DOME, the software that enables the sharing of models and simulations, is briefly described.

Next, the collaboratory’s implementation is described. After establishing the goals that motivate the collaboratory’s design, there is a detailed explanation of each of the collaboratory’s features, accompanied by corresponding images of the user interface. Following the description of features, there are some general use scenarios that describe how the average user might interact with the collaboratory.

Finally, the taxonomy of collaboratories proposed by the SOC, that was discussed in Chapter 2, is revisited. The taxonomy is revised to include this new model-sharing collaboratory, and a definition specific to this new collaboratory that fits the mold of the SOC’s work is established.

Chapter 4 - The User Group

Having established an understanding in Chapter 3 of what exactly constitutes the proposed new collaboratory, Chapter 4 discusses in detail the target user group of the new collaboratory. Because the user group is a key to the collaboratory’s success or failure, it is critical to have as clear an understanding of the user group as possible. As such, this chapter provides a detailed description and analysis of the composition of the target user group.

First, the user group, which consists of the contributors to and readers of the Journal of Industrial Ecology (JIE), is described. A brief background of the field of Industrial Ecology is provided, and the reasons for picking the JIE community as an example group are explained.

Then, the community’s “collaboration readiness” is discussed, to determine if the group is mature enough, from a collaboration standpoint, to accept the new collaboratory. Both its contributing authors and its readership are surveyed in detail; it is concluded that the majority of both parties have participated in collaborations, and should be accepting of the new collaboratory.

The group’s “technology readiness” is also discussed, to gauge if the community has had sufficient exposure to technology to be comfortable with adopting this
new collaboratory. The JIE community and its computer use habits are examined on a macro level; it is concluded that because of its members' almost exhaustive exposure to Internet technologies, the group should be capable of using this Web application.

Finally, two use scenarios, specific to the Journal of Industrial Ecology, are provided, as examples of how the collaboratory is designed to be used.

Chapter 5 - Future Testing and Conclusions
This final chapter first provides a brief summary of the work presented in this thesis. But before reaching any definitive conclusions about the success of this new collaboratory, it is this chapter explains that a user test must be conducted. A pilot of this collaboratory must be deployed within the community of the Journal of Industrial Ecology, and feedback must be obtained accordingly.

There are several questions that must be answered based on results of the pilot, before attempting to deploy the collaboratory as a free, publicly-available download. Namely, the willingness of participants to openly share the models and simulations from their research must be verified. Whether this collaboratory actually increases efficiency of scientific collaboration over the current state should also be corroborated. Feedback regarding the collaboratory's ease of use should be obtained, and adjustments made accordingly. Finally, it should be determined if the collaboratory meets its overall design goals, and if it is indeed helpful in general.

It is explained that, after determining answers to these unknowns, the design and implementation of the collaboratory can finally proceed. Additional work is needed to bundle the collaboratory into a single, easy-to-use, downloadable software package, which would be the ideal final deliverable of this project.
Chapter 2 Background

The desire to improve the effectiveness and efficiency of collaboration in research is the motivating factor behind this thesis. Therefore, it is critical to understand both the role of collaboration in research, and the role of technology in collaboration, before addressing in detail the Web application created as part of this work. Neither collaboration nor technology has been introduced to the research sphere without controversy; the following sections are intended to help better understand the challenges that each has faced.

2.1 Collaboration in Research

2.1.1 The Evolution of Collaboration and its Meaning Today

Collaboration in research has existed for over a century, and considerable effort has been devoted to its study by modern-day scientists and researchers. However, the goal of this section is to merely provide a whirlwind tour of the history of collaboration, and how it has evolved to its present-day form. Since there is no single agreed upon definition of collaborative work, it will prove helpful to look briefly at the origin of collaboration; from there, a meaning of collaboration relative to the context of research can be understood.

The earliest evidence of collaboration in research surfaces within French chemistry work of the early 1800s (Beaver, 2001). However, it was not until after World War I that collaboration started to grow rapidly within the field of science (Beaver, 2001). When the field of science professionalized around the same time, and scientists began to work together as epistemic equals, collaborative research became functionally significant in the scientific scene (Wray, 2002).

Obviously, collaboration has not always been prevalent in scientific research. Prior to the breakout of collaborative research, much scientific research was the result of collective efforts. In collective research, although many hands perform experiments, it is ultimately the role of a single researcher to direct the project, compile the data and take responsibility and credit for the results. In collaborative research, however, “credit and responsibility are shared” between all the scientists involved (Wray, 2002).
In its early years, collaboration in scientific work took place between individual scientists working together on a single project; collaboration was not a premeditated act, but more of a practice that arose out of necessity when more than one person was needed to complete a project. More recently, however, the notion of scientific collaboration has evolved and become less specific; it now connotes a “group structure” in which lab projects are taken on by teams of scientists from the start (Beaver, 2001).

Some researchers have defined collaboration in general terms that do not limit the way in which communication and information transfer occur. Fox and Faver describe collaboration as the communication and exchange of results, a process that engenders teamwork, cooperation, and interdependence (Fox & Faver, 1984). Beaver describes collaborative work as research performed as a group, in contrast with research performed by a single individual (Beaver, 2004). Another very general, accepted meaning of collaborative research is “working closely with others to produce new scientific knowledge or technology” (Bozeman & Corley, 2004).

Yet other researchers have sought to define in very specific terms what constitutes collaborative research. In an effort to devise a method by which to quantitatively measure the extent of collaboration, Katz and Martin (Katz & Martin, 1997) examine the two extremes between which a collaboration lies. They explain that a collaborator is neither as broad as anyone who provides input into the project, nor as specific as only the scientists who contribute directly to all the main tasks; the definition of a collaborator lies between the two extremes. Katz and Martin assert that this in-between position usually includes, although not exhaustively: those who work together during the duration of the project and make frequent or substantial contributions; those whose names appear on the project proposal; those responsible for one or more main elements; those responsible for a key step; or the original person who proposed the project.

Clearly, there are multiple interpretations of the meaning of collaboration in research and scientific work. The differences in semantics can be put aside for the moment, however, since one single definition is not crucial. Instead what is important is the fact that collaborations exist throughout the scientific world, with varying degrees of formality, and defined by boundaries that are increasingly blurred. A collaboration can take on a variety of meanings, depending on the researcher; technologies that hope to cater to this constantly evolving, and increasingly less-precise, research practice must remain open-minded and remember that the world of scientific collaboration is a dynamic one.

Researchers have also invested considerable effort in coming up with quantitative ways to measure the extent of collaborative participation in a given research endeavor. One of the more commonly used measures of collaboration is based on the assumption that co-authorship of publications is an indicator of collaboration between scientists and scholars; this assumption is made by many researchers, including Fox and Faver. However, other researchers consider this assumption insufficient, and have sought to further explore the area of
collaboration measurement. Katz and Martin question this bibliometric analysis, arguing that although it is assumed that co-authorship is an indication of collaboration, there exists little proof to support this assumption. Ultimately though, they do not provide any concrete suggestions for a more appropriate measurement model.

Despite the differences in opinion found throughout the study of collaborative research, there is one indisputable aspect of collaboration. The amount of collaborative work in science is rapidly growing; it is increasingly prevalent throughout the field, and as a result, individual work is decreasing. The next section will briefly discuss researchers’ theories as to why this increase in collaboration is occurring, and at such a rapid rate.

Finally, it should be noted that researchers are in disagreement about the future implications of the growth of collaboration. Some of the earlier literature on collaboration supports the theory that collaborative research will eventually cause the extinction of individual research; Wray mentions a prediction by Beaver and Rosen (Beaver & Rosen, 1979b) that the scientific future will see “the virtual demise of the lone researcher”. However, Wray refutes this prognosis, asserting that the individual researcher and single-authored papers serve a role in science that is just as significant as that of the collaboration. For example, he suggests that the individually-authored paper plays a key role in the development of young scientists, and that it is an opportunity for young scientists “to prove themselves, both to themselves and other scientists.” They will not, he adds, remain motivated or interested in the sciences without this opportunity. Wray’s assertion is a reminder that all sense of individualism should not be forgotten when considering technologies for collaboration; after all, the individual is still the fundamental unit of research.

2.1.2 The Motivation Behind Collaboration

As mentioned in the previous section, the extent of collaborative research in science today is growing undisputedly. Many researchers have sought to understand why this increase is occurring, and at such a rapid rate. Asking what is causing this growth is analogous to questioning the motivations behind collaboration. Multiple researchers’ opinions of the motives behind collaboration will be summarized below.

It should be noted that the motivations behind collaboration differ from the benefits of collaboration, even though not all literature on the topic makes the distinction. The motivations are what instigate the collaborations, whereas the benefits are the outcomes of the collaborative effort. Often, perceived gains in the form of outcomes, or benefits, may present themselves as motivating factors in the minds of researchers. However, those studying collaboration must clearly maintain the distinction between the motivations and the benefits.
Much of the literature on collaboration lists various factors that often motivate collaborative work, although each author approaches the matter slightly differently. Beaver and Rosen together were some of the pioneers in the study of collaboration; they authored three papers (Beaver & Rosen, 1978), (Beaver & Rosen, 1979a), (Beaver & Rosen, 1979b) in the late 70s that were, at the time, the most comprehensive publications on collaboration in science. More recently, Beaver (2001) has readdressed their work and summarized their findings in the context of modern-day collaborations. He authoritatively lists the 18 points in Table 1 as the motivations for collaboration.

Table 1: Beaver's reasons for which people collaborate (Beaver, 2001)

<table>
<thead>
<tr>
<th></th>
<th>1. To have access to expertise</th>
<th>10. To retool, learn new skills or techniques</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2. To have access to equipment, resources, or “stuff” one doesn’t have</td>
<td>11. To satisfy curiosity and intellectual interest</td>
</tr>
<tr>
<td></td>
<td>3. To improve access to funds</td>
<td>12. To share the excitement of an area with new people</td>
</tr>
<tr>
<td></td>
<td>4. To obtain prestige or visibility; for professional advancement</td>
<td>13. To find flaws more efficiently, reduce errors and mistakes</td>
</tr>
<tr>
<td></td>
<td>5. For efficiency; multiple hands and minds; easier to learn tacit knowledge.</td>
<td>14. To remain more focused on research, because others are dependent</td>
</tr>
<tr>
<td></td>
<td>6. To make progress more rapidly</td>
<td>15. To reduce isolation, and recharge ones energy and excitement</td>
</tr>
<tr>
<td></td>
<td>7. To tackle “bigger” problems</td>
<td>16. To educate (a student or oneself)</td>
</tr>
<tr>
<td></td>
<td>8. To enhance productivity</td>
<td>17. To advance knowledge and learning</td>
</tr>
<tr>
<td></td>
<td>9. To get to know people, create a network</td>
<td>18. For fun, amusement, pleasure</td>
</tr>
</tbody>
</table>

In a similar manner, Katz and Martin (Katz & Martin, 1997) assemble a list of six factors that may serve as motivators to a scientific collaboration. Many of their reasons include, or elaborate on, the points that Beaver lists in his work. Firstly, they say that costs of conducting research have increased to levels that are prohibitive to the individual scientist; in order to receive funding for the instrumentation needed to conduct research, scientists have been forced to pool resources and collaborate more closely. Secondly, travel and communication have become less costly and more readily available in recent years. Thirdly, they assert that science is a social field that relies on interaction with fellow scientists. Fourthly, Katz and Martin believe that there is an increasing need for specialization, requiring that many scientists collaborate to have sufficient knowledge to understand and tackle a single, large problem. Fifthly, they believe that interdisciplinary fields are growing in importance, and that collaborations across both disciplines and sectors are necessary to obtain the range of fields needed to solve problems. Finally, they observe that political factors motivate some scientific collaborations. They state, for example, that there has been a movement for scientists in Western Europe to collaborate with their colleagues in Eastern Europe, to ideally strengthen the political and cultural ties between the two previously distinct European halves.

Sociologists Fox and Faver (Fox & Faver, 1984) sought to determine “scientists own perceptions of their motivations for collaborating,” by conducting hundreds
of one-on-one interviews with scientists. Although they, too, list several motivating factors, they do not distinguish between “motivations for” and “advantages of” collaboration; this lack of distinction leaves the reader with an ambiguous understanding of their explanation for the growth of collaboration in science. Nonetheless, it is worth mentioning their interpretation of the reasons causing the growth phenomenon. Firstly, they conclude that scientists believe that collaboration allows them to “increase their efficiency and enhance the overall quality of their work.” When research tasks are large, collaborations allow for a division of labor to appropriately and efficiently divide tasks among scientists. Secondly, Fox and Faver believe that scientists often suffer from “detachment and aloneness,” and that collaboration with other scientists is a way to alleviate this work-induced isolation. Finally, they believe that collaboration comes with a commitment to others, and that this commitment motivates scientists to work harder, stick to research schedules, and generally maintain a higher level of energy when working.

Beaver, Katz and Martin, and Fox and Faver all list individual factors, any of which may, according to their theories, act as motivation for a scientist to collaborate. However, none of these authors uses his or her reasons to assemble a single, cohesive argument as to why scientific collaborations have increased so significantly in the past few decades, and why they continue to grow in importance today.

Wray, however, adopts a more holistic view of the trends seen in collaborative research, and assembles a comprehensive philosophical argument that explains not only the motivation behind collaboration but what has sustained its growth in significance within the field of science. He uses the logic of “functional explanations”, which he defines as explanations that “identify specific causal effects of a practice or institution and then argue that the practice exists in order to promote those effects,” to first explain why collaborative research plays a significant role in scientific communities, then to explain how it persists because it is effective, and finally to establish that initial collaboration occurs prior to its resulting success.

The first part of Wray’s argument establishes why collaboration is significant to scientific research by listing five ways in which collaborative work allows scientists to “realize their epistemic goals.” These five reasons are similar to the individual motivating factors given by all the authors mentioned above, although Wray’s reasons are only part of his argument. He argues firstly that collaboration appears to increase the quality of research performed. Secondly, he explains that collaboration has allowed research that would not otherwise have been possible by any one single scientist. Thirdly, he explains that the collective participation of many scientists is a way to ensure that research findings will not be lost or forgotten. Fourthly, he believes that there is a positive correlation between collaboration and productivity, and that collaborative research is “partially responsible for the rapid growth of scientific knowledge.” Finally, he believes that collaboration is an important aspect in the training of young scientists.
Wray then proceeds to argue that collaboration persists “because it enables scientific communities to realize their epistemic goals.” In fact, he explains that collaboration not only persists after it begins, but it is becoming increasingly popular in many areas of science. Even though groups of collaborators are dynamic, and constantly “split, merge, and disappear,” collaboration continues to persist. He also provides evidence to support the theory that if there is a viable alternative, scientists are inclined not to collaborate. However, he explains that collaboration can only persist in the proper environment where “(i) substantial resources are required for which there is competition, and (ii) the community of researchers are epistemic equals.”

Finally, Wray sets out to determine the initial cause of collaboration, to prove that collaborative work occurs causally before its success. He makes the logical point that the initial cause of collaboration is probably not scientists’ individual effectiveness at research. Through some examples, he concludes that collaboration was never forced upon scientists prematurely, and that only after the research environment changed such that collaboration was indispensable did it become the norm. Only after scientists begin to collaborate and witness the benefits of collaboration themselves will subsequent collaboration be induced, and eventually become the norm in their field. This causality allows Wray to concisely state his main argument for the prevalence of collaboration in science: that “in certain fields, those in which scientists must compete for access to resources in order to engage in research effectively, collaborative research has become the norm.”

### 2.1.3 Who Collaborates

To have a full overview of collaboration in science, it is useful to understand not only the motivation behind why people collaborate, but also who collaborates. At first glance, it may seem ambitious to attempt to generalize, over all of science, the people involved in collaborative work. However, researchers have approached this topic from several different angles, and have created multiple interesting classifications that characterize those who collaborate.

One trait about collaboration in science that continues to intrigue researchers is the fact that collaboration does not take place to equal extents across fields. It is consistently observed that collaboration occurs less in social sciences and humanities than it does in other areas of science (Wray, 2002). It has also been documented that theoreticians collaborate less than experimentalists (Beaver & Rosen, 1978). Wray suggests that this disparity is observed because the variety of funding available across fields leads to a variety in the extent of collaboration necessary.

Researchers have also attempted to characterize the types of collaborations that occur in science, in terms of who collaborates with whom, and also in terms of the goals of each collaborator. Katz and Martin propose a scheme that divides the different groups of people who collaborate, and the resulting types of
collaborations that occur. Reproduced in Table 2 is the breakdown, in tabular form, as published in their paper; this table will be used as a basis for describing a test user group later in this thesis. They note that, although collaborations can be broken down by group level and distinction between intra- and inter- forms, the individual person is the fundamental unit of collaboration; any collaboration comes down to the cooperation between two or more individuals, and person-to-person relationships are key.

Table 2: Katz and Martin’s model of who collaborates with whom (Katz & Martin, 1997)

<table>
<thead>
<tr>
<th>intra -</th>
<th>inter -</th>
</tr>
</thead>
<tbody>
<tr>
<td>individual</td>
<td>between individuals</td>
</tr>
<tr>
<td>group</td>
<td>between individuals in the same research groups</td>
</tr>
<tr>
<td>department</td>
<td>between individuals or groups in the same department</td>
</tr>
<tr>
<td>institution</td>
<td>between individuals or departments in the same institution</td>
</tr>
<tr>
<td>sector</td>
<td>between institutions in the same sector</td>
</tr>
<tr>
<td>nation</td>
<td>between institutions in the same nation</td>
</tr>
<tr>
<td></td>
<td>between institutions in different sectors</td>
</tr>
<tr>
<td></td>
<td>between institutions in different countries</td>
</tr>
</tbody>
</table>

Bozeman and Corley (Bozeman & Corley, 2004) conducted a detailed statistical analysis of “collaboration strategies” among scientists, yielding a way to categorize collaborators based on their scientific personalities. They developed a taxonomy that includes six collaboration strategy types, each of which encapsulates the motivating factors of the collaborator it describes. Table 3 showcases their collaboration strategy types.

Table 3: Bozeman and Corley’s Six Collaboration Strategy Types

<table>
<thead>
<tr>
<th>Collaboration Strategy Type</th>
<th>Chooses a collaborator based on ...</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Taskmaster</td>
<td>... work ethic and whether the collaborator sticks to schedules</td>
</tr>
<tr>
<td>2 Nationalist</td>
<td>... if fluent in their own language, if of same nationality</td>
</tr>
<tr>
<td>3 Mentor</td>
<td>... motivation to help junior colleagues and grad students</td>
</tr>
<tr>
<td>4 Follower</td>
<td>... imposition by administration requiring the collaboration, and if collaborator has a strong reputation</td>
</tr>
<tr>
<td>5 Buddy</td>
<td>... length of time of acquaintance, quality of past collaborations, whether the collaborator is fun and entertaining</td>
</tr>
<tr>
<td>6 Tactician</td>
<td>... skill set complementary to their own</td>
</tr>
</tbody>
</table>

The two classification schemes – that of Katz and Martin, and that of Bozeman and Corley – are mutually independent. They approach the question of who collaborates from two different angles, and present very different schemes for classifying the affiliations of those who collaborate, to better understand the
types of people involved in collaborations. Katz and Martin’s scheme proves useful when analyzing at a high level the collaborations that take place among a large community, whereas Bozeman and Corley’s scheme is helpful when analyzing the personal motivation of individual collaborators. Each provides a unique way to look at those involved in collaborations.

2.1.4 The Benefits of Collaboration

In most of the literature that addresses scientific collaboration, there is discussion about the “benefits” and the “costs” of collaboration. However, the meaning of the “benefits” of collaboration is ambiguous in the context of collaborative work, and should be further explained before proceeding. It is of interest to determine to whom these benefits are beneficial, and from whose point of view. The two questions are likely related, since the benefits often depend on the motivation behind the collaboration. For example, if the collaboration is politically motivated (as explained in the preceding section), then perhaps the collaboration is not directly beneficial to the individual scientist performing the work.

Recall that there is a distinct difference between the benefits of collaboration and the motivations for collaboration. Whereas motivations, or purposes, for collaboration are the driving forces that instigate a collaborative effort, the benefits are the positive outcomes of the resulting collaborative work.

Each piece of literature that has been examined so far addresses what are perceived as “benefits” of collaboration; however, each has differing motives for looking at these advantages. Before summarizing the most commonly cited advantages of collaboration, the reasons for each author’s examining the benefits will be discussed.

Three works examine the advantages of collaboration from similar points of view. Beaver (2001) addresses the advantages and disadvantages to both science and the overall scientific institution of “big science”, or the new organizational structure of science that favors large laboratories and collaborative work. Similarly, Katz and Martin discuss the advantages of collaboration to individual collaborators as part of their attempt to holistically characterize the collaboration scene. Finally, Fox and Faver also look at the advantages and disadvantages of collaboration to the individual scholar, with the goal of better understanding scientists’ perceptions of collaboration, and the “larger effects of collaboration on the scientific and scholarly enterprise.”

Wray and Beaver (2004) both examine the epistemic authority of collaborative research, and treat the benefits and costs of collaboration accordingly. Wray addresses the advantages of collaborative work to support his “functional explanation” for the prevalence of collaboration in science; he states five ways in which collaboration allows researchers to meet their “epistemic goals.” It is assumed here that researchers’ meeting their goals is favorable, and that these five points are “benefits” of collaboration. Similarly, since he seeks to determine
if a “team of researchers in a lab is superior to a single, lone researcher,” Beaver lists four examples in which collaborative research does appear to have greater epistemic authority; these examples can be generalized as benefits of overall collaborative work in science.

Why do the benefits and costs of collaboration matter to this discussion? The answer is twofold. First, the goal of this research is to facilitate scientists’ collaborative efforts. To do so, one must understand not only why researchers choose to interact and communicate with one another, but the consensus of what is beneficial to collaboration. Additionally, much of the previous work on scientific collaboration has studied “formal” or “traditional” collaborations; little effort has been devoted to the informal or casual online communities that will be examined here. It should be beneficial to synthesize and foster the benefits of formal collaborations for members of online communities who do not collaborate with one another to the same extent as those researchers studied in the existing literature.

The most commonly cited benefits of collaboration, from the literature previously discussed, are listed below. The benefits are categorized into one of two groups: either benefits that are advantageous to the scientist personally, such as ones that might help advance his career or enhance the enjoyment of his work, or ones that directly benefit the project or work toward the betterment of science overall. Of course, the two categories are interrelated, and ultimately not so distinct, but using such a distinction provides a first-order perspective with which to consider the perceived benefits of collaboration: who benefits, and why?

- Advantages of collaboration that personally benefit researchers:
  - Provides a way to transfer knowledge or skills between researchers, especially social and management skills that are not taught or learned formally. Opportunity to learn something about others’ specialties.
  - Helps overcome “intellectual isolation” or “deprivation” with intellectual companionship; helps form work and personal relationships with others.
  - Helps generate a wide network of contacts in the scientific community; provides more resources to ask for help and advice.
  - Helps visibility of one’s work; work can be diffused into a wider network, and will likely be cited more frequently and have a greater impact.
  - Provides a source of stimulation and creativity.

- Advantages of collaboration that benefit the project, or science overall:
  - Lends extra authority to the results of the work, because each expert adds “methodological facets” that add up; intersubjective verifiability.
  - Sharing of knowledge, skills, and techniques means talents are used more effectively; all areas of a problem can be tackled.
Multiple minds mean multiple views, new ideas, new insights, or new perspectives not otherwise seen on one's own.

More hands mean more labor. Increased efficiency and enhanced overall quality of work by capitalizing on individuals' specializations and the separation of tasks.

Work commitments to others sustain motivation. Increases likelihood of sticking to schedules and agendas. Increases energy in a project.

Increases likelihood of finding novelties or errors, and see what others cannot, because each collaborator has an "outsider's viewpoint" with respect to different aspects of the project.

These perceived advantages to collaboration should be borne in mind both when discussing the perceived costs of collaboration, and later, when considering ways in which to facilitate collaboration by capitalizing on the advantages already benefiting the individual scientists and their research projects. Not all researchers participating in collaborative work experience all of the benefits listed above; however, understanding the range of benefits derived from collaboration is critical to helping the scientific community as a whole.

2.1.5 The Costs of Collaboration

The costs of collaboration are often studied in conjunction with the benefits of collaboration; they are the antitheses of the advantages, and often the reasons that make collaborative work appear less desirable. As with the benefits of collaboration, one must consider to whom these costs are disadvantageous; the people affected might include the individual researcher, the project's principal investigator, the provider of the project's funding, or even the scientific institution as a whole.

The costs of collaboration are often more tangible than the advantages, and the relative significance of the costs weighed against the benefits varies from project to project. Nonetheless, the following list of costs encompasses the most frequently cited perceived disadvantages to collaborative scientific efforts.

- Costs of collaboration
  - Additional financial costs: travel and subsistence costs; equipment and material might be moved and set up multiple times.
  - More time spent:
    - preparing proposals jointly or securing joint funds
    - jointly defining research problems and planning an approach
    - moving between locations for different parts of the project
    - keeping all collaborators informed of progress
    - resolving differences of opinion
    - jointly writing up results and resolving disagreements
  - Increased administration required to manage the project.
Differences in customs or systems of the participants may lead to clashes:
- management culture clashes
- financial systems
- rules on intellectual property rights

Personal and socio-emotional costs: developing and maintaining good working relationships with one's collaborators requires time and "emotional energy".

Slow, uncooperative partners can delay or jeopardize a project and lead to frustration.

Difficulty properly allocating credit to individuals, when looking at co-authored papers. Who did what?

A loss in overall quality of work, because quality control is more difficult.

The above list reflects the costs of collaboration as perceived by most of the people involved in the effort. Similar to the benefits, not all collaborative endeavors are plagued by all of the above costs. Nonetheless, it is imperative to understand what participants deem the disadvantageous aspects of collaboration, so that the application that is developed here can work toward the goal of lowering at least some of the perceived costs of collaboration. In fact, many of the points mentioned above can be alleviated with the application that will be proposed in upcoming chapters.

The relative strengths of the benefits and costs of collaborative efforts are not fixed; there exist many factors that affect the balance of each in the eyes of the collaborators. The strength of these factors is what helps scientists weigh the pros and cons of, and ultimately decide whether to pursue, collaborative work. For example, the type of research being conducted, and the approach being used both come into play. (E.g., is the research observational or experimental?) The degree of competition in the field often affects the perception of collaboration. If the work challenges established methods or views, collaboration and group effort are often more desirable. Finally, whether the collaboration is within or between departments or institutions has an effect; closer physical proximities can reduce time and money costs.

2.2 Collaboratories

The increased role of new technologies has become clear in the realm of scientific collaboration; like in many other areas of science, a technological facelift was inevitable. As its name suggests, the collaboratory embodies the merging of current technologies with collaborative science. It is a concept that was first developed in the late 1980s, but that has continued to evolve since its inception and has been the focus of many academic studies in recent years. The following sections will summarize the evolution of the concept of the collaboratory, provide an overview of relevant work that has been conducted in the area of collaboratories to date, and discuss issues that must be considered when
designing collaboratories for use with today's collaborative efforts. These historical points should be understood before addressing the collaboratory that will be developed in the upcoming chapters of this work.

2.2.1 The Evolution and Meaning of the Collaboratory

William Wulf, a professor of computer science, coined the term “collaboratory” in 1989, by combining the terms “collaboration” and “laboratory” (Wulf, 1989). At the time, collaboration in science was the focus of much academic attention, as were the up-and-coming personal computer and Internet, too. Since its inception, the concept of the collaboratory has been the focus of many research studies, and its denotation has continually evolved as collaboratory implementations have been tested and revised.

The term “collaboratory” was first coined with the goal of enabling collaboration without the confines of a physical laboratory, using new technologies to work around previous constraints. As the name suggests, Wulf originally envisioned the collaboratory:

as a center without walls, in which the nation’s researchers can perform their research without regard to physical location—interacting with colleagues, accessing instrumentation, sharing data and computational resources, and accessing information in digital libraries (Wulf, 1993).

Wulf foresaw a new application in communications for computing technologies that had, up until that point, only been used in science for running models and simulations. He imagined software that would allow scholars to communicate remotely with non-collocated colleagues, to remotely use instrumentation and machinery to perform experiments, and to remotely perform tests and analyze data with the same ease as what they were presently doing locally (Wulf, 1993).

Over the following decade, several software pilots were deployed that attempted to embody Wulf’s “distributed laboratory” vision of the collaboratory. As software implementations evolved, so did the meaning of the collaboratory. In his benchmark paper that describes collaboratories as a “new form of scientific organization,” Finholt (Finholt, 2002) revises Wulf’s 1989 definition of the collaboratory to stress its importance as an “organizational entity.” The key attributes of the collaboratory are, according to Finholt, the “human and behavioral aspects” that support “rich and recurring interaction.” Although still part of many collaboratories, the essential elements are no longer the physical aspects like the sharing of equipment, but the opportunities for interaction, sharing, and collectively creating.

Bos, Zimmerman, Olson, and others at the University of Michigan’s School of Information have been studying collaboratories quite extensively for the past six years. In 2002, they began a five-year project entitled the Science of Collaboratories (SOC), the goal of which was “understanding the technical and
behavioral principles that can lead to better, more successful design of collaboratories in the future” (Olson, 2004). At the start of their project, the group agreed upon the following definition of collaboratory, which has since been upheld by researchers as the most appropriate:

A collaboratory is an organizational entity that spans distance, supports rich and recurring human interaction oriented to a common research area, and fosters contact between researchers who are both known and unknown to each other, and provides access to data sources, artifacts, and tools required to accomplish research tasks (Bos et al., 2007).

The SOC definition of collaboratory is much closer to Finholt’s revised definition than it is to Wulf’s original; it, too, stresses the importance of human interaction, but adds the key point that researchers involved in a collaboratory can be either known or unknown to each other at the start of their collaboration.

Definitions aside, Finholt identifies a model for the role of collaboratories in the larger picture of science that succinctly summarizes how science has evolved over the past several decades, and where collaboratories fit into this evolution. He recognizes that, in the middle of the 1900s, the organization of science revolved around the concept of the “invisible college”—a system by which “the bulk of new knowledge is created by a small core of elite researchers working among themselves.” Finholt notes that, in fact, at one time, fifty percent of scientific publications were being produced by 16% of practicing scientists.

With the restructuring of the scientific organization, the trend has shifted to that of “big science,” also described as “distributed intelligence.” In this model, scientific effort is mobilized such that work is distributed among many scientists, and a larger percent of the “scientific workforce” is able to partake in the creation of new knowledge. This increased distribution of work is one driving force behind the need for distance technologies like collaboratories. Thus, the paradigm shift that has seen scientific organization evolve from the model of the invisible college to “distributed intelligence” is the same shift that has introduced the collaboratory and its many incarnations in recent years.

2.2.2 Relevant Past Work on Collaboratories

Between the time when Wulf first coined the term “collaboratory” and the start of the SOC project, over 100 collaboratories emerged in science and engineering (Bos, Olson, & Olson, 2003). The implementations of each differed greatly by project, as did the degree of success experienced by each; although some succeeded, many failed, and for a variety of reasons. Interested in the different approaches to collaboratories, and more importantly, in determining the most successful social and technical practices for sustainable, successful collaboratories, Bos, Olson, Olson and others at the University of Michigan embarked on their SOC project.
The Science of Collaboratories project was a large endeavor, with many goals. Most immediately relevant to the collaboratory design that will be presented later in this thesis was the project’s goal to “identify organizational patterns, somewhat similar to design patterns... which could be used by funders and project managers in designing new collaborations” (Bos et al., 2007).

One of the more significant outcomes of the SOC venture was the establishment of a taxonomy system that categorizes collaboratories as one of seven types. To come up with the seven types of collaboratories, the researchers involved in the SOC located 212 online collaboratories and conducted a “landscape sampling” of them, looking ultimately for three characteristics: novelty, success, and prototypicality (Bos et al., 2007). Using the results of their survey, they created a seven-category classification system; in their paper on the project, they provide a definition of the collaboratory type, an example of this type of collaboratory, key technology issues, and key organizational issues for each of the seven collaboratory types. Table 4 summarizes these key attributes for each of the seven collaboratory types.
<table>
<thead>
<tr>
<th>Name</th>
<th>Definition / characteristics</th>
<th>Example</th>
<th>Technical Issues</th>
<th>Organizational Issues</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Shared Instrument</td>
<td>* Increases access to a scientific instrument  &lt;br&gt; * Often, provides remote access to expensive instrument</td>
<td>Keck Telescope atop Mauna Kea summit, Hawaii – made remotely accessible.</td>
<td>• Synchronous communication  &lt;br&gt; • Remote-access technology  &lt;br&gt; • Security and handling of large datasets  &lt;br&gt; • High-end electronic notebooks</td>
<td>• Allocating access  &lt;br&gt; • Technical support; maintaining social relationships with technicians  &lt;br&gt; • Maintaining contextual knowledge.</td>
</tr>
<tr>
<td>2. Community Data Systems</td>
<td>* Information resources created, maintained, or improved by a geographically distributed community.  &lt;br&gt; * Semi-public and of wide interest.</td>
<td>Protein Databank—single worldwide repository for processing and distributing 3D structure data of protein and nucleic acid molecules.</td>
<td>• Data standardization  &lt;br&gt; • Modeling and visualization techniques</td>
<td>• How to motivate contributors to supply public goods?  &lt;br&gt; • Large-scale decision-making methods required</td>
</tr>
<tr>
<td>3. Open Community Contribution System</td>
<td>* Open project.  &lt;br&gt; * Aggregates efforts of many geographically separate individuals toward common research problem.  &lt;br&gt; * Contributions in form of work, not data.</td>
<td>Open Mind Project – volunteers take Web surveys. Aggregated data made available to projects requiring such data (the collaborative Web site does data collection).</td>
<td>• Cross-platform system that’s easy to learn and use  &lt;br&gt; • Behind-the-scenes standardized data formatting</td>
<td>• Maintaining quality control of large, distributed user group (community vetting)  &lt;br&gt; • Reaching and motivating contributors</td>
</tr>
<tr>
<td>4. Virtual Community of Practice</td>
<td>* Network of individuals who share research area and communicate about it online.  &lt;br&gt; * Share news, advice, techniques, pointers to other online resources.  &lt;br&gt; * Not focused on undertaking joint projects.</td>
<td>Ocean US—electronic meeting place for researchers studying oceans, with focus on US coastal waters. Uses bulletin boards / listservs, online workspace. Run by for-profit company.</td>
<td>• Usability  &lt;br&gt; • Deciding whether to emphasize asynchronous or synchronous technologies (e.g. bulletin board vs. online symposia)</td>
<td>• Maintaining energy and participation rates, esp. with changing set of participants  &lt;br&gt; • Choosing all-volunteer or for-profit management</td>
</tr>
<tr>
<td>Virtual Learning Community</td>
<td>Distributed Research Center</td>
<td>Community Infrastructure Project</td>
<td></td>
<td></td>
</tr>
<tr>
<td>---------------------------</td>
<td>-----------------------------</td>
<td>---------------------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Goal to increase knowledge of participants.</td>
<td>- Functions like university research center, but at a distance.</td>
<td>- Develops infrastructure to further work in a particular domain (e.g. software tools, standardized protocols, new scientific instruments, educational methods, etc.)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Does not necessarily conduct original research.</td>
<td>- Aggregation of scientific talents, efforts, resources beyond level of individual researchers.</td>
<td>- Often interdisciplinary.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Often formal education, but also in-service training or professional development.</td>
<td>- Unified by topic area of interest.</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Ecological Circuirty Collaboratory** - empiricists, theoreticians and their students in ecological sciences share goal of educating young ecologists with empirical research and quantitative modeling.

- Disparity in technology infrastructure across institutions.
- Software design trade-offs (1-to-many broadcast vs. small groups in parallel)
- Operating System issues

**Information and the Host Response to Injury** - medical research including hospitals, academic medical centers, informatics, statistics centers. 7 core groups: each has director, investigators, experts. Non-area experts (e.g. biologists) provide multi-disciplinary character.

- Standardization of data
- Long-distance technical support
- Technologies for workplace awareness (try to achieve convenience of collocated collaboration)

**Grid Physics Network** - plan to implement first petabyte-scale computational environment for data-intensive science. Focus on creation of tools to manage “virtual data.” Key deliverable is software package for managing virtual data. Many groups involved. Computing and storage resources distributed across the country.

- Standardizing data and data collection protocol
- Management of very large datasets
- Data provenance (keeping track of editing and transformations that have occurred on datasets)

**Key Challenges and Issues**

- Aligning educational goals and assessments so learners from multiple sites have their needs met
- Gaining and maintaining participation among diverse contributors
- Standardizing protocols over distance
- Facilitating distributed decision-making
- Providing long-distance administrative support
- Resolving cross-institutional I.P. issues
- Addressing career issues of younger participants
- Negotiating goals among disciplinary partners (e.g. asking whose agenda is paramount?)
- Choosing academic managers vs. private sector management
- Addressing career issues of young scientists who participate
Of highest interest to the work in this thesis are the technical and organizational issues listed for each of the collaboration types. These points are concise, distilled challenges that have plagued, often fatally, many of the early collaboratory attempts. Many of the issues are common to more than one collaboratory type, and often the organizational and technical issues listed for a given type are not the only ones that designers and developers will encounter.

Bos, Zimmerman, Olson and co. present an additional scheme by which to categorize the seven classifications. For each of the seven collaboratory types, they identify the type of resource to be shared (tools, information, or knowledge), and the type of activity to be performed (aggregating or co-creating). Using this information, they create a tabular representation of the seven collaboratory types, reproduced in Table 5.

<table>
<thead>
<tr>
<th>Type of Activity to be Performed</th>
<th>Type of Resource to be Shared</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Aggregating</strong></td>
<td><strong>Tools</strong> (instruments)</td>
</tr>
<tr>
<td>across distance</td>
<td>- Shared Instrument (1)</td>
</tr>
<tr>
<td>(loose coupling, often synchronous)</td>
<td></td>
</tr>
<tr>
<td><strong>Co-creating</strong></td>
<td><strong>Information</strong> (data)</td>
</tr>
<tr>
<td>across distance</td>
<td>- Community Data System (2)</td>
</tr>
<tr>
<td>(tighter coupling, often synchronous)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Infrastructure (7)</td>
</tr>
</tbody>
</table>

This breakdown is significant because it provides one relatively uncomplicated way to contextualize the wide array of collaboratory projects that have been developed in the recent past; differentiating between asynchronous and synchronous technologies, and from the sharing of tools, information, and knowledge is usually an elementary task.

Knowing the type of resource to be shared by collaboratory participants is important, because different types of sharing require “different technologies, practices, and organizational structure(s)” (Bos et al., 2007). Sharing tools or instruments might involve coordinating the use of an expensive piece of equipment, like the Keck Telescope in Hawaii mentioned in Table 4. When information is the resource, users might share data that has been gathered; when knowledge is the resource, they might share new findings, ideas, theories, or other creations.

The distinction between aggregating and co-creating, which is usually equivalent to the difference between employing asynchronous and synchronous technologies, is significant, because it addresses the types of interactions that take place between collaboratory participants. For collaboratories that are based
predominantly on aggregating activities, participants need not work as closely with their co-collaborators as they must when they are co-creating; the activities are more loosely coupled, and technologies that are asynchronous, for which events needn’t happen in a certain order, are sufficient. Often, synchronous technologies, or ones that are sensitive to the time order in which things happen, are harder to employ and support.

Bos, Zimmerman, Olson and co. assert that as one moves from the top left of the table to the bottom right, the collaboratory becomes more difficult, because sharing knowledge is more difficult than sharing information, and sharing information is more difficult than sharing physical instruments. Likewise, co-creating is more difficult than simply aggregating (Bos et al., 2007). This table, and the conclusions reached by its creators, will be readdressed later in this thesis when examining a new collaboratory construct.

One reason why collaboratories continue to interest researchers is because of their unanticipated failure rates. Very few have successfully maintained active membership; most end up effectively dead shortly after their inception. As Finholt notes, compared to the “breakout success of the Web... collaboratory use has been confined to a much smaller number of users” (Finholt, 2002). Similarly, the creators of the taxonomy of collaboratories describe the unexpected failure of collaboratories as a shock because “modern studies of science have repeatedly emphasized the social nature of scientific communities” (Bos et al., 2007).

Olson and Olson, who have extensively studied the role of distance in scientific collaborations, explain that although most attempts to use technology when collocated work is not possible fail, those that succeed result in “Herculean efforts to adjust behavior to the characteristics of the communication media” (Olson & Olson, 2000). Work is often reorganized to accommodate the constraints of the distance technologies available. User motivation is also a significant source of failure among collaboratories; first-time users see no incentive to share their ideas with a larger group if their ideas will then become common or no longer unique.

Despite the barriers that past failures pose, researchers have made recommendations and identified specific challenges that, if resolved, should make future collaborations more promising. For example, Finholt explains that because typical group work assumes a shared space, compensating for the loss of a common setting is the largest barrier to collaboratories. Therefore, it is paramount that collaboratory users be very explicit about stating information that is otherwise tacit in a collocated work situation (Finholt, 2002).

Bos, Zimmerman and co. highlight what they consider the three most critical barriers to collaboratory success. First, although information can be easily transmitted over distances, knowledge is still difficult to transfer. Second, as has always been the case, scientists work independently, and often enjoy a higher degree of independence than do other employees of corporate firms. This culture of independence poses challenges to the aggregation of scientific effort. Finally, cross-institutional work remains difficult because of administrative challenges.
Legal issues related to intellectual property are hard to resolve, and funding is often granted in ways that prevent multi-site collaboration. This effect is most visible among universities, which rarely operate in more than one state (Bos et al., 2007).

2.2.3 Future Collaboratory Design Considerations

The previous section addressed some difficult aspects of distance collaboration that will continue to test the robustness of future collaboratories. There is one additional, critical concept that should be borne in mind before endeavoring to design new collaboratory technologies that remedy the failures of collaboratories past.

The important concept to keep in mind, which has appeared as a core message in most works about the future of collaboratories, is that there are two primary characteristics of a collaboratory that must be addressed before attempting to deploy a new project, both of which heavily correlate to the collaboratory’s projected user group. These characteristics are the user group’s “collaboration readiness” and its “technology readiness.”

The collaboration readiness of a group refers to the extent to which the potential collaborators are motivated to work with one another (Finholt, 2002). Because collaboratory users can be both known and unknown to each other, the collaboration readiness in this case involves understanding the extent to which the potential group of users already collaborates. In Bos, Zimmerman and co.’s paper on collaboratory taxonomies, one of the key aspects of each collaboratory type discussed is its key organizational issues (Bos et al., 2007); this point is synonymous to the collaboration readiness required of a group before employing a given collaboratory structure.

In response to this concept of collaboration readiness, Olson and Olson provide the prescription that, in communities that do not already have a culture of collaborating and sharing, there should be no attempt to introduce groupware or remote technologies (Olson & Olson, 2000). From this, one can extrapolate that the designers of distance technologies should not deploy new collaboratories among groups of potential users who do not already collaborate with others. This assertion will be one key point of examination when a potential user community is examined later in this thesis.

As the term suggests, a group’s technology readiness refers to the presence of adequate technology infrastructure and expertise, both explicit and implicit, to support the distance collaboration (Finholt, 2002). In other words, the designer should consider the current technology use among the members of the potential community, and ask if this use has provided sufficient exposure to the technologies required to successfully use the new collaboratory. Will the users be comfortable with the technologies they will be required to use? What technological design hurdles must be overcome to employ the collaboratory in a
user-friendly way? These questions are analogous to the key technology issues addressed in the taxonomy paper (Bos et al., 2007).

Olson and Olson also provide a prescription for how to address a community’s technology readiness. They advise that advanced technologies should only be introduced in small increments (Olson & Olson, 2000). Designers should keep this step-wise progression in mind when developing new collaboratories; the technological aptitude of the user group should be analyzed and then used to gauge the probable capability for using novel applications of new technologies. This examination of technology readiness will also be used in later chapters of this thesis.
Chapter 3  The Proposed Collaboratory

The coupling of the reorganization of science as a distributed workforce with technological developments and the sudden surge in the Internet’s popularity is conducive to the birth of many new collaborative technologies. Numerous interpretations and incarnations of distance technologies appropriate to the increased distribution of scientific work have emerged since the coining of the term “collaboratory” fifteen years ago. However, there are still many avenues in the realm of collaboratories that have yet to be explored.

Specifically, there has yet to be a technology successfully deployed that allows scientists to openly share the concrete models and simulations behind their work. Until now, researchers have been left to read about the works of others, without a way to explore first-hand and build upon other researchers’ computer models. This chapter proposes a collaboratory paradigm that would allow scientists, researchers, and any other interested parties to both share their computer models and simulations over the Internet, and use and synthesize those of others to create new models.

3.1 The Theory and Motivation Behind the Collaboratory

The idea behind this collaboratory is influenced by the recent surge in popularity of community-oriented, sharing-based Web sites on the Internet. Web sites like YouTube (YouTube, 2008) and Wikipedia (Wikimedia Foundation, 2008), that allow the general public to participate by sharing videos and encyclopedic information, respectively, on publicly-viewable Web pages, are currently very high-traffic sites that attract large numbers of participants. The goal of this collaboratory is to provide a similar way for scientific researchers to share their work over an open, Internet-based platform with as much ease as is enjoyed by users of YouTube and Wikipedia.

3.1.1 A Paradigm Shift

In his publication on collaboratories, Finholt (Finholt, 2002) describes a paradigm shift that has occurred in the organization of science as an institution. Briefly described earlier in Chapter 2, the trend he identifies is science’s moving away
from the “invisible college” model that gained popularity in the 50s and 60s, in which small numbers of elite scientists conduct most of the significant scientific work being produced at a given time. Instead, he observes, science has adopted a “distributed intelligence” model, in which many scientists, in many locations, participate in scientific research, thereby distributing the work being conducted and increasing the participation of each member of the scientific workforce. Individuals’ contributions are aggregated and shared with interested members of the larger community.

This paradigm shift hints at a new mentality in the realm of scientific research, off of which this suggested new collaboratory attempts to capitalize. It is proposed here that, in addition to a shift toward a new model of scientific organization as described by Finholt, there is also an emerging new paradigm in research. This new research paradigm focuses on the open, sharing-oriented nature of Internet communities, and increasingly of academic communities. In fact, because of the current trajectory of the role of the Internet in science, academic communities are becoming more and more like Internet communities themselves.

Academic communities have always had a culture of sharing results in the form of journal publications and conferences, but it is increasingly popular, and possible, to share more than just published results. The drive to collaborate with others makes the sharing of a scientist’s original work with other scientists both possible and desirable. The open-source nature of new technologies and scientific collaboration is moving research away from being a competitive institution, and toward being one based around openly sharing work with others.

This paradigm shift toward the open sharing of research work with other scientists and researchers opens many doors for collaboratory technologies. The collaboratory proposed in this thesis capitalizes precisely on this shift, and on researchers’ understanding the necessity and benefits of sharing their work with others. This collaboratory assumes that the work scientists and researchers wish to share is in the form of computer models and simulations. The collaboratory allows researchers to publish these computerized models on the Internet through a Web interface, and then allows other researchers to run these computer models from their own computers using only a Web browser. Users can share their own models, and build on the models of other users to create new, more complex models that suit their research interests more exactly. The specific functionalities of the collaboratory will be described in more detail later in this chapter.

### 3.1.2 Existing Influences

Given the high rate of failure of collaboratory technologies, as discussed in Chapter 2, when creating this new collaboratory, it was helpful to consider Web trends that have already proven sustainable and successful. The recent successes of many existing, community-oriented Web sites have largely motivated the creation of this new collaboratory. These influential Web sites and Web
standards are, for the most part, not academic in nature; instead, they are used by the general public for the sharing of knowledge and creations with the rest of the general public. What follows is a brief description of some of these Web sites and Web practices, and how their attributes have influenced the design of this new collaboratory.

**Creative Commons**

The Creative Commons (Creative Commons, 2008b) movement was started to allow authors (of many kinds of media) to retain some rights to their work, while still encouraging others to use it and build off of it. As explained on the Creative Commons website, authors can “change their copyright terms from ‘All Rights Reserved’ to ‘Some Rights Reserved’.”

There are four different Creative Commons conditions that grant authors varying levels of ownership over their work. The “Attribution” condition allows others to “copy, distribute, display, and perform [the] copyrighted work – and derivative works based upon it – but only if they give credit the way [the authors] requests.” The “Noncommercial” condition grants others similar rights, but for noncommercial purposes only. The “No Derivative Works” condition allows others to “copy, distribute, display, and perform only verbatim copies of [the] work” but “not derivative works based upon it.” And the “Share Alike” condition allows others to “distribute works only under a license identical to the license that governs [the original] work.” Creative Commons a number of licenses that are combinations of the above conditions.

The Creative Commons movement has set a standard that recognizes that not all authors want their original work to be limited by the constraints of a traditional copyright. In fact, many authors want to contribute to a larger repository of work, to encourage the creation of similar work by fellow community members. The website of ccMixter (Creative Commons, 2008a) is one such example; community members both compose original pieces of music to share with the community, and modify existing music that has been shared with the community, to create new pieces of music of their own. There is an understanding that community members will both give (share compositions) and take (remix and reuse). The success of the ccMixter website is proof that such a community-oriented sense of sharing can be successful. It is evidence that a context in which an individual creates and then shares with a larger community is a sustainable paradigm. It is hoped that such a context can be created within this academically-oriented collaboratory, too.

**Wiki Software**

The increasingly ubiquitous “wiki” is another software technology whose use influenced the creation of this collaboratory. The website wiki.org (BoLeuf & Cunningham, 2002), which is maintained by Ward Cunningham, the creator of the first wiki, describes the wiki as “a piece of server software that allows users to freely create and edit Web page content using any Web browser. Wiki supports hyperlinks and has a simple text syntax for creating new pages and
crosslinks between internal pages on the fly.” The website also notes that wiki is “unusual among group communication mechanisms in that it allows the organization of contributions to be edited in addition to the content itself.” Wikipedia (Wikimedia Foundation, 2008) is one of the most well known wiki software implementations to date.

There are many interesting aspects of the wiki that are incorporated into the design of the proposed collaboratory. Most obviously, it is reassuring that wiki software has been successful at drawing and maintaining participants. For example, as of this writing, the English-language portion of Wikipedia contains 2,276,587 articles and has 6,648,555 registered users. 207,920,799 different edits have been made to Wikipedia’s articles by its users (Wikipedia, 2008). Users are willing to openly share their knowledge with the greater community, with few immediate benefits to themselves. Hopefully, this behavior can be replicated within the new collaboratory.

Additionally, wiki software has the ability to be downloaded, installed, and maintained by any interested user or user community. For example, the MediaWiki software package (MediaWiki, 2008) is stand-alone software that can be freely downloaded from the Internet, and installed and run by anyone who has a software server. MediaWiki is the same software package that is used by Wikipedia; it is stable, and constantly undergoing improvement. The wiki’s ability to be downloaded, installed, and maintained by any individual is another trait that is replicated in this collaboratory software; the proposed collaboratory software is intended to be downloaded and installed as a self-contained software package that can be downloaded and hosted by any interested user community with a software server.

**Web Photo Albums**

Internet users openly and publicly share media other than just information for encyclopedia articles. For example, the growth in popularity of the digital camera has created a need for an easy, Internet-based means to share digital photos; as a result, several Web sites have emerged that allow users to easily upload and publicly share digital images. The Picasa Web Albums site (Google, 2008) allows users to upload their photos into albums that are viewable as Web sites with any Internet browsers. Users can opt to install on their local computers a small application that enables drag-and-drop photo selection and then uploads the photos to Picasa’s web servers, or they can directly upload individual photos one-by-one using their Web browsers. Users are given one gigabyte of storage space on the site’s server. Once a user’s photos are saved on Picasa’s Web servers, he can manipulate his albums entirely through the Web interface. Albums have unique domain addresses, so users can easily direct friends to their albums by sending direct URL links.

Web photo albums like Picasa demonstrate that users can easily share over the Internet media that they have stored on their local computers. Furthermore, it is proof that other users can view and interact with this media through a Web interface alone, without using any additional software. A similar model is used
in the proposed model-sharing collaboratory; users' computer models and simulations are uploaded to a remote server that hosts a repository of models (like Picasa's servers, which host repositories of photos). These models can then be viewed and run by other users through a Web page, using only an Internet browser (like how Picasa photo albums are viewed by other users).

Open-Source Software

One final, significant influence on the creation of this collaboratory is open-source software (OSS). In its essence, OSS is software whose code is open and publicly available to anyone. The Open Source Initiative (OSI) (Open Source Initiative, 2008b) maintains the definition of the open-source principles, and oversees its implementations. The most notable aspect of open-source software is that the source code is openly available. Additionally, OSS licenses must also permit the free redistribution of the software, and must allow modifications and derivative works. According to the OSI website, OSS development "harnesses the power of distributed peer review and transparency of process." It also promises to yield software of "better quality, higher reliability, more flexibility, [and] lower cost."

There are many successfully OSS products that are in high-use today, including the "Apache" Web server (Apache Software Foundation, 2007), Mozilla’s "Firefox" Web browser (Mozilla, 2008), and "Linux," an OSS implementation of the UNIX operating system (Linux Online, 2008). Like Creative Commons, the OSI provides an assortment of licenses (see Open Source Initiative, 2008c) that software writers can apply to their code, each of which complies with the complete Open Source Definition (see Open Source Initiative, 2008a). Although OSS has its critics, it is hoped that the open-source mentality that has caught on among the software development community will become equally pervasive among academic communities; this mentality will help draw and retain participants in this model-sharing collaboratory, which has a similar open-source character.

3.2  The Technology

The technology behind the model-sharing collaboratory being proposed in this thesis is not new. The proposed collaboratory is, in fact, an implementation of software technology that was developed over several years in the CADlab (CADLAB, 2006), a laboratory in the Department of Mechanical Engineering at MIT. The technology, called DOME, was originally implemented as a Java-based software application installation on the user's local computer.

A Web-interface was later created for the DOME technology, through a trial implementation called PEMS Web (Sukkasi, 2007). The Web interface allows users to run DOME technology using only a Web browser. It is this Web interface that was used to create the proposed collaboratory technology in question; both the DOME software and its Web interface will be described in the sections to come.
3.2.1 An Existing Model

In his doctoral thesis, Sukkasi (Sukkasi, 2008) discusses at length the increasingly prevalent act of creating, synthesizing, and sharing multimedia on the Internet by common people, or non-experts. He proposes a model for “Commons-Oriented Information Syntheses” (“COIS”), which is made up of the sum of three activities: creation, participation, and publication. Creation activities are acts of information creation or information synthesis by non-experts; participation activities refer to participation in a community by non-experts; and publication activities refer to the publication of this created information by non-experts that contributes to the community’s commons. Sukkasi states that “the engagers do not just create. Instead, they engage in creation, participation, and publication as a whole.”

In addition to explaining the details of his COIS model, and giving several examples of existing COIS environments (Wikipedia, YouTube, and ccMixter are each considered COIS, for different reasons), Sukkasi explains in detail a “prototypical COIS environment” that he has implemented for sharing environmentally-oriented computer models and simulations. This prototype, entitled PEMS Web (Sukkasi, 2007), is the influence behind this model-sharing collaboratory; the PEMS Web application was the first to employ the technologies that are used in the model-sharing collaboratory.

The technologies employed in PEMS Web are the ones that would also be used to implement the proposed collaboratory. Tomcat Web server technology (The Apache Software Foundation, 2007) is used to host the Web application itself, and is also the server-side technology used in the proposed model-sharing collaboratory. The PEMS Web infrastructure also consists of Web-based user interfaces, both for browsing and searching through the application’s model repository, and for running an existing model. These Web interfaces are generated dynamically, based on the requests of the users. Asynchronous JavaScript and XML (AJAX) is a set of Web development techniques that allow Web-based applications to respond to user inputs asynchronously and behave more similarly to desktop applications; AJAX techniques are also used to implement PEMS Web, to allow the user interfaces to update dynamically, based on the inputs and requests of the user at a given time. Finally, PEMS Web employs a user-interface-to-model communicator, to translate information between formats that the user can understand and that the models can understand. This translation is done using the DOME infrastructure, and will be discussed in further detail in the next section. For a more detailed description of any of the technologies used to implement PEMS Web, Sukkasi’s dissertation can be consulted.

Although the technologies used to implement this collaboratory are largely the same as those used in PEMS Web, this collaboratory is novel because it is the first application of the technology for use by a specific, targeted user group. This implementation is critical to verify the efficacy of the COIS model discussed in Sukkasi’s dissertation, and to determine whether the COIS model can be used in
situations where academic material, such as computer models and simulations, are shared, synthesized, and published in a research-oriented commons.

3.2.2 DOME

DOME is the primary enabling software behind this collaboratory technology. DOME stands for Distributed Object-based Modeling Environment, and is a “modeling infrastructure that is intended to create a global community, or marketplace, of individuals offering access to simulation services related to their own specialties, much as the WWW has enabled world-wide access to information” (CADLAB, 2003). DOME was created as a software package to facilitate product design, by providing a means for integrating software models.

The DOME software facilitates real-time communication between multiple computer-based design models and simulations (Borland & Wallace, 1999). DOME’s many features provide flexible options for users to create and integrate models. For example, users can build object-oriented models that can handle both discrete and continuous variable types, or can utilize any of a selection of sub-models from a provided catalog. These capabilities are the supporting fundamentals behind the ability to build and publish models on the collaboratory over the Internet.

The DOME client application is a program in DOME that allows models hosted on a DOME server to be accessed and run over the Internet. This client application has an application programming interface, or API, that provides a way to interact with the DOME software without using the graphical user interface. Instead, users can write code to interact with the DOME application. It is through this API that the Web code interacts with the DOME software (Sukkasi, 2007). More details about the DOME server technology that supports the use of DOME and DOME models over the Internet can be found in Sukkasi’s thesis.

3.3 The Implementation

The proposed model-sharing collaboratory was designed to allow users to share their research, beyond just their results in textual form (as is currently the norm, via journal articles and other scientific publications). The proposed collaboratory allows scientists and researchers to share over the Internet the actual computer models and simulations that they create as part of their research, with other scientists and researchers who have interests similar to their own. Sharing actual work, like computer models and simulations, will increase the efficiency of scientific research, because researchers can build directly on the work of others, without having to first replicate work that has already been done.
3.3.1 The Collaboratory Definition, Goals, and Key Concepts

The proposed collaboratory was created in the context of the increasing distribution of scientific intellect and intelligence discussed at the start of this chapter. As discussed in Chapter 2, many collaboratory technologies have emerged since the mid-1990s, the majority of which have not survived. The collaboratory in question hopes to both capitalize on the failures of past collaboratories, and on the successes of other, current Internet phenomena, to be more successful and sustainable in the long term.

**Top-Level Description**

The proposed model-sharing collaboratory is a Web application, which is ideally public, and is intended for geographically-distributed users who are both known and unknown to each other, who share common interests (e.g. research areas). Users contribute to the collaboratory both by sharing their original work, and by using, creating, and building off of others' work. The Web application is, therefore, an active repository of original work (in the form of models, simulations, and relevant publications) and the nesting place of new, synthesized work. It also acts as a gathering place for its users to hold discussions about the models and simulations published in the collaboratory. The proposed collaboratory software can be downloaded, installed, and maintained by the users themselves. The concept of open-source sharing, using, and contributing is a strong driving factor in the use of the collaboratory.

**Capabilities**

The proposed collaboratory has been designed with many capabilities, to fulfill a number of functions that are desirable for distributed research collaboration. A brief listing of the collaboratory's capabilities are below; more detail will be given in the *Description of Features* section that follows.

The collaboratory:

- Allows users to publish models and simulations, alongside their research papers.
- Supports the publication of many types of computer models and simulations (e.g. Excel models, Matlab models) and can be configured to support many more. Allows both open-source sharing and more hidden publication methods.
- Allows users to synthesize multiple models to build more complex ones that they can then use, and share (publish).
- Supports discussion between members via bulletin boards that are linked directly to specific models.

**Goals**

The proposed collaboratory was created with many goals in mind, reflecting what is known about past collaboratory successes (and failures), and the needs of scientists participating in distance collaborations. These goals include:
• To strengthen ties and links within the research community (through increased awareness of what others are doing, and increased interaction between community members).
• To encourage open sharing and co-creating with one's fellow scientists and researchers.
• To foster group independence.
• To facilitate communication between researchers who have a common passion, whether they know each other or not.

3.3.2 Description of Features

The proposed collaboratory's many features fulfill the capabilities and goals of the Web application that are discussed above. The following descriptions explain the specific characteristics of this model-sharing collaboratory, why they were chosen, and the thought supporting their implementations. Many of the features that are described have been implemented; however, some are higher-level attributes that have yet to be realized at the time of this writing, and will need to be implemented before the collaboratory is deployed.

Software Installation and Hosting

The proposed collaboratory software will be available as a stand-alone software package that can be freely downloaded from the Internet, similarly to how a wiki software package can be downloaded presently. Any interested research community can freely download and install the software, provided that they have a computer server available to host the website. It is suggested that the community have a designated webmaster with sufficient server technology knowledge that he or she can easily maintain the server and user database.

Publication Coupling

The proposed collaboratory is designed to go hand-in-hand with existing research communities; because such subject-specific communities already exist within academic journals, it is intended that the collaboratory first be deployed within the readers of and contributors to an academic journal. Given this tie to an existing publication, it is important to consider how academic journals currently disseminate their publications.

Because almost all journals now publish electronically (either in addition to paper publication, or exclusively), it is ideal if the collaboratory can combine with the website that currently provides the online publications. The number of websites that the subscriber must visit to interact with the publication should be kept to a minimum. Therefore, in addition to hosting the user's models and simulations, the collaboratory will handle the periodic electronic publications of the journal.
Figure 1 shows the homepage of the pilot collaboratory. The website, which was made for the Journal of Industrial Ecology (JIE) (MIT Press, 2008), lists all of the journal’s publications to date, and includes a link to each separate publication. This way, users can immediately select which journal to browse, even if they intend only to look through the journal’s articles. They need not interact with any of the models included in the collaboratory.

**Article View**

The proposed collaboratory’s homepage, shown in Figure 1, includes two tabs toward the top of the page. These tables allow users to choose one of two views through which they will interact with the website: “article view” or “model view.” The article view is closest to the standard electronic representation of a journal publication; it lists publications by date, and articles within the publication by topic.

When using “article view,” the user can select the publication issue of interest to him. Upon selecting the issue, he is taken to an issue-specific page that lists all of the articles available in the issue. This is shown in Figure 2.
Figure 2: The issue page in "article view"

Users can select an article they wish to view, as they would normally do when browsing an electronic journal. Instead of directly receiving a portable document format (PDF) file containing the article they have selected, users are taken to an article-specific Web page that contains more information about the article. It is at this level that users first notice the novel capabilities of the collaboratory.
The Changing Metabolism of Cities
Spring 2007, Vol. 11, No. 2, Pages 43-59

Christopher Kennedy, associate professor in the Department of Civil Engineering at the University of Toronto.
John Cudliffe, former graduate student in the Department of Civil Engineering at the University of Toronto.
Joshua Engel-Yan, former graduate student in the Department of Civil Engineering at the University of Toronto.

Abstract
Data from urban metabolism studies from eight metropolitan regions across five continents, conducted in various years since 1965, are assembled in consistent units and compared. Together with studies of water, materials, energy, and nutrient flows from additional cities, the comparison provides insights into the changing metabolism of cities. Most cities studied exhibit increasing per capita metabolism with respect to water, wastewater, energy, and materials, although some cities showed increasing efficiency for energy and water over the 1990s. Changes in solid waste streams and air pollutant emissions are mixed.

The review also identifies metabolic processes that threaten the sustainability of cities. These include altered groundwater levels, exhaustion of local materials, accumulation of toxic materials, summer heat islands, and irregular accumulation of nutrients. Beyond concerns over the sheer magnitudes of resource flows into cities, an understanding of these accumulation or storage processes in the urban metabolism is critical. Growth, which is inherently part of metabolism, causes changes in water stored in urban aquifers, materials in the building stock, heat stored in the urban canopy layer, and potentially useful materials in urban waste dumps.

Practical reasons exist for understanding urban metabolism. The vitality of cities depends on spatial relationships with surrounding hinterlands and global resource webs. Increasing metabolism implies greater loss of farmlands, forests, and species diversity; plus more traffic and more pollution. Urban policy makers should consider to what extent their nearest resources are close to exhaustion and, if necessary, appropriate strategies to avert exhaustion. It is apparent from this review that metabolism data have been established for only a few cities worldwide, and interpretation issues exist due to lack of common conventions. Further urban metabolism studies are required.

Figure 3: The article-specific Web page

Figure 3 shows that the article-specific Web page contains publication information about the article, such as the author(s) and publication date. It also contains the article’s abstract, if one is available, and a .pdf file of the article’s full text. Finally, the Web page includes new functions that are specific to the models that have been published in conjunction with the article. Users can choose to try any of the article’s published models, or enter a discussion forum specific to the article in question.
When the user selects a model to try, a model-specific Web page is generated for the model. As seen in Figure 4, input fields are created for the different input variables that the user can change. When the user has inputted his desired values, he can elect to "run this model!" by clicking the run button. After the model, which is stored on the collaboratory's server, has been run, the results are returned to the user in the output variable fields.

When the user scrolls down on the model page, he can see other features that are related to the model. Figure 5 shows that the reader can also access and participate in a discussion forum specific to this model. He can also see what
keywords have been used to tag this model, view any other models with a common tag, and add a new tag himself. Finally, he can see child models that have been made using this model as a starting point. He can opt to view any of these child models.

If the user chooses to view the discussion forum, he is taken to a discussion page that is specific to the model, like the one shown in Figure 6. This specificity allows users to freely discuss any aspect of a model, with minimal confusion. (For example, it is clear which model is the model in question.)


Figure 6: The model-specific discussion forum page

**Model View**

Recall from the homepage (Figure 1) that there are two view options: “article view” and “model view.” The model view provides users a less-conventional means for interacting with the collaboratory. This view is more appropriate if the user is interested in directly browsing the different models that have been published in the collaboratory, instead of seeing which ones are related to which articles. If this is the case, he selects the “model view” tab at the top of the homepage. The page is then reloaded, from the model perspective, as shown in Figure 7.
Journal of Industrial Ecology

This area contains a collection of all the models publicly available that accompany articles published in the Journal of Industrial Ecology. Here, you can browse the models, search for a particular model, or integrate two or more models to develop a new model that suits your interests.

- search database: search all available models by keyword
- see all models: view a list of all available models
- integrate!: integrate two or more models!

Figure 7: The model view of the collaboratory

The model view gives users three options for looking for models. Depending on how specifically the user knows what he is looking for, he can choose any of the three options. If he elects to “search database,” he is taken to a search page that searches the model repository for models that match his search criteria.

Figure 8 shows the search page, which provides options to perform both a basic search and an advanced search. The advanced search is more appropriate for users who have specific model characteristics in mind. The basic search is better suited for those users who only have a general idea of what they are looking for, or who are looking for a broader list of results.
Journal of Industrial Ecology

Figure 8: The model search page

Had the user chosen the “see all models” option from the model view main page (Figure 7), he would have been taken to a list of all the models in the model repository, as seen in Figure 9. He then has many options for sorting the list of models; he may sort by name (both ascending and descending), chronologically (both oldest and most recent first), or by contributor.
Finally, the user has the option to integrate one or more models that are already published in the collaboratory, and to create his own model. He can do this by selecting the “integrate” button in the model view main page, and then selecting from which model(s) he would like to build. He can synthesize a new model by linking the output parameters of one existing model to the input parameters of another existing model, or by using mathematical operators on existing parameters to create new parameters (either input or output).

Once he has created his new, integrated model, he can save it within the collaboratory’s model repository, and publish it in the collaboratory Web site. It will then appear like all other existing models in the repository when future users search for models, and have its own page when users opt to try it. Additionally, this integrated model will appear as a “child” of the base model, on the original model’s page.
Figure 10: The integration option for synthesizing new models

Navigation
In both the article view and model view interfaces of the Web site, navigation of the site is facilitated by the use of a breadcrumb trail at the top of the page. The breadcrumb trail keeps track of how many levels into the Web site the user is currently located, and allows him to back up to any higher level at any time. The breadcrumb trail is shown in more detail in Figure 11.

Figure 11: The breadcrumb navigation trail leading to a model page

The features of the collaboratory that have been discussed in this section will be mentioned again in the following section, when specific use scenarios and potential user groups are described.

3.4 The Use
As is the case with any collaboratory technology, or any new technology in general, the technology was designed and implemented with specific users in mind. In the case of this collaboratory, the intended users are members of communities that already exist around academic publications. It is assumed that
these users have similar interests (either professionally or casually), because academic publications typically have a topical focus. Nonetheless, the use of the collaboratory technology is certainly not limited to such journal-oriented user groups, and may be beneficial to other communities as well. The following sections address potential user groups of the collaboratory, and general use scenarios that might be encountered by users of the collaboratory.

3.4.1 Potential User Groups

The pilot implementation of the proposed model-sharing collaboratory was designed for a community that revolves around an active academic publication, the Journal of Industrial Ecology (Yale University, 2008). The next chapter will discuss, in detail, why this particular academic community was targeted as a pilot. However, the collaboratory can be successful with user groups other than the specific one targeted by this pilot. The following discussion addresses some more general categories of users that could benefit from this collaboratory technology.

Any journal publication that is interested in expanding the quality of research it is able to share through its publication may benefit from this collaboratory technology. One novel feature of the collaboratory is that it allows media, other than just text, to be published as part of its publication. Therefore, scientists and researchers can easily publish their computer models and simulations in conjunction with the papers that report on their findings. These multimedia publications would allow others interested in making advances in similar fields to investigate first-hand what people with similar interests have been working on. Ideally, in such a situation, future collaborations or partnerships will be established between interested parties, making research more efficient and more productive. These users need not be members of the same sector; for example, scientists in research labs may discover that they have similar research interests to workers at a corporate company, and that their research may be beneficial to practical applications that are currently in development.

This proposed collaboratory can also be useful to researchers within a scope smaller than an entire academic community. For example, labs within the same department or departments within an institution could use the collaboratory in a way similar to its intended use for academic publication; they could still summarize their research findings in text form, and publish accompanying computer simulations for colleagues to use at their leisure. Similarly, scientists and researchers at private research firms might find an application of the collaboratory technology useful within their company, as might members of professional societies. The Journal of Industrial Ecology was chosen over these other, more focused research communities, because it encompasses a broader scope of readers and authors, from many sectors; it was hoped that the pilot would have greater potential.
A less obvious application for the proposed model-sharing collaboratory technology is within the classroom. Since the technology allows users to share text and computer models side-by-side, the collaboratory could be a useful tool for teaching students how scientists use computer models. Students could learn about computer modeling, and write a small simulation for one aspect of the class-wide project. Then, all the models could be published in the collaboratory, and students could link the models to create one larger, more complex simulation. The ability to create a large simulation would teach valuable lessons in working with constraints and approximations to replicate natural events. Furthermore, because the collaboratory is accessible over the Internet, students would be free to explore the use of the models outside the classroom; this ability would hopefully encourage them to explore the capabilities of scientific modeling beyond their classroom experiences.

3.4.2 General Use Scenario

There are two general approaches to the use of this new model-sharing collaboratory, each of which will be described below. The first use scenario is from the perspective of a reader (and perhaps past author) who browses the contents of the journal in question each time a new issue is published. The second use scenario is from the perspective of a researcher who is interested in finding a particular model hosted by the collaboratory, for use with his work.

Scenario 1: The reader

Recall that this collaboratory is associated with a research publication (a journal). The user is a regular reader of the journal, who, in the past, has also published his findings in the journal. Since he reads this journal to keep current with the research in his field, he desires to browse this issue's publications online.

He visits the journal's homepage, which is part of the collaboratory. There, he is able to browse the most recent online publication. By selecting the most recent publication, he is taken to a Web page that lists the contents of the most current publication, including the titles and authors of each submission. The user selects an article of interest to him.

By selecting a particular article, the user is brought to a "homepage" for the specific article. At this page, the user can learn more about the article, including details about the article's authors, read the abstract, or download a PDF of the entire article. The collaboratory also allows the user to access novel new features previously unavailable in electronic publications. The user can view and use computer models discussed in the paper and uploaded by the paper's author(s), or enter a dynamic discussion page with other readers of the article or other users of the article's model(s).

After browsing the contents of the paper to better gauge its subject matter, the user browses the models and simulations that have been made available by the article's authors.
The user can select any of the models; after choosing one, he is taken to a page that is dynamically generated for the model he has selected. At the model-specific Web page, the user can change the model's input parameters as applicable, and run the model to see how his variables affect the model's outputs. He can change the input parameters as often as he wants, allowing him to experience first-hand the model he has read about in the paper, and to "get a feel for" the behaviors encapsulated in the model. The user can run as many different models as he would like, each time receiving the resulting outputs directly on his screen.

If the user has some feedback about the model, or has an idea for a discussion point pertaining to the model or the article that it accompanies, he can opt to contribute to the ongoing discussion forum specific to this model. At this discussion page, he can see what others have thought about the article and/or its models, respond to their comments, and leave his own feedback for others to read.

**Scenario 2: The model-seeker**

This time, the user is a researcher who is interested in finding a robust computer model that has already been created, tested, and published online, to integrate into his own computer simulation. He is concerned not with reading the literature associated with the model, but with finding a model appropriate for his experiments. For his search, he uses the collaboratory's "model view".

When he enters the model view, he has three options available to him: a search, a list of all the models published in the collaboratory, and an option to integrate existing models. Since he wants to first gauge what is available to him, he opts to browse through all the available models. Choosing this option brings him to a list of all the models available in the collaboratory, which can be sorted by name, date, and contributor. When he finds a model that interests him, he can select it and be taken directly to the model's Web page.

If instead he decides that he wants to perform a search to find a smaller subset of models that might be relevant to his work, he can return to the model view homepage, and this time choose to search the model database. This search Web page allows him to choose between a basic search or an advanced search. Regardless of the type of search he chooses, his query will generate a list of relevant model hits on a new page.

Finally, if the user decides that no one available model is adequate for his work, or that he would like to add functionality to an existing model, he can choose to integrate models that are already available in the database. Selecting this option from the model view homepage takes him to a dynamic Web page that walks him through the steps of integrating two or more models with similar parameters. At the end of the integration process, in addition to running his new model, he will have the option of publishing it in the collaboratory, too, so that future users can take advantage of his newly synthesized model.
3.5 In the Context of the Collaboratory Taxonomy

The background on collaboratories that is provided in Chapter 2 discusses a taxonomy system identified by researchers as part of the Science of Collaboratories (SOC) project. The taxonomy breakdown, which was presented in Table 4, categorizes existing collaboratories into one of seven groups, each of which has different characteristics, technological issues, and organizational issues. This collaboratory is unique, because the technology employed by this collaboratory is unlike that used by any of the collaborative types defined in the SOC’s taxonomy. In fact, this collaboratory clearly does not fit into any of the seven types of collaboratories described by the SOC. The following sections will revisit the seven types of collaboratories described by the SOC, explain why their categorization system is not appropriate for the collaboratory in question, and propose how this new model-sharing collaboratory would be defined according to the SOC’s criteria.

3.5.1 The Seven Taxonomies, Revisited

The Science of Collaboratories (SOC) project identified seven types of collaboratories. Please refer to Table 4 for a detailed description of each of the seven collaboratory types and their characteristics. The following sections will refer to each of the seven types by name, and briefly describe the similarities of each type with the new collaboratory introduced in this thesis. The main differences between each type and this collaboratory will also be discussed, to demonstrate that no one category alone is sufficient to describe this new collaboratory technology.

Shared Instrument

The “shared instrument” type of collaboratory exists to allow increased access to scientific instruments that are typically expensive to own and operate, and/or are located in remote areas. These types of collaboratories require having available very thorough and reliable technical support, since most users are not collocated with the instrument.

A key similarity between the proposed model-sharing collaboratory and collaboratories of the shared instrument type is the concern surrounding remote social relationships. For example, the users of a shared instrument collaboratory must be trusting of the technicians who maintain the instrument remotely, and in some cases set up the user’s experiments for them. Although they will probably never meet in person, the user relies heavily on the technician. Similarly, in the proposed collaboratory, trust is required to share ones work with a group, especially when the group consists largely of people one has never met or whose identities one might not even know. Developing this sort of trust requires establishing remote relationships with ones peers, similarly to the remote relationships required for shared instrument collaboratories.
There are obvious differences between the shared instrument collaboratory and the model-sharing collaboratory. The main goal of the shared instrument collaboratory is to physically allow the use of a remotely located tool. This model-sharing collaboratory has no such concept of remote access. Nor does this collaboratory have to deal with the synchronous communication issues that often challenge developers of shared access collaboratories.

**Community Data Systems**

The "community data systems" type of collaboratory is effectively a large, semi-public database of information that is created, maintained, and continually improved by a geographically distributed community. Its primary technical concerns are with large-scale data standardization, and its primary organizational issues involve figuring out how to motivate contributors to contribute to a repository of public goods.

The community data systems type of collaboratory is similar to the proposed collaboratory because both are self-maintained, by members of a geographically distributed community. Both will encounter similar issues related to handling organizational issues remotely. Also, they are both semi-public and open to people with interests similar to one another. And both must motivate users to contribute to a public good.

One significant difference between the community data systems type of collaboratory and the one in question is that the community data systems type is a repository of information; data is put in the collaboratory to be accessed. However, in the model-sharing collaboratory, information and simulations are continually revised, reworked, and improved by any or all of the collaboratory’s participants. Information stored in the collaboratory is not static; rather, it is very dynamic and constantly evolving.

**Open Community Contribution System**

The "open community contribution system" type of collaboratory is typically an open project that accepts contributions in the form of work, not data. This type of collaboratory aggregates the efforts of many geographically distributed individuals toward a common research problem. These types of collaboratories usually involve large, distributed user groups, so quality control maintenance must be done on a large scale. Similarly, contributors must be continually found and motivated.

The open community contribution system is similar to the proposed model-sharing collaboratory because it is an open effort. In both types of collaboratories, anyone is invited to participate. Further, both types of collaboratories accept contributions in the form of work; the model-sharing type of collaboratory accepts contributor’s new computer models and simulations, or ones they have synthesized from existing models. Also, both types of collaboratories involve a geographically-distributed community of users. As a result, both types will have similar challenges dealing with a diverse, distributed audience.
The two types of collaboratories are different, however, because they cater to different user groups. The open community contribution system type of collaboratory has two distinct types of users: the users who provide data samples, and those who use or analyze the data samples. Simply put, one group of users are the “testers,” and the other group of users are the “scientists.” This type of user distinction does not exist in the model-sharing collaboratory; all users are allowed to, and even expected to, both share original work and synthesize that of others. There is no distinction between those who provide the work and those who use it; they are one in the same. Furthermore, although both types of collaboratories involve the sharing of work, the model-sharing type of collaboratory involves the sharing of entire computer models and simulations, not just data. Although just the output data from the models and simulations could be provided (equivalent to what is shared in the open community contribution system type), the model-sharing collaboratory goes one step further and shares the original simulation itself.

**Virtual Community of Practice**

The “virtual community of practice” type of collaboratory is typically a network of individuals with a common research interest, who communicate about their interest online. Participants share news, advice, techniques, and links to other resources. However, the collaboratory is not focused on undertaking joint projects. The collaboratory may encounter usability challenges, and must strive to keep users’ enthusiasm and participation rates high.

The most obvious similarity between the virtual community of practice collaboratory and the proposed model-sharing collaboratory is the fact that both involve geographically distributed groups with similar research interests, whose members communicate through Web-based means. For example, the members of an academic journal may participate in a virtual community of practice collaboratory by sharing information related to their research, news, upcoming conferences, or updates about the progress of their personal work.

However, the two types of collaboratories also differ on a fundamental level. In a virtual community of practice, no actual sharing of work takes place. The discussion in Chapter 2 about the types of resources shared in different types of collaboratories differentiates between sharing information (data) and sharing knowledge (new findings); the sharing of knowledge is considered the more difficult of the two. The virtual community of practice type of collaboratory facilitates the sharing of information only, whereas the model-sharing type of collaboratory promotes the sharing of knowledge, in the form of original work. As a result of this difference, users will participate in each of the collaboratories for different reasons, and with different intentions. The virtual community of practice type of collaboratory is not a substitute for the model-sharing type of collaboratory, because no actual knowledge synthesis can take place in the virtual community of practice.
Virtual Learning Community

The "virtual learning community" type of collaboratory aims to increase the knowledge of its participants; however, it does not necessarily involve conducting original research. Rather, this type of collaboratory offers formal education, in-service training, or professional development to its users. Its primary organizational challenges involve simultaneously meeting the needs of many learners from multiple sites. Technological issues often arise because additional software is used to implement the collaboratory, requiring compatibility with different infrastructures and operating systems.

There is little overlap between the virtual learning community type of collaboratory and the proposed model-sharing type of collaboratory. The goals of each are distinctly different. Nonetheless, each collaboratory exists with the goal of enriching the knowledge of its users.

The technological issues that arise in the virtual learning communities should not appear in the model-sharing type of collaboratory. In fact, the technological issues of the virtual learning communities are good examples of why the model-sharing collaboratory was designed as a Web application; applications that require the installation of separate software are almost always accompanied by operating system compatibility problems, maintenance problems, or general use problems. Web applications, like the model-sharing collaboratory, can be run from any computer, over any Web browser. The fact that no additional software is needed should eliminate a significant portion of the technological issues encountered by other types of collaboratories, like the virtual learning communities.

Distributed Research Center

The "distributed research center" type of collaboratory functions like a university research center, but with geographically distributed participants. It aims to aggregate scientific talents, efforts, and resources, beyond levels possible by individual researchers. This type of collaboratory has a unified area of interest, and involves a large amount of human-to-human communication. It requires technologies for workplace awareness, to try to replicate the convenience of collocated collaboration, and a strong network of long-distance technical support. The distributed research center is the most organizationally-ambitious collaboratory, for it must gain and maintain participation of a diverse group of contributors, must facilitate geographically distributed decision-making, and must address legal issues like cross-institutional intellectual property rights.

The distributed research center type of collaboratory is essentially a traditional collaborative research effort, just at a distance. It is, perhaps, the collaboratory closest in implementation to Wulf's original definition of a collaboratory. This type of collaboratory is also probably the most similar to the model-sharing collaboratory that is the focus of this thesis. Both are "research efforts" with a common goal among their participants.
However, the distributed research center type of collaboratory differs from the proposed model-sharing type because in the distributed research center, all the details are planned from the get-go. Collaborations are carefully calculated in advance, and users' participations are not only anticipated, but expected. The collaboratory is not an environment to “casually” or spontaneously share computer models and simulations, for others to use as they see fit. Compared to the distributed research center, the model-sharing type of collaboratory seems more like a repository of resources available for use as needed, that is very dynamic and constantly changing. Furthermore, at a given time in the model-sharing collaboratory, there is not necessarily any active collaboration on a common research project.

**Community Infrastructure Project**

The “community infrastructure project” type of collaboratory is an effort to further work in a particular domain. For example, this type of collaboratory might work toward developing software tools, standardized protocols, new scientific instruments, or new educational methods. This type of collaboratory is often interdisciplinary, and accordingly, requires the organizational infrastructure necessary to negotiate goals between its various disciplinary partners.

The community infrastructure project works to develop software and digital infrastructure; this is similar to the proposed model-sharing collaboratory in question, because it also deals with aspects of software. Code and models from various disciplines might be shared in the collaboratory, to be combined with others or to be built upon. The collaboratory creates new models, simulations, and code resources and infrastructures for its community to use, which is technologically similar to the community infrastructure project type of collaboratory.

However, the two types of collaboratories differ because the community infrastructure project is more like a meta-collaboratory or a home to meta-research. Community infrastructure projects share and develop code in preparation for a project or final implementation. However, in the model-sharing type of collaboratory, code and simulations complement individuals’ pre-existing research projects; it is this actual research code that is shared and built upon. Compared to the community infrastructure projects, the model-sharing collaboratory can be considered abstracted by one fewer level.

**3.5.2 The Activity and Resource Grid, Revisited**

Recall from Chapter 2 that, as part of its taxonomic categorization of existing collaboratories, the SOC created a table that attempts to differentiate between the types of collaboratories based on the dominant resources and activities involved. This chart is shown in Table 5. On one axis of the chart are the types of resources shared in the collaboratory: tools (instruments), information (data), and knowledge (new findings). The opposing axis represents the type of activity
performed in the collaboratory: aggregating (which involves asynchronous technologies and work that is loosely coupled), and co-creating (which involves synchronous technologies and work that is tightly coupled).

The resulting grid of six combinations is used to differentiate the seven types of collaboratories. There are some shortcomings to this classification scheme, however. The SOC notes that the categorization is only an approximation, and is made based on the dominant types of resources and activities of each collaboratory type, not the only types of resources shared and activities performed in each type of collaboratory. A given collaboratory may share more than one type of resource, or perform both aggregation and co-creation activities, but an assumption is made that only one of the two is dominant.

This approximation is not appropriate for the proposed model-sharing collaboratory in question. As has already been discussed, no single type of the seven predefined collaboratories encompasses all of the features of this new collaboratory; likewise, the classification table shown in Table 5 is not at all appropriate for this collaboratory, for neither its resources shared nor its activities performed can be whittled down to a single attribute. The following discussion explains why the SOC’s table is not appropriate for this new model-sharing collaboratory; in essence, the new collaboratory transcends the assumed categorization criteria inherent in the table’s construction.

### Activities Performed

The SOC’s grid (Table 5) assumes that only one of two types of activities is performed by a given collaboratory: either aggregation activities or co-creation activities. The table does not accommodate a collaboratory that supports both types of activities to equal degrees, which is the case for this new model-sharing collaboratory.

Aggregation takes place in the proposed model-sharing collaboratory when users share their existing simulations, models, code, journal articles, appendices, and other media that may be relevant to the research that they have already conducted. The collaboratory serves as a central place of aggregation for all of the users’ already-completed work. The aggregate work hosted by the collaboratory from the get-go is the fundamental building block of all subsequent activity in the collaboratory.

Co-creation in the collaboratory takes place based on the aggregated media shared by the users at the start. The shared models, simulations, code, and articles are used to create and synthesize new models, simulations, code, and even research projects, all within the collaboratory. Other users co-create by adding their insights, experiences, and efforts to the aggregated media.

The SOC asserts that it is “generally more difficult to co-create than to aggregate” (Bos et al., 2007). If this assertion is true, then it is probably harder still to achieve both aggregation and co-creation to similar extents, simultaneously, within a single collaboratory. Then again, it may not be more challenging, because having
a base aggregate may provide a strong set of starting resources with which to co-create; such a base aggregate is something missing from the existing co-creating-type collaboratories.

**Resources Shared**
The SOC's grid (Table 5) assumes that each type of collaboratory shares one of three types of resources: tools (instruments), information (data), or knowledge (new findings). The table does not accommodate a collaboratory that supports two or more types of resource sharing to similar degrees. On an abstract level, the proposed model-sharing collaboratory supports sharing of all three types of resources.

Tools are shared within the collaboratory, because different scientists' and researchers' models can be considered tools to others. After they are shared, these models and simulations are used by other scientists and resources to build additional features into their own models and simulations. The tools of others provide scientists with capabilities and know-how that they would not otherwise have had on their own.

Information, both direct and indirect, is embedded in the models and simulations themselves. Users should be able to analyze the models and simulations of others, to extract information about what work has already been done in a given area of interest. They can obtain information on how systems work, see who has already worked on what types of research, observe how these researchers have approached the work, and deduce why it was successful.

Finally, new findings are shared because the models themselves are new findings being shared with the rest of the community, similar to the distributed research center type of collaboratory. New findings can be shared within this collaboratory, and new results can be borne from the collaboratory via the assembly and synthesis of multiple published works (where the works may be models, simulations, papers, etc.).

Given the fact that the proposed model-sharing collaboratory could physically be placed atop the existing 6-cell grid proposed by the SOC, it is reasonable to assert that the chart of resources and activities is not applicable in this situation. Instead, an eighth, more all-encompassing collaboratory type will be defined.

### 3.5.3 An SOC-like Definition

It has been established in the last few sections that the SOC's previously defined typologies for existing collaboratories are not appropriate for this new model-sharing collaboratory. Since the proposed collaboratory fits into none of the seven taxonomy categories, an eighth will now be defined that would best fit this new collaboratory. The same criteria as used by the SOC will be considered in defining the type's properties. Think of the following sub-sections as an eighth row of Table 4.
Definition
Many traits define collaboratories of this type. In particular, this collaboratory type:

- Is a public web application for geographically-distributed users with a common interest (e.g. a common research area).
- Hosts Web-interfaces for models and simulations.
- Allows users to contribute by publishing their own models and simulations, by using those of others and providing feedback, and by building on those of others to make and publish new models and simulations.
- Acts as a gathering place for users with common interests to discuss models, simulations, and articles published within the collaboratory.
- Is driven by the open-source concept.
- Is created, hosted, and maintained by its users.

Technical Issues
In addition to the technical issues that plague all collaboratories, many of which are mentioned in Table 4, other technical issues include:

- Its usability: maximizing ease-of-use is a priority.
- Its ease to install and maintain by users.
- Making it a completely Web-based, browser-neutral and platform-neutral Web application.
- Handling asynchronous communication technologies that can support both aggregating and co-creating.

Organizational Issues
In addition to the organizational issues that plague all collaboratories, many of which are mentioned in Table 4, other organizational issues include:

- Public goods – how to motivate contributors? This is similar to the “community data systems” type of collaboratory.
- Intellectual property issues – who retains what rights?
- How to maintain energy and participation rates of participants, so that they actively share, use, and build? This is similar to challenges of the “open community contribution system,” “virtual community of practice,” and “distributed research center” types of collaboratories.
- How to maintain neutrality toward users’ professions? This collaboratory serves researchers, workers in industry, and members of many other sectors. It must be usable and understandable by all, and must foster cross-sector communication by all.

The seven types of collaboratories that the SOC have defined in their taxonomy scheme may currently be inclusive of the variety of collaboratories that exist. However, the classification scheme will not remain complete or exhaustive forever. New technologies and organizational paradigms will force additions to and rethinking of, the system. The addition of this eighth collaboratory type is
just one such example of changes that will need to be made in the near future. The dynamic nature of collaboratories, and the technologies that support them, will ensure that researchers will not lose sight of this field of study.
Chapter 4 The User Group

The goal of this thesis is to identify a test user group for which this new collaboratory should be both useful and successful. Because the collaboratory must be appropriate for the user community, there is little value in trying to deploy the new technology among a poorly-matched community; it will not take off. Past research has examined what factors help predict whether a given user community and a given collaboratory will be compatible with each other.

This chapter addresses these factors, to justify why this new collaboratory is appropriate for the Journal of Industrial Ecology (JIE) community, and likewise, why the JIE community may be likely to adopt the new collaboratory technology. Note that the Journal of Industrial Ecology is the official publication of the International Society for Industrial Ecology (International Society for Industrial Ecology, 2008); for the purpose of this study, the JIE community is comprised of both the readers of the articles and the contributors to the journal (i.e. the articles' authors).

4.1 The Importance of the User Group

Before deploying a new distance technology, it is critical to both have a concretely-defined user group, and to understand the members of the user community. As previously mentioned in Chapter 2, Olson and Olson (2000) raise many points relevant to the difficulties of introducing new distance technologies. They explain that the acceptance of a new technology is influenced by many characteristics of both the users and the group as an entity. Specifically, it is important to gauge if the user group has sufficient “collaboration readiness” and “technology readiness” to adopt the new distance technology; both of these terms were defined and discussed in Section 2.2.3.

Given the importance of understanding the user group in question, it is crucial that the test group for this collaboratory be quantified; exactly who the users are should be determined. The subsequent sections of this chapter will, therefore, provide an overall, largely quantitative survey of the make-up of the Journal of Industrial Ecology community. The survey will attempt to answer the following questions:
• What percentage of the articles are co-authored? (Co-authorship can be used as a loose measurement of collaboration.)
• How diverse or geographically distributed are the collaborations that exist? How many of the co-authored articles are co-authored by people in different departments, or at different institutions? Are any international?
• How diverse is the community as a whole? How many countries are represented? With what frequency?
• What sectors are represented by the journal’s community members? How many are from universities? Research labs? Industry?

Answering the above questions will help draw two fundamental conclusions. First, the answers will help determine how much this particular community collaborates already, and to what extent. In other words, they will help determine if the community has sufficient “collaboration readiness”. Secondly, the answers will help judge how open the community members are to using technology. Do they have sufficient “technology readiness”?

Ultimately, the answers to these questions, coupled with additional observations, will help to determine if such a collaboratory tool is appropriate for this user group. Only if the conclusion is affirmative should the pilot application be deployed within the community.

4.2 The Test User Group: The Journal of Industrial Ecology

As mentioned above, the proposed user group for this new collaboratory is the Journal of Industrial Ecology community, and it is assumed that the community is made up of two fundamental parties: the contributors (the authors of the articles), and the readers. There may be overlap between the two communities (i.e. authors of the articles may also make up the readership), but the two groups will be analyzed as if mutually independent, with little detriment to the results. The following sections will discuss in more detail the history of the journal (via the history of the field of industrial ecology), how the community members identify themselves, their motivation, and the reasons why the community was chosen as a potentially successful user group with which to test this pilot Web application.

4.2.1 Description of the JIE

Background
The Journal of Industrial Ecology is the official journal of the International Society for Industrial Ecology. Although the society formally opened its doors to membership in 2001, the journal has been in publication since 1997, publishing four issues annually. The journal is owned by Yale University and published by the MIT Press.
Industrial ecology is a field that emerged in the early 1990s, from a paper authored by Robert Frosch and Nicholas Gallopoulos, both researchers at General Motors. The paper, ultimately entitled "Strategies for Manufacturing" (Frosch & Gallopoulos, 1989) was originally titled "Manufacturing – the Industrial Ecosystem View," but was not accepted. The paper uses the earth’s natural ecosystems as a model for manufacturing and industrial processes, and suggests that a similar “ecology” model be put in place for manufacturing and industrial processes; a so-called “industrial ecology” system might be much more efficient and economical, and might greatly reduce end-of-process wastes that are currently being generated, the paper suggests.

Today, the field of industrial ecology seeks to “understand how the industrial system works, how it is regulated, and its interaction with the biosphere; then, on the basis of what we know about ecosystems, to determine how it could be restructured to make it compatible with the way natural ecosystems function” (Erkman, 1997).

**Motivation and Target Audience**

According to the International Society for Industrial Ecology (International Society for Industrial Ecology, 2008), the society “promotes industrial ecology as a way of finding innovative solutions to complicated environmental problems, and facilitates communication among scientists, engineers, policymakers, managers, and advocates who are interested in how environmental concerns and economic activities can be better integrated.” Its mission is to “promote the use of industrial ecology in research, education, policy, community development, and industrial practices.”

Similarly, the Journal of Industrial Ecology states on its website (MIT Press, 2008) that it “seeks to reach both a professional and an academic audience and to cross disciplinary and national boundaries... The journal serves a diverse audience including managers, policymakers, professionals, academics, and technical researchers.”

So, it can be concluded that the journal caters to people across many disciplines and sectors (e.g. managers, policymakers, academics, professionals, etc.) and assumes minimal common experience across its membership. Here, it is assumed that the contributors or authors represent a narrower range of disciplines than does the readership, since, for example, policymakers are less likely to publish journal articles than they are to write bills and advocate change. Nonetheless, as will be shown, the contributors still span a broad range of sectors.

**4.2.2 Why choose the JIE?**

The collaboratory described in this thesis is founded on the concept of sharing, exchanging, building and communicating openly throughout a community. The wider the diversity of the participants, the more diverse the knowledge they
have to share with the community. Therefore, the JIE community was selected as a promising target for a test of this Web application because it seeks a diverse group of readers and contributors.

The community is comprised of a diverse group of people from many backgrounds, who are all working toward a common goal. The members are, therefore, likely to cooperate with one another when put in an open, sharing-oriented environment. Furthermore, participants in the field of industrial ecology know how diverse the field is, so they anticipate the need to interact with others, both in and beyond their native sectors. The participants in industrial ecology should be fully aware that no one person knows everything in this field.

As will be shown, the journal’s contributors are largely scientists or engineers in academia, from independent research organizations, or from industry. Their work is largely computer-based, and technology is presently proliferating rapidly throughout the research community. Therefore, the participants are highly likely to be proficient with computers. This proficiency suggests that they will be willing to take the next step toward even further use of technology in their work.

Given the propensity of the JIE community to adopt the new collaboratory technology, the following sections will take a further step to examine the make-up of the user group. A survey of the contributors to the journal will be conducted on both macro- and micro-levels; the readership will also be studied. These studies will help determine more confidently that the community is collaboration-ready. Similarly, the overall technology readiness of the community will be analyzed on a deeper level, to confirm that the group is prepared to adopt the new technology itself.

### 4.3 Collaboration Readiness

The concept of “collaboration readiness” was discussed in Section 2.2.3. To recap, the collaboration readiness of a group refers to the extent to which the potential collaborators are motivated to work with one another (Finholt, 2002). Because the JIE community is made up of both contributors to the journal (i.e. the articles’ authors) and the readers of the journal, the collaboration readiness of each of the parties will be examined below. At the end of both discussions, conclusions will be made about the collaboration readiness of the JIE community as a whole.

#### 4.3.1 The Contributors

Because “collaboration readiness” generally refers to the willingness to share, by both individual participants and whole organizations, determining the extent to which contributors currently collaborate is one way to measure the collaboration readiness of the users. Contributions to the journal can be looked at on two levels: a macro-level, and a micro-level.
The macro-level view of the contributors gives a feel for the overall make-up of the journal. Examining the contributors as a whole provides insight as to how useful the collaboratory would be to the overall JIE community. The micro-level view provides a look at the individual groups of co-authors. Examining the types of collaborations that already exist between the authors that contribute to the JIE will provide an understanding of how the collaboratory might benefit already-existing collaborations, and facilitate knowledge transfer that already takes place through other means. Both the macro-level and micro-level compositions will be addressed in turn.

The Macro-Level Composition

At the time of this study, ten full volumes of the JIE and three issues of the eleventh volume had bee published; these existing publications were the basis of the following analysis. The following analysis refers to “articles” in the journal; these articles consist of all written contributions in a given issue: the columns, the forum articles, the research and analysis articles, the applications and implementations articles, and the book reviews. The following statistics represent the macro-level composition of the JIE contributors.

49.2% of the articles in the journal are co-authored. Of those co-authored articles:

- 47.7% are co-authored by 2 people
- 28.7% are co-authored by 3 people
- 12.8% are co-authored by 4 people
- 4.5% are co-authored by 5 people
- 3.5% are co-authored by 6 people
- 1.0% are co-authored by 7 people
- the remaining 2.1% are co-authored by 8 or more people.

The contributing authors represent 32 countries. The breakdown of the author affiliation by country is:

- 40.6% from the United States
- 12.3% from the Netherlands
- 8.6% from the United Kingdom
- 6.7% from Japan
- 3% from each of Sweden, Germany, Australia, and Switzerland
- 2% from each of Austria, Finland, Canada, and Norway
- 1% from each of China, Denmark, France, and India
- fewer than 1% from each of Italy, the Philippines, Spain, Belgium, South Korea, Vietnam, Indonesia, Portugal, Thailand, Singapore, Argentina, Brazil, Malaysia, South Africa, Sri Lanka, and the United Arab Emirates.

The breakdown of the contributing authors’ institutional or sector affiliations is:

- 70.7% from universities or institutes of higher education
- 18.5% from research organizations or consultancies
• 7.0% from industry, commercial business, or professional societies
• 3.7% from government agencies or outreach and advocacy groups.

Figure 12: The percentage of co-authored articles per publication, over time

Although it is understood that the authorship of a paper is not an absolute measure of collaboration or indicative of collaborative work, co-authorship is accepted as one of the easier ways to gauge collaboration. In this thesis, it is assumed that co-authorship is the result of a collaborative effort. It should also be noted that, in the case of the JIE, the collaborative effort is not necessarily a research effort, because not all articles are research papers; some are co-authored opinion columns or book reviews.

On average, half of the articles published in the JIE are the result of collaborative work. Figure 12 shows the percentage of co-authored articles per issue. Although there are large fluctuations with each issue, it is clear that the fraction of co-authored articles to singly-authored articles increases over time (at a rate of approximately 0.69% per publication). This increasing trend indicates that within the JIE community, collaboration in research, science, and industry is increasing; technologies that help the collaborative process may be appropriate to this user group.

The authors of the journal’s articles are distributed throughout the world. Although the JIE is an American-based journal, only 40% of the authors are based in the United States. This global distribution indicates that the overall community generally consists of a remotely-distributed population, and that remote collaboration techniques should be employed for work or discussions that take place between community members.
Many, but not all, of the contributing authors are affiliated with universities. The remaining 30% of the contributors come largely from industry, independent research institutes, and consulting firms. This distribution of sectors suggests that little can be assumed about the contributors as a whole; the technology they use in their workplaces varies greatly, as do their daily tasks and interests. Not all users are immersed daily in a research environment.

Conversely, the diverse backgrounds suggest that users potentially have lots to learn from their fellow community members. Collaboratory technology could facilitate the transfer of knowledge between smaller, more specialized groups. The collaboratory could facilitate correspondence between users employed in industry and users who perform related research.

**Micro-Level Composition**

The micro-level composition is an examination of only those articles that are co-authored. The goal of the micro-level analysis is to determine the variety and types of collaborations that exist already between the journal’s contributors. Each co-authored article was categorized by the collaborators involved in the collaboration, based on Katz and Martin’s model, which was presented earlier in Table 2.

<table>
<thead>
<tr>
<th></th>
<th>intra -</th>
<th>inter -</th>
</tr>
</thead>
<tbody>
<tr>
<td>individual</td>
<td>between individuals or groups in the same department</td>
<td>between individuals</td>
</tr>
<tr>
<td>department</td>
<td>between individuals or departments in the same institution</td>
<td>between institutions</td>
</tr>
<tr>
<td>institution</td>
<td>between institutions in the same nation</td>
<td>between institutions in different countries</td>
</tr>
</tbody>
</table>

Some modifications to Katz and Martin’s model have been made; the result is shown in Table 6. Note that the group and sector levels are absent. Because of ambiguities in the data, the categories used to group the collaborations are less specific than those proposed by Katz and Martin. For example, because “sector” is difficult to define in this context, it was not included as one of the hierarchical levels. Likewise, “group” is not used, because there is not sufficient resolution in the data to determine if authors from the same department are from within the same research group. Therefore, the possible tiers included either intra- or inter-departmental, institutional, or national collaboration efforts.

It should also be noted that, given the data resolution available at the time of this study, the difference between inter-department collaborations and intra-institution collaborations is not discernable. The distinction between the two arises in the data shown only because some authors identify themselves with a school or department within an institution, whereas others simply list an institutional affiliation. The two effectively have the same meanings—that is,
both represent collaborations between departments within the same institution. Also note that “institution” is the general categorical name for any university, business, firm, organization, or other highest-level affiliation of an author.

Of the 49.2% of all JIE articles that are co-authored:

- 28.6% are intra-department collaborations
- 3.5% are intra-institutional collaborations
- 6.6% are inter-department collaborations
- 34.5% are inter-institution collaborations
- 26.9% are inter-nation collaborations.

The inter-group collaborations are of particular interest, since their members come from multiple departments, institutions, and even nations; inherent in each inter-group collaboration is an effort made by the participants to actively reach outside their immediate proximity and find others with whom to work. The scale of the inter-group collaborations is affected by two quantities: the total number of co-authors, and the total number of different “groups” that make up the collaboration.

A ratio, named the “inter-group collaboration factor,” was created to compare and rank the inter-group collaborations. The factor is calculated by dividing the number of different “groups” by the number of co-authors involved. For example, a paper written by 3 authors that come from a total of 2 different universities would have an inter-collaboration factor of 2/3, or 0.667 on an inter-institutional level. A paper written by 4 authors, each of whom comes from a different country, would have an inter-group collaboration factor of 4/4 or 1 on an inter-national level. This factor was used to break down the inter-department, inter-institution, and inter-nation collaborations listed above. The results are shown in Figure 13.

![Figure 13: The inter-group collaboration factor for inter-group collaborations](image-url)
33.4% of the journal’s articles are the result of inter-group collaboration, or collaboration across some established organizational boundary. It is obvious that the inter-group collaboration factor will have a misleadingly high frequency of 1, because all two-person inter-group collaborations automatically involve two different groups!

Nevertheless, the results demonstrate that over one third of all the articles published in the Journal of Industrial Ecology are the result of a cross-organizational collaborative effort. Each such collaboration brings with it administrative and logistical challenges, yet the prevalence of these collaborations indicates that the motivation of the participants exists to overcome such difficulties. This motivation increases the likelihood that a collaboratory that would facilitate these long-distance relationships would be accepted by the JIE community, and suggests that the collaboration-readiness of the contributors is certainly sufficient to accept the new collaboratory.

4.3.2 The Readership

It is difficult to precisely determine who constitutes the readership of the Journal of Industrial Ecology. Since it is the official publication of the International Society for Industrial Ecology (ISIE), and included in each membership subscription, for the purpose of this study it is assumed that the members of the ISIE constitute the journal’s readership. Of course, there are other readers as well; institutions and individuals may subscribe separately to the journal, without belonging to the society. However, it is very difficult to obtain data on these independent subscribers.

Unlike for the contributing portion of the JIE community, enough data is available only to analyze the macro-level composition of the readership of the JIE community. At the time this analysis was conducted, there were 428 unique members of the ISIE listed in the member directory on the society’s Web site (International Society for Industrial Ecology, 2008).

Data on the readership was gathered from the information provided in the member directory on the website. Each member has access privileges to a separate members-only section of the website, which allows a member to view the member directory, and update his own member profile. A member is required to provide his name, email address, country, organization, and sector affiliation, and may optionally provide a list of interests within the field of industrial ecology. These attributes are visible to all other members. The readership analysis conducted is based on these attributes, as included for each listed member of the society. Note that if a member never log onto the Web site to update his member information, only minimal information is provided: a member’s name and email address. Consequently, not all attributes are known for all members.
390 members, or 91.1% of the total readership, provide a country of affiliation. 35 countries are represented. The breakdown of member affiliation is:

- 42.3% from the United States
- 8.2% from Canada
- 6.2% from the United Kingdom
- 5.9% from Japan
- 4.1% from China
- 3.9% from Norway
- 3.6% from Germany
- 3.3% from Australia
- 2.8% from The Netherlands
- 2.6% from Sweden
- 2.3% from each of Finland and France
- 1.5% from each of Austria, Brazil, and Spain
- 0.8% from each of Portugal and Switzerland
- 0.5% from each of Armenia, Denmark, India, Ireland, The Philippines, Taiwan, and Thailand
- 0.3% from each of Belgium, Egypt, Hungary, Italy, South Korea, Malaysia, Mexico, Micronesia, Nepal, South Africa and the United Arab Emirates.

379 members, or 88.6% of the total readership, provide a sector affiliation (from the following 6 options available to them). The breakdown is:

- 68.1% academic
- 7.9% corporate
- 7.7% research institute
- 6.3% government
- 6.1% consulting
- 4.0% nonprofit / advocacy.

Only 123 members, or 28.7% of the total readership, indicate a list of interests. Of those, the following interests (from a fixed list) are indicated, along with the percentage of members who list the interest included:

- Industrial Ecology Tools and Modeling: 59.6%
- Material/Substance Flow Analysis: 55.3%
- Eco-Efficiency: 50.4%
- Indicators, Metrics, and Corporate Sustainability: 45.5%
- Industry Technology Change and the Environment: 43.1%
- Life Cycle Management: 43.1%
- Design for Environment: 41.5%
- Sustainable Manufacturing: 40.7%
- Input-output Analysis in LCA and MFA: 39.8%
- Industrial Ecology in a Global Context: 39.0%
- Corporate Environment Management: 38.2%
- Managing Energy and Greenhouse Gases: 38.2%
- Industrial Ecosystems & Eco-Industrial Parks: 36.6%
- Regulation and Policy: 36.6%
- Social Dimensions of Industrial Ecology: 36.6%
- Life Cycle Design Planning and Assessment: 35.8%
- Sustainable Cities and Regional Metabolism: 35.8%
- Thermodynamics of Energy and Material Systems: 26.8%
- Product Oriented Environmental Policy: 25.2%
- Spatial Dimensions of Industrial Ecology: 17.9%
- Biomaterials and Biocomplexity: 13.0%

Because very few members actually supplied all the information requested of them, some assumptions were made to fill in missing data and obtain a larger data sample. For example, although many users did not explicitly provide a country of affiliation, their email address domain names provided enough information to determine their country of work or residence. Similar extrapolations were made to more completely determine users' organizational affiliations and sector affiliations.

Although there are some clearly dominant groups among the categories of readership affiliations, there is also diverse representation of many affiliations across the categories. For example, although members from the United States make up 42.3% of the readership, readers still represent a total of 35 countries of varying levels of economic and social development. The journal's content reaches readers throughout the world, where consequences or implications of the journal's articles may have vastly different embodiments.

Similarly, even though people working in academia constitute 68.1% of the readership population, there is also significant representation from corporate, research, government, consulting, and non-profit/advocacy sectors. The statistics show that the journal does meet its multi-disciplinary goal of reaching readers from a variety of sectors, and can realistically serve as a forum for "continuing exchange of information and opinions" (MIT Press, 2008).

The diversity represented by the country and sector affiliations suggests that the composition of the readership of the Journal of Industrial Ecology is suitable for distance technologies. Readers are largely not collocated with other readers, and could benefit from Internet-based technologies that allow them to experience first-hand the models that they read about in papers. Similarly, the journal's readers could benefit from distance technologies through the introduction to work of others from sectors other than their own. The collaboratory would allow corporate users to easily experience the work being conducted in academic and research areas, to see what might be used in their specific applications in the future. Specific relationships between research labs and corporations need not be established in advance; rather, if a corporate representative interacts first-hand with technologies being developed that might be applicable to her company's work, she can then initiate a partnership with the respective research lab.

The data obtained from the expressed interests of the readers is more difficult to synthesize than the country and sector affiliations, because a significantly smaller
portion of readers provided data on their interests. Only 28.7% of the members listed in the directory indicated one or more interests; for the purposes of this preliminary survey, it is most instructive to assume that this portion of the readership is a representative slice of the entire JIE readership community, although the veracity of this assumption is difficult to verify.

Each of the areas of interest available for members to select was chosen by some number of members. The most frequently chosen area of interest was "industrial ecology tools and modeling," selected by 59.6% of the respondents who specified interests, and the least frequently indicated area of interest was "biomaterials and biocomplexity," selected by 13.0% of the respondents who specified interests. The average number of interests listed per respondent was 8, and the mode number of interests listed per respondent was 6 (by 15.5% of the population). However, the range of interests indicated was as wide as possible: two respondents listed all 21 possible choices as areas of interest, and four respondents listed only one area of interest. This range of responses confirms that there are many interests represented, both personal and professional, among the readers of the journal. It is expected that the model-sharing collaboratory technology would facilitate communication and collaboration between users with similar interests, and provide users with different or complementary interests and novel ways to collaborate and work together more efficiently and productively.

4.3.3 Collaboration Readiness Conclusions

The surveys conducted of both the contributing population and the readership of the JIE community suggest that the JIE community is indeed sufficiently collaboration-ready to accept this new collaboratory. Already, over half of the publications in the journal are the result of collaborative work, and in many cases, the collaborations span not only multiple countries, but multiple continents. Similarly, the readership represents a wide range of professional sectors, from across the planet. The population is global in makeup, mindset, and scope. The community is ripe to embrace new distance technologies that will further aid existing collaborations, and that will encourage new collaborations that would otherwise not have been possible to initiate.

4.4 Technology Readiness

The concept of "technology readiness" was discussed in Section 2.2.3. To recap, the technology readiness of a group refers to the readiness of the group’s user habits and infrastructure to accept technology for distance work. It is generally agreed that novel, advanced technologies should be introduced to new communities in small steps, and only when the community is ready to accept and use the new technology. Therefore, it is important to examine the extent of current technology use by the members of the JIE community, to gauge the JIE community’s technology readiness.
4.4.1 Current Technology Use by the JIE Community

Current procedures of the JIE as an organization, and of its individual members, were surveyed to gauge the technology use of its community. The following practices were observed:

- Most authors provide personal email addresses with their correspondence information, and some provide URLs of relevant personal or lab-specific Web sites.

- Papers are submitted to the journal in electronic formats only. The editorial board of the journal is an international body consisting of 60 people from 12 countries; it is assumed that the members of the editorial board use electronic means to communicate amongst themselves and to perform their editorial duties.

- Letters to the editor, and in response to previously published works, are submitted electronically, and published online only. Links are provided to the letters from the article to which they correspond; to view the letters, users must visit the journal’s Web site. Letters are posted immediately after they are received, so the website is updated dynamically. It is unknown how frequently users visit the Web site. Authors of original works are required to respond to at least two letters written in response to their original work (if applicable).

- Electronic supplements, or e-supplements, are provided with some articles, and are only available on the Internet. The embodiment of these e-supplements varies, but they have previously included supporting information such as complete data sets, derivations, appendices, and Chinese translations.

The above sampling provides a feel for the community’s current use of technology on a macro level. Without questioning individual authors about their research habits and technical prowess, the technology use that is assumed by the community’s leaders can be characterized. A technology comfort threshold is clearly inherent in the JIE organization. For example, without surveying any individual users, it is known that each contributor or contributing collaboration probably has a connection to the Internet with which the contributor(s) can submit work, browse other contributions, and read responses to his or her past work.

Additionally, a press release (Journal of Industrial Ecology, 2004) issued by the JIE when it first started publishing its abstracts online in Chinese indicates the importance of including its Chinese community members, since China plays a growing role in the development of the field of Industrial Ecology. The underlying assumption in the press release is that, since the abstracts are only available over the Internet, the journal’s colleagues in China have Internet access.
Of the many collaborative groups that exist within the JIE community, 26.9% are international. It is assumed that international collaborations require long-distance communication between the members of the endeavor. Common means of communication between individual researchers and small groups of non-co-located researchers include telephone communication, audio and/or video conference calls, emails, and electronic meetings (often through Microsoft NetMeeting) (Olson & Olson, 2000). Each of these requires a degree of comfort with current technology standards, or at minimum a facility for adopting and employing new technical skills.

4.4.2 Technology Readiness Conclusions

Olson and Olson consider the use of email a good benchmark for technological advancement. Organizations with inadequate habits and infrastructures for adopting technologies have usually not yet adopted email. Clearly, the JIE user group in question has, at the very least, grasped the use of email.

Olson and Olson provide an example of one research community that had sufficient technology readiness to accept a new collaboratory. A group of geographically dispersed physicists began their collaborative efforts proficient in email, telephone, and fax use. Over time, their collaboratory developed to include online access to instrumentation, and eventually, their entire project became Web-based. Minimally, the JIE community has the same level of technological prowess as the physicists described above, and more realistically, the community has much more extensive exposure to Web technologies. This comparison to the physicists, and the additional known information about the practices of the JIE, suggest that the JIE user group should have sufficient technological readiness to accept this new collaboratory technology.

4.5 A JIE-Specific Use Scenario

As an extension of the general scenarios for the use of the collaboratory, there are also two basic use scenarios for the JIE community. Each of the two scenarios is described step-by-step below. The first use scenario is from the perspective of a researcher who has published results with the journal in the past; the second perspective is that of a researcher who is interested in finding a particular model hosted in the collaboratory, for use with his work.

Scenario 1: The reader
Imagine a researcher who has published findings in the Journal of Industrial Ecology in the past, and who reads its articles each quarter. He is aware of other scientists and researchers who also contribute to the journal, and whose work is of interest to, and often related to his. He would like to explore their research more in-depth and first-hand.
He visits the journal’s homepage to browse the most recent online publication. He selects the volume and issue of interest (Figure 1).

He is then taken to a page that briefly lists the content of this quarter’s journal, including the titles and contributing authors of each work. He selects an article that is of interest to him (Figure 2).

He is brought to the “homepage” of the article he has just chosen (Figure 3). At this page, he can read more about the article he just selected, including details about the contributing authors, and the abstract.

Because the goal of this website is to enable users of the site and readers of the journal to more easily share data and ideas, there are additional features he can access now. He can:

(a) download a PDF of the entire article, or
(b) use the computer model(s) discussed in the paper and uploaded by the paper’s author(s), or
(c) enter a discussion page with other readers of the article / users of the article’s model(s).

He looks at the models available to him, and decides that he first wants to investigate the model that compares three different estimates of extraterrestrial radiation on a given day, since it seems potentially relevant to what he is working on. So he clicks "use this model."

He is now taken to a page that contains the model he just selected (Figure 4), and that is dynamically generated specifically for him on the spot. He sees that there are two "input variables" that he can enter: day number and solar constant, each of which has predefined units. He enters both, and chooses "run this model!"

Soon, his new outputs are sent back to him on the same Web page, under the "output variables" heading. He can try changing the input parameters as often as he wants, enabling him to use first-hand the model that he has read about in the paper.

After trying this model, he decides that he wants to try the other model that has been published in conjunction with the paper. He selects the article’s name again in the breadcrumb trail at the top of the page (“the changing metabolism of cities”) (Figure 11), and he is brought back to the article’s homepage. This time, he selects the model that calculates power generated by a multi-turbine system.

Now he is brought to a different model page. He sees that this model has a variety of user inputs, including two binary input options. He can read a description of the model at the top of the page, before setting the input parameters to those of his choosing. When he is ready, he chooses "run this model!" Again, the model runs in the background, and sends the resulting power output back to him on his screen.
This time, the user decides he wants to leave a comment about the model that he has just run. So he selects the option to "discuss this model", found right below the model interface itself (Figure 5). He is brought to a discussion page (Figure 6) that is specific to the article itself, so he can see what others have thought about the article and/or its models.

Scenario 2: The model-seeker
Now, the user is a researcher who is interested in finding a model that has been published online to integrate into her own computer simulation. Her main concern is not with reading the literature associated with the model, but with finding a model that's the right fit for her experiments. For this, she explores the JIE's model database through the 'model view' (Figure 7).

She sees that she has three options available to her: a search, a list of all models, and an integration option. She decides to begin her search by browsing through all the available models, so she clicks the "see all models" button.

She is taken to a list of all the models available to the JIE community (Figure 9). These have been published online by the various authors of the journal's past articles. She can sort the articles by name, date, and contributor. This top-level view lets her see with which article the models are associated, without having to select the model itself. When she finds a model that interests her, she clicks its name, and is taken directly to the model page.

If instead the user decides that she wants to perform a search to find a smaller subset of models that might be relevant, she can return to the main model-oriented homepage by again choosing the "model view" tab. This time, she selects "search database".

She is brought to a search page (Figure 8) that allows her to choose from a basic search, or an advanced search, depending on the specificity of the model she has in mind. Her query will return a list of model hits on a new Web page.

Finally, if she decides that no one model is adequate enough for her work, or that she would like a model that does something different than any of the models provided already in the JIE model database, she can choose to integrate models already available in the database. She returns to the "model view" homepage, and selects the "integrate!" button.

She is taken to a page that will walk her through the steps of integrating two or more models that have similar parameters (Figure 10). At the end of the integration process, not only will she have the option of saving and running her new model, but she will be able to publish it on the Web site, too, so others will be able to use her new model in the future.
Chapter 5  Future Testing and Conclusions

The final chapter of this thesis seeks mainly to provide directions for the follow-up work that must be performed before this collaboratory can be formally deployed and used by any user group. The work is still largely unfinished, and there remains much to be completed. Before discussing the future work, the main contributions of this thesis will be summarized. Then, key questions that must be answered through user testing will be discussed. Answers to these unknowns must be obtained from user feedback and extensive testing of the collaboratory. Finally, conclusions about this collaboratory technology will be provided.

5.1 Contributions of this Thesis

This thesis has provided the first steps toward deploying a new collaboratory technology to foster increased and more efficient collaboration between scientists, researchers, and industry workers. The thesis provided discussion of three main points: an overview of the current situation of collaboration and collaboratory use in research; the technical details of a new collaboratory technology; and a survey of the targeted user group, and why the collaboratory should be successful among this user group. Each of these main points of contribution is summarized below.

5.1.1 A Survey of Current Collaboration and Collaboratories

Collaborations

First, this thesis provided both a description of the evolution of collaboration in scientific research, and the current state of collaborative efforts in science and engineering. A general meaning of the term “collaboration” was provided, although throughout the literature, many definitions can be found, each with a slightly different connotation than the others. Based on a literature review of existing work on scientific collaborations, there was a discussion of scientists’ motives to collaborate; the difference between motivations and benefits was stressed.

There was also a discussion of the types of people and groups that collaborate. Based on existing literature, a few different viewpoints were taken to examine who collaborates. First, the question was examined from a view proposed by
Katz and Martin (Katz & Martin, 1997), whereby collaborators are identified by the organizational or institutional group to which they belong (e.g. a laboratory, a research institute, or even a nation). A chart (Table 2) was provided that breaks down collaboration types based on how many people, and from which organizational groups, participates in a given collaborative endeavor. This chart is revisited later in the thesis. Then, the breakdown of types of people who collaborate took a more fundamental viewpoint, and examined the personalities of each collaborator, to establish types of people who collaborate, and their motivations (as per the discussion in (Bozeman & Corley, 2004)).

Finally, based on the literature, a summary of the benefits of collaboration was provided. The benefits were broken down into those that are advantageous to the researcher(s) personally, and those that benefit a larger project or that work toward the betterment of science.

Collaboratories
This thesis introduced the concept of the “collaboratory,” a relatively new term coined in the 1990s, that encapsulates the use of computers and modern technology to facilitate the collaboration of scientists and researchers who are geographically dispersed. First, a discussion of the evolution of the collaboratory was provided, and its currently accepted definition was discussed. As defined by the Science of Collaboratories (SOC) (Bos et al., 2007), a collaboratory is

an organizational entity that spans distance, supports rich and recurring human interaction oriented to a common research area, and fosters contact between researchers who are both known and unknown to each other, and provides access to data sources, artifacts, and tools required to accomplish research tasks.

A discussion of the past work on collaboratories was provided; the history of research on collaboratories is not nearly as old or rich as that on collaboration. Collaboratories are, in the larger picture, still in their infancy. However, the SOC has provided the definitive work on collaboratories to date, by providing a very thoroughly researched taxonomy of seven overall types of collaboratories. However, as was discussed in this thesis, the SOC taxonomy system does not apply to all collaboratory types (including the one proposed in this thesis), and will require continual reevaluation as technologies continue to evolve.

Finally, this thesis provided a discussion of considerations that must be borne in mind when designing new collaboratory technologies. Most notably, the user group, and its capabilities must be considered very carefully, for the success or failure of a collaboratory is largely based on the appropriateness of the technologies for the target user group. As such, the “collaboration readiness” and “technology readiness” were stressed as two important factors that must be examined for any user group that might consider adopting a new collaboratory.
5.1.2 The Creation of a New Collaboratory

This thesis also endeavored to create a new collaboratory technology for use by both existing research communities, and those that have yet to establish themselves. The goals of the collaboratory proposed here are twofold. First, it seeks to increase the efficiency of research collaborations, by increasing the ease by which researchers can communicate amongst themselves, and by minimizing the repeated work that often takes place when one researcher picks up where another leaves off. Secondly, it seeks to foster an open, commons-oriented attitude of sharing. The collaboratory is based on the notion that researchers will be willing and eager to share their work with others, with the goal of faster, more efficient advances in their field.

A detailed description of this collaboratory was provided. The collaboratory is Internet-based, and can be run by a user using only a Web browser. Images were provided of the collaboratory’s interface, and the collaboratory’s many features were discussed; both descriptions and reasons behind the choice of features were provided.

Finally, based on the findings of the literature discussed in Chapter 2, the importance of the user group was stressed. Although it is hoped that this collaboratory technology will be adopted by a wide variety of user groups, it was designed primarily with academic journal-based research communities in mind. The characteristics of this type of user group were discussed, and features of the collaboratory that were designed for this community were examined.

5.1.3 A Study of the Targeted Test Group

Lastly, this thesis established that it is critical to understand as much as possible about the targeted user group before designing a new technology for it and its members. It was explained that the success or failure of a collaboratory is dependent on whether or not a group is ready to accept the collaboratory technology. As such, the user group was defined as the Journal of Industrial ecology community: both its contributing authors and its readership.

This user group is both novel and an appropriate fit to the collaboratory at hand, because its members come from a diverse background. Because of the diverse nature of the field of Industrial Ecology, the journal does not cater only to academics or researchers. It also seeks the participation of members of industry, policy makers, and volunteers, for example. Members of the community are aware of how diverse the group is, and of the large number of different viewpoints their peers can offer. The importance of collaboration with others from different fields and of different specialties should be especially clear among this group, since Industrial Ecology is such a complex and heterogeneous field.

An extensive study of the make-up of the contributors and the readers was provided, both from a collaboration standpoint and a technology-use standpoint.
It was concluded that the members of the Journal of Industrial Ecology community are indeed both collaboration-ready and technology-ready. If (and when) the model-sharing collaboratory in question were to be deployed among this user group, its chances of success would be promising.

5.2 Future Work

This thesis is only the beginning to the successful deployment of this new model-sharing collaboratory. The work that has yet to be completed on this project is significant; the author wishes to stress the importance of the work that remains and her hope is that it will be completed in the near future. The following sections address key questions that must be answered, and stress how critical an extensive user test will be before this collaboratory can be formally deployed for public use.

5.2.1 Testing

A test of this new collaboratory among its target user group is critical to assessing how well the collaboratory lives up to its expectations and fulfills its design goals. Therefore, a strong relationship must be established between the Journal of Industrial Ecology community and the developers pursuing the deployment of this collaboratory. The test within the user group will help answer many key questions, and will provide much-needed feedback that is critical to the ultimate success of the collaboratory as a freely available software download. The following sub-sections address the types of questions that should be answered before proceeding with the deployment.

Willingness to Share?

The success of this collaboratory depends largely on the participants' willingness to share their research and original work openly with colleagues within their community.

The collaboratory assumes that users will be inclined to contribute to the community's commons by openly sharing the models and simulations from their research, similarly to how users share their musical compositions in ccMixter, how open-source code and programs are freely shared and exchanged, and how users openly share their knowledge in the form of encyclopedia articles on Wikipedia. (All three of these examples were discussed in more depth in Chapter 3.)

However, it is unknown whether researchers and scientists will be as willing to openly share their professional work with others (who are possibly unknown to them), as the common people are open to sharing their creations with Internet communities of similar interests. Referring to the COIS model mentioned earlier, the scientists, researchers, and industry representatives are considered "lead users" more than they are "common people." The open, commons-oriented
behaviors witnessed in the COIS model may no longer be applicable with professionals.

Deploying a test implementation of the collaboratory within the Journal of Industrial Ecology community will reach a large spread of sector representatives, and will help to determine which users, from which professions, are more inclined to openly share their work with others.

Improved Efficiency?

One of the original fundamental goals of this collaboratory is to reduce the repetition that occurs in collaborative scientific research; often, one researcher must repeat some of the experimentation or modeling of the previous researcher, before picking up where he left off. Ideally, the collaboratory will provide a way to increase the efficiency of collaborations, by providing researchers a means of easily using the models and simulations of others.

Whether or not this is actually the case remains to be seen. By testing the collaboratory with the JIE community, it will hopefully become clear whether less time is spent on unnecessary, repeated work than currently takes place, due to the collaboratory technology. It will also be interesting to gauge whether or not researchers perceive any increased efficiency or time savings in their research as a result of use of the collaboratory.

Ease of Use?

The collaboratory has not yet been tested with a specific user group. Therefore, it is critical to receive feedback on the overall user experience with the collaboratory. Although it has been designed to be easy to use and intuitive, it is unknown whether users perceive it as such.

Therefore, feedback regarding the collaboratory’s ease of use and intuitiveness, to both share one’s own work and use or synthesize the work of others, is needed. This feedback should be used to determine what improvements can be made to the design of the collaboratory. Is the layout logical to the user? Should features be added? Are there any current features that are unnecessary or distracting? Is anything unclear?

Overall Helpfulness and Success?

Finally, the obvious question to answer through testing is whether the collaboratory helps the community, and if it is an improvement over the current state of the members’ collaborative research. It is not immediately obvious what questions should be asked to answer this question, but it is suspected that the implications of the overall user feedback will provide sufficient feedback.

The overall success of the collaboratory can be gauged by determining whether users enjoy the experience of sharing and synthesizing through the collaboratory, if users collaborate more than they used to, if users have met new collaborative partners as a result of the collaboratory, and even if users feel they are spending their time more efficiently. Have they discovered things they did not know
before? Do they find articles and publications easier to comprehend with accompanying, interactive models and simulations? The answers to all of these questions should help to determine whether the collaboratory meets its goals and design intents.

5.2.2 Refinement and Deployment

After receiving user feedback from a pilot test, it should be obvious what changes need to be made to the content of the collaboratory, and to its features, before it can be publicly deployed. Ideally, after the changes have been made, a second version of the test implementation can be deployed, again within the pilot group, to both verify that the changes have led to improvements in the overall experience with the collaboratory, and to determine where there is still room for improvement.

Recall that, in its final form, this collaboratory will be available as a stand-alone application, similarly to how Internet users can currently download and install wiki software. Before the collaboratory can be deployed, much more work must be done on the software development side, to convert what is currently a one-time, single-use implementation of the collaboratory to a robust, reproducible, stand-alone software package.

A beta version of this software package will then be available for download and testing by interested users. After obtaining feedback about the package installation processes, making changes as appropriate, and working out any remaining bugs in the download, the collaboratory application will be made freely available for download by any interested user groups. From there, the collaboratory will be in the hands of the users to install, maintain, and use as they see fit. Hopefully it will be as useful to them as the designers envisioned it to be.

5.3 Conclusions

Collaboration continues to become increasingly widespread throughout scientific research. Scientists and researchers are not only increasing the frequency with which they collaborate with others, but they are expanding their collaborative horizons. Advances in technology and the augmented prevalence of the Internet allow scientists and researchers to collaborate with a wider range of people, located arbitrarily far away from their research locales.

Collaboratories, the collaborative organizations that use distance technologies to work beyond the conventional physical boundaries of collocated collaborations, represent a relatively young notion that is still maturing and evolving. Nonetheless, collaboratories are earning a name for themselves in the research sphere, by facilitating—in a number of different ways—collaborations that span distance, institutions, and even nations.
Collaboratories, coupled with the Internet, have a very promising future. Continuing advances in technology should allow collaboratories to reduce the inefficiencies and frustrations often associated with distance work, and to improve the ease with which scientists and researchers communicate amongst themselves and with others.

One such collaboratory was proposed in this thesis; it is hoped that, when fully implemented, this collaboratory will introduce a new paradigm to the research community: a paradigm that encourages the open sharing of ones of work with ones peers, to improve the efficiency and ease with which collaborations take place. This open, community-oriented mindset, coupled with the technology to support it, has the potential to change the way research is approached and conducted.
Bibliography


Wulf, W. A. (1989). *The national Collaboratory–A white paper*

