Chapter 1 -- Introduction

In the twenty years that elapsed between the award of my Bachelor's degree in Geography and my entry into graduate school, I made my living practicing software engineering. I had eventually specialized in the design of embedded systems, systems whose behavior was managed by a computer. The software for these systems always arises from an elaborate behavioral model. Designers of such systems invariably anthropomorphize their system, attributing its behavior to a being that has a personality that is sometimes proper and placid, and at other times wretchedly wicked. When the system decides to be wicked, it's time to get out the software knives and do a little brain surgery.

Why couldn't a model of some small facet of human behavior be built using the same tools? Dismissing the problems of infinite complexity and the open system nature of human behavior, I decided that it could be done. The model of the system would have to be based on observations of reality. Causality would have to be modeled on theory derived from observation and experiment. A final model, a simulation equivalent to the behavioral model of software developers, would provide a validating test. Certainly such models would be superior to simplistic black box math models that used aggregated data and were empirically derived without recourse to cause.

While doing my Master's work, it became quite obvious that the major difficulty presented in such modeling was the immense amount of work to build such a model. A behavioral model of an activity such as human migration[1] would have to connect the best work that had been done over the past century, and more. Furthermore in order to incorporate the work of contemporary scholars, it would have to be maintained constantly. Clearly, models of this nature would have to be properly scoped and developed within a well-funded, long-term, large-scale collaborative effort guided by a team of dedicated researchers. Such a team would need a set of tools to make their work manageable, and would also need an organizational system to keep the team working

harmoniously. This dissertation is a description of the design and use of such a system, the Research Web, and its tools.

1.1 Importance of Asynchronous Collaboration in Research

Over the years scientific research has become more difficult because the research that produces new scientific knowledge is supported by a long and ever more complex body of established knowledge. The hypotheses that suggest new lines of research emerge from a professional reading of the existing body of knowledge. Original research pushes beyond existing knowledge and thus requires more specialization and/or synthesis and often more effort, including mastering the research legacy of the problem. Increased effort requires either a larger research team or more time. If one subscribes to the view that research costs and results are roughly proportional to person-hours expended, then time becomes the limiting resource, and a larger team must be assembled. This team then needs to be drawn into an effective collaboration. There is a long-standing trend toward larger collaborations[2,3,4]. The scale of research is often much larger than in the past, and continues to grow each year.

> *... the complexity and subtlety of the constitutive problems that must be faced by the sciences we populate, the diverse requirements for both relevant information and understanding that they now impose, and the rapid emergence of a wide range of sophisticated technologies that are potentially of great value to us, all converge to make obsolete the "Mom and Pop enterprise" (i.e., the single investigator with a few relatively inexperienced graduate students) that virtually all of us have operated in the past and that most researchers still operate.*
>
> --- William Bevan[5]

Specialization in scholarship has fragmented disciplines, so a good deal of interdisciplinary communication is often necessary. Not only have specialties been created by fragmentation, they have been created by new discoveries, by integration of disciplines, and by scientification of art and technology[6]. Specialists often find themselves isolated due to the increased specialization of scholarship. The isolation is an outgrowth of academic staffing policies; there is simply no need or economic justification

for more than one specialist at most universities[7]. As a result, groups of specialists, hence potential collaborative teams, find themselves geographically dispersed. As a consequence of geographic dispersal, people are operating on very different UMT time schedules and are frequently unavailable for synchronous communication.

The recent, and continuing, burst of innovative communication technology has provided research teams with capabilities that augment conventional scholarly collaboration. Software to facilitate most aspects of scholarly research is in some stage of planning or implementation. Activity in both the software and hardware to support research is flourishing. Driven in part by this intense engineering activity, scholars have been providing the intellectual grounding for improving collaborative technology. Much of this research, especially that associated with commercial software, is pedestrian automation of existing methods, but some research has laid open the intellectual basis of collaborative activity. The intellectual pieces are beginning to emerge; but scholarly collaboration cannot realize its potential if recent research and new technology are not utilized.

> *The change from paper-based to electronic text is one of those elementary shifts – like the change from manuscript to print – that is so revolutionary we can only glimpse at this point what it entails.* ---- Jerome J. McGann[8]

Over the past half century, English has become the language of science. In several nations the English language is taught to all students in the common schools (e.g. Holland and Norway). English is the language of instruction in elite Universities in India, Russia and China. English as a Second Language has become an educational industry throughout the world. This slowly unfolding event has, by lowering a primary communication barrier, opened many more opportunities for collaboration. The Internet has provided the spark that might ignite an age of collaboration.

Politically and philosophically, the tenor of the times has changed to a more inclusionary and collaborative imperative. Elitism is recognized and usually banished from discussions affecting the public. Science, engineering and technology permeate our lives and are now subject to the same public scrutiny as political issues. Funtowicz and Ravetz[9] characterize much of today's science as being "Science for the Post-normal Age." This new science is applied to problems that are either highly risk-laden or have very high stakes. Both these situations call for a dialog that is extended beyond the elite to all groups and citizens holding a significant stake in the issue. Post-normal science is a more hopeful reaction to the Kuhnian "normal Science" than is the more fashionable, cynical and pessimistic postmodernism. The environment proposed in this dissertation is designed for post-normal science.

1.2 Key Assumptions, Novelty, and Bias

This dissertation proposes the adoption of methods that apply new technologies and management techniques to the existing world of scholarship. Some very basic assumptions about the nature of scholarship were used to design the collaborative environment. The environment that emerges is not revolutionary, but introduces some novel methods that will require some minor adjustments to behavior in scholarly collaboration. I owe it to myself and to the work to discuss my biases, which are, I believe, benign.

1.2.1 Assumptions

The two major assumptions in this work are that scientific research is framed in documents[10], and that scientific progress is made through critical dialog about those documents[11]. Many scholars and philosophers support these assumptions[12]. Two additional propositions, supported below, are of great importance: the importance of reflection and the importance of recording information transactions.

The software components for the collaborative environment are designed to be compatible with the World Wide Web; they are therefore, going to be effective only as long as the WWW remains the dominant information-sharing facility. Obsolescence of materials on the WWW and tools designed for WWW support is certain to be evolutionary due to the enormous body of content in place today. An example of the evolution is seen in the current challenges to HTML (Hypertext Markup Language) as the formatting language for web pages. XML (eXtensible Markup Language) is the favored language now, but HTML will continue to be accepted by web browser software for the foreseeable future.

When discussing collaboration and communication this research looks to the near future when most researchers will be facile with the equipment and software. Access problems due to economic or infrastructural limitations are given little attention under the assumption that such problems will be temporary, and also that most scientific researchers are properly equipped. On the other hand, this research does not espouse the most advanced technology, but rather the technology readily accessible to most researchers: access to the Internet and WWW, and a competent desktop computer.

This research focuses on collaboration supported by asynchronous communication. Synchronous communication, especially face-to-face verbal communication, is today considered the "Gold Standard" of communication. This research joins others[13,14,15] in challenging that assumption by showing not only advantages of asynchronous communications, but also the shortcomings of synchronous communications for the purpose of scientific collaboration or learning. The complementarity of synchronous and asynchronous methods is discussed.

> *… communities that combine both f2f [face-to-face] and CMC [asynchronous] systems would be able to bond better and share values more effectively than communities that rely upon only one or the other mode of communication.*
>
> --- Etzioni and Etzioni[16]

The growing tendency of research teams to be geographically, and thus temporally, dispersed is thoroughly discussed. The time geography of these teams is described and used as a basis for the assumption of a growing need for asynchronous collaboration and communication. This research will lead to the definition of an environment for asynchronous distributed collaboration based on existing research and extended with theoretical support. The collaborative team and support personnel will form a "Network of Excellence" enabled by the proposed collaborative environment.

The human behavior exhibited during collaboration must be analyzed in order to design the environment and tools for the new environment. This behavior is studied in psychology, social psychology, sociology, management and political science. Each component of the proposed environment will be understood through scientific realism[17,18]and general systems theory[19,20,21], but is examined in greater depth and less abstraction by means of conceptual frameworks developed by scholars who have studied the embedding system of each component. Each tool developed in this research is considered an artifact cooperating with other tools in the environment and is examined on its own terms. It is shown that the components rest conformably within realism and general systems theory and fall within a common conceptual framework.

> *Social science is the science of social systems.* For this reason, it will have to use the approach of general systems science. --- Ludwig von Bertalanffy[19]

Interaction is the principal measurable attribute of collaboration. In order to evaluate collaboration, I analyze the records of interactions captured by the tools. Each tool having an interactive capability is equipped with programming to store a record of interactions in a log file. The time dimension of interaction can be extracted from the log files. The content of the interactions is permanently recorded in the textual dialog produced by the tools. The content of the dialog is analyzed using two qualitative coding schemes: Bales codes[20], and the structurational argument coding scheme[21].

1.2.2 Conflict with existing methods

The preference herein for asynchronous collaboration methods as opposed to synchronous methods is heretical and needs to be explained. One does not need to look far into the past to find the greatest advance in asynchronous methods since the movable type printing press: the Internet. Scholarship has been built on both synchronous methods, principally face-to-face conversation, and the asynchronous methods of postal services and scholarly journals. The Internet has marginalized the postal services, and will shortly do the same to the hardcopy journals. This dissertation proposes that a similar marginalization, or at least a marked improvement, needs to be made to synchronous methods. Why? Synchronous methods of collaboration have several major defects: they force all participants to schedule participation; they are the enemy of reflection, forcing all participants into rapid response and implied acceptance; synchronous methods are usually verbal, and thus difficult to record; synchronous dialogue or monologue cannot be searched [yet]; and finally, synchronous methods favor the rhetorically skilled and powerful, not always the knowledgeable.

- The emphasis on asynchronous communication is not due to weaknesses of synchronous technology, but rather the suitability of asynchronous dialog to scholarship.
- Asynchronous communications allows permanent documentation of dialog, and access to that dialog.
- Asynchronous communication gives plenty of time to reflect and compose.
- Synchronous communication provides advantage to the rhetorically facile, and it disadvantages those in isolated time zones.
- Use of synchronous dialog cannot be disallowed, but is discouraged unless the synchronous dialog is documented in minutes or archived e-mail.

The emphasis this dissertation places on asynchronous collaboration and communication will shock many scholars accustomed to communicating in synchronous modes. Remember that, at the time of writing, the Internet had only been in widespread use for a

decade, and the WWW for eight years. Before the Internet Age, effective asynchronous collaboration was practically impossible. The ideas presented in this dissertation are in conflict with the collaborative methods most of today's scholars were taught.

Large-scale, long-term collaboration is an unusual idea to established scholars. The goal of collaboration in the past and today, especially in the social sciences and humanities, is the production of a single research paper. This is short-term, small-scale collaboration. The goal of a large-scale, long-term collaboration is learning about the issue domain, the development of theory and hypotheses, and the production of a stream of research papers.

1.2.3 Bias

There is, of course, the presumption of objectivity in any scholarly research. I have thought about my biases and attempted to reduce them, with some success. I do admit to being impatient with the glacial reaction of the scholarly community to the opportunities offered by the new technology. Despite this impatience, I've been able to examine the reasons for the conservatism exhibited by scholars. That examination opened my mind to the beauties of the existing techniques that scholars have developed over the centuries. Thus my designs do incorporate the best of the established methods, and offer some improvements. In other words, the technology is made to enhance the established methods of scholarship, not to supplant them.

The search for objectivity in examination of social issues in computing is, according to Mowshowitz[22], rather hopeless, since the research is motivated by questions of policy. The best one can do is to avoid the excesses of the positions that one must take. The worst danger is the tendency to suppress debate. Since an expanded version of this dissertation will be published on the WWW and every section made annotatable by anyone, I certainly have avoided that problem. Moshowitz outlines five positions that reflect contemporary thought on social issues in computing: technicism, progressive individualism, elitism, pluralism, and radical criticism. The social issue at stake here is

the use of information technology in dialog; in our case the dialog is scientific argumentation.

Technicism is a view of the issue that equates the computer and its software system with an instrument of progress. One symptom of technicism out of control is the tendency to apply the system to problems that do not need it. The view that the issues deserving of research are either exceedingly complex, or very specialized, argues for the use of all the power we can apply to the research process. Yet, for the small-scale, short-term collaboration leading to the publication of a single research paper, current practices are adequate. While I certainly advocate the application of computing technology to collaborative research, I avoid the principal pitfall of technicism**:** placing the social activity under control of the system. The system serves the team, and the individual team members. There are some behavior modifications requested of the members, such as recording the essence of private conversations that contribute content to the dialog. These modifications are nothing more than what is called for in responsible team behavior.

Progressive individualism is a view that doubts the abilities of technical systems, social groups, and human judgment -- all in the favor of individual initiative. The views put forth in this dissertation all favor the ability of each individual to criticize any position held by institutions or individuals. Open individual criticism is seen to be the engine of refinement of research and its products, thus the proposed technical system empowers the individual and simultaneously protects the research team from domination by powerful members. It is my hope that the system serves the researchers without constraining their actions.

Elitism favors either technicians or managers. The danger in this outlook is that there may be controls built into the system that preferentially empower a single group. Certainly the proposed environment has controls built into the system. They take the form of restrictions of the genres that can be employed in the team's documentation.

Every page on the web site is a WWW page. All e-mail in the archives is a WWW page. Fortunately the WWW page is one of the most flexible communication genres ever invented. This flexibility, and the almost universal ability to annotate the documents, saves us from elitism. Furthermore, the proposed system invests technical services and know-how in a facilitator who is subordinated to the scientific leadership.

Pluralism supports both conflict and consensus. The ability of all to participate is restricted by required membership in the research group. The members of the group may represent themselves or others, but are each constrained by several rules of dialogic behavior. While rules are put forward, individuals are also constrained by the characteristics of the system: open criticism is a great leveler, and the asynchronous nature of the dialog reduces the power advantages held by the powerful or those skilled in rhetoric. The proposed environment does not require consensus, nor does it require conflict.

Radical criticism is a position that challenges the other four positions defined by Moshowitz. This view is characterized by an appeal to determinism or devolutionism. Determinism centralizes power in the hands of the technicians and managers and, in conjunction with a powerful system, subjugates the humans that the system is to serve. The proposed environment tends to turn each member into both a beneficiary and a servant. Devolutionism insists on redistribution of social power. While the environment proposed in the dissertation does limit power and hopefully encourage universal participation, it seeks to work within the existing power systems: the ethos of science, the institution, and cultural practices.

1.3 Examples of Large-scale Collaboration

While there are several described examples of large-scale, long-term collaboration, most of these enterprises are organized by funding agencies and concentrate their management on the organizational aspects of the collaboration, leaving the management of the

substantive research collaboration to the teams[23,24]. Many more examples exist, but their public realization is just a public area describing the mission, the members, and the research results. There is certainly research, perhaps in intranets, behind these public sites, but no indication of a collaborative character, or methods employed. I have found no examples of large-scale long-term research sites equivalent to Research Webs.

Collaboration initiated by funding agencies proceeds in a common pattern: first is the identification of candidate topics by an elite group from the management of the funding agency; next meetings are held to recruit scholars and plan the research effort; finally the plan is executed by research teams and research results are published and discussed. The management of substantive research is left to the research teams. In short, the agency provides funding and guidance; but its contribution to substantive research is largely limited to the organization of meetings.

1.3.1 National Center for Geographic Information and Analysis (NCGIA)

The NCGIA was founded in 1988 with a National Science Foundation (NSF) grant. In the last 13 years the NCGIA has supported numerous Initiatives, Conferences, workshops, and curriculum development committees; and has provided financial support for many scholars and research associates. Most of NCGIA's research activities are now under the aegis of Project Varenius.

> *The objective of NCGIA's new research plan, entitled Project Varenius, is to advance geographic information science through basic research, education, and outreach. The research is motivated by scientific, technical, and societal concerns. First, the research serves science and scientists in two ways, focusing on areas in which our knowledge of formalizable geographic concepts is currently incomplete, and contributing to the development and refinement of tools and methods that scientists can use to study geographically distributed phenomena. Second, the research provides basic understanding of geographic concepts, which is required for the production of new technologies. Third, the research examines the impacts that these technologies have on individuals, organizations, and society, and that other digital technologies have in the context provided by geographic space.* --- Varenius Project Description[25]

Like other funding organizations, the NCGIA identifies and debates candidate topics inside a select group of scholars. This group contains the geographic information scholars of the invisible college of Geography, many of which are from the NCGIA. The candidate topic is then discussed in specialist meetings where the focus is sharpened and plans are made for funding and promotion of research outside the NCGIA. Research is conducted by participating scholars and is reported and discussed in workshops, seminars and in the literature. No provisions are made by the NCGIA to facilitate substantive collaboration at the detailed level.

1.3.2 Cochrane Collaboration

This very large and successful distributed enterprise is devoted to the establishment and practice of evidence-based medicine. The principal product of the Cochrane Collaboration is a set of systematic reviews of clinical trials that examine specific health problems or intervention practices[26]. While the work is nowhere near complete and will take decades to provide the information all physicians and health care workers need, the Collaboration is a very active and growing concern.

It is in the Systematic Reviews where collaboration takes place. Each Review is prepared by a group of collaborating authors called a Cochrane Reviewing Group. The

work of the group is very highly formalized in order to maintain a rigid set of quality requirements. The formal methods are published in *The Reviewer's Handbook*[27]. The methodology that guides the work is complete and well established; but there are, in the implementation, many opportunities for criticism and discussion.

The Reviewer's Handbook does not specify methods to resolve disagreements or how to engage in collaboration. The Review Group is required, however to document the methods used to make decisions on selection of studies and how they resolve disagreements. The Handbook suggests using outside colleagues to help identify unpublished studies that might be examined for inclusion.

When one examines the details of how the Review Group must go about its business, it becomes clear that many of the tools and methods used in the Research Web could be applied to the Systematic Reviews. For instance, the Handbook suggests the use of Procite, a personal bibliographic manager, to establish a bibliography of studies that may be examined. The Annotated HyperBibliography, an important tool of the Research Web, is based on a personal bibliographic manager. Use of the Annotated HyperBibliography would allow very rapid access to the studies, with their abstracts (usually structured abstracts), and perhaps full text accessible at a click; and would provide the reviewers with an annotation area that might serve as an information base sufficient to exclude or include a large portion of the identified studies. As the study progresses several sections of the final Review will be drafted. DocReview, the critical apparatus of the Research Web, could serve as a tool to attach annotations to the draft for use in writing later editions. The information repository function of the Research Web would serve to organize and store much of the documentation, such as formal letters of request for information.

It appears that the Cochrane Review Group is exactly equivalent to the Research Web's authoring team. The Cochrane Review itself could be DocReviewed after release, thus

opening the product to annotation and suggestion from the user community. Since the Reviews have a mandated peer review prior to release, that would be another use for DocReview. Bero and Rennie[28] point out many mandated demands for review and make a very strong case for continual review and updating. Presenting the Cochrane Review as a RW Essay, a hypertextual augmentation of research reports, would be quite appropriate; but their use would require some changes to the document formatting and methods of distribution (the reviews are currently distributed on CD). The hypertextual nature of the RW Essay would allow direct linking of the Review to its data.

1.3.3 MacArthur Foundation: Research Networks

In 1980 the MacArthur Foundation embarked on the establishment of "an experiment in scientific organization." Following two years of study, two Research Networks were established to study facets of mental health. The constellation of Research networks reached sixteen in 2000[29]. The Research Networks are organizations that support interdisciplinary long-term collaboration in research domains of particular interest to the Foundation[30].

The management of a Research Network is very flexible, featuring a close association of a scholar-representative of the Foundation and the Network Chair, who is selected by the Foundation. Members of the Network are selected solely on the basis of their scholarly potential, not on any institutional or geographic affiliation. Funding from the Foundation is intended to support the infrastructure: seed money, administrative support and bridging. The Members are expected to acquire funding for the substantive research through the normal grant proposal process.

During early organizing meetings, the scope of the issue domain is settled. The issue domain must be specific enough to attract members intellectually, but general enough to provide room for individual growth. The Networks have a *core group* of between seven and sixteen member who are organized into *subgroups* formed on the basis of research

interests, rather than discipline, institution, or geographical region[31]. Membership in a subgroup is quite stable, though not permanent or exclusive. Networks also create *temporary working groups* designed to fulfill a task defined by a subgroup. These groups usually include investigators outside the core membership. Some temporary working groups may be organized to perform pilot studies, or exploratory research under the direction of one of the members of a subgroup.

The methods for substantive collaboration are left to the members of the Network, and what exactly occurs in that collaboration has not been described. There is fundamental agreement with the organization and funding of the Research Networks and the collaborative research methods of the Research Web. The working groups of the Research Networks correspond exactly with the authoring groups of the Research Web. The Network Chair corresponds to the RW's Scientific Coordinator. The Core Group corresponds to the RW's Principal Investigators, plus the Scientific Coordinator.

Placing the responsibility for modeling with the Network Chair could easily rectify the absence of modeling as a principle in the MacArthur Research Networks. As the Network chairs have administrative assistants assigned to remove the details of management from their shoulders, so might a modeler be provided to remove that task from the leadership by taking information directly from the temporary working groups. The modeler could also be assigned to assist the working groups with the detailed models of their subdomain.

1.3.4 Daimler-Benz Foundation: Ladenburger Diskurs and Kollegs

The Gottlieb Daimler and Karl Benz Foundation was founded in 1986 with the intention of examining the interrelationship between humankind, environment and technology. To that end, the Foundation has established a 'think tank', the Ladenburger Diskurs, to identify topics for further examination in actual research to be carried out by interdisciplinary research associations, the Ladenburger Kollegs. There have been seven

Kollegs formed, five of which have completed their work. One of the recently completed Kollegs studied organizational learning in rapidly changing environments[32]. This Kolleg had seventeen projects lead by 40 social scientists, managers, consultants and administrators from 11 countries.

The Ladenburg Diskurs establishes potential topics, recruits researchers, and submits a proposal to the Foundation. If the Foundation accepts the proposal, a Ladenburger Kolleg is formed. The Kolleg and its projects are funded by the Foundation and the Foundation provides facilities for several meetings of the Kolleg at the Foundation's Estate in Ladenburg. There are no provisions made for collaboration at the project level, as the Foundation is concentrating its efforts on the introduction and implementation of communication processes and publication of findings[33]. Based on the history of the Kollegs, it is also apparent that most Kollegs are designed to have a determined lifetime and specified output. Research Webs, on the other hand, have an indeterminate lifespan and do necessarily have predetermined products, except for the models of the issue domain.

1.3.5 What's Missing?

Research Webs need a home, a virtual location that would be the equivalent to a research laboratory or Institute. For social science issues, the Institute model, with its physical collocation property, is simply incompatible with the spatially fragmented nature of academic research. Distributed institutes are unlikely to be formed simply because it is the nature of Universities to want to "keep the money at home." Favoring researchers at a given University flies in the face of an imperative to recruit scholars who are qualified, available and committed to study of the issue domain.

Where is the model for a large-scale, long-term, well-funded, distributed research community? I see two close approximations: Research Webs administered by endowed chairs and the MacArthur Research Networks. In both of these models the central

administrator is a respected scholar, and is free to recruit anywhere. Since the non-labor expenses of running a Research Web is small, existing University infrastructures will support the technical requirements. Labor in both cases is provided by grants obtained by members of the research team. Seed money is provided by either the endowment or the Foundation.

1.4 Geographical Aspects of the Research

There are several geographic factors that influence this work: the potential application of the concepts to academic research in geography; the influence of spatial distance on collaborative research teams; and the effects of geographic differentiation, especially cultural, linguistic, and economic differentiation.

1.4.1 The Benefits of Collaboration for Geography

Morrill, while president of the American Association of Geographers, stated in his Presidential Address[34] that, " if geography is a meaningful part of knowledge seeking and if it deserves to survive, it has to concentrate on creating a coherent body of theory that others recognize as significant." He suggests that research can best be strengthened in centers of excellence devoted to the development of theory in single issues for long periods of time. Such centers would concentrate intellectual power into a critical mass, enable much-needed collaboration and feedback among the participants, and would support and encourage visiting scholars and postdoctoral students to become the next generation of research leaders. Morrill's speech was given on the cusp of Internet acceptance, so was cast in the physical presence mode of Institutes. This dissertation suggests the application of the new technology of the WWW to create "cyber-places" that will defeat the tyranny of distance within the very discipline that is based on distance. Not only will cyber-places provide immense savings in time and travel expenses; but they will enable a wider audience to participate, usually from their own offices or homes and on their own schedules, changing a research setting that Morrill characterized as having "astonishingly little feedback or interaction."

Most of the sub-disciplines of Geography are international in scope. Systematic studies such as transportation and economic geography are taught and practiced all over the world. While international collaborations to write journal articles and edited books are common in geography, there is little evidence for collaboration on a larger scale. Since most of these collaborations are joined to produce a single paper, they usually dissolve after the paper is accepted. Occasionally a set of authors collaborates on a series of papers over the years, but usually the collaboration is short-term and limited in scope. A notable exception to small-scale collaboration in Geography is the National Center for Geographic Information and Analysis (NCGIA) initiatives. These initiatives are cooperative agreements between scholars and teams of scholars from many institutions. There are elements of collaboration and coordination at the more abstract levels of their broad issue domains; but true scientific collaboration is present in some specialized projects within the initiative's framework. The initiatives are more "incubators" of collaborations than they are themselves collaborations.

Relatively large-scale and long-term collaborations are common in the physical and natural sciences (e.g. the LTER, Long Term Ecological Research), less common in the social sciences, and nearly absent in the humanities[35,36]. These scholarly collaborations assemble around the physical presence of large machines in physics, and around well-funded laboratories in medicine and biochemistry. The poorly funded social sciences, including geography, could join in large-scale long-term collaborations by creating their laboratories in cyber-places called Research Webs (see Chapter 3).

1.4.2 The Influence of Geographic Distance on Scientific Collaboration

The international character of science mentioned above is not the only factor putting distance between scholars. The economics of higher education are such that in many academic departments there is only one position open for the specialist[37,38]. The hiring of single specialists will provide the instructional coverage, but will not create the critical

mass for research work[39]. There are even elements of competition for students – a department may choose to avoid a specialty if a nearby university is serving the local needs well. Thus even within a given country the specialized scholars are dispersed. Distance effects are expressed as communication expenses and quality impacts. Unless the scholar is well equipped and facile with communication methods other than face-to-face conversation, the communication is severely degraded and collaboration unlikely. The cost required to participate in face-to-face meetings is often too great for the limited budgets of most social science collaborations. The airline fare alone for one intercontinental trip is more expensive than all the hardware and software necessary to equip a researcher for computer-mediated collaboration over the Internet!

Another important effect of geographic dispersal is the difficulty of coordinating synchronous communication. In countries with multiple time zones, like the United States, Canada, and Russia, in each day there are at most only a few hours when one can expect colleagues to be in their offices simultaneously. When one adds a few inevitable time commitments and a mid-day meal, telephone calls are difficult to complete unless scheduled in advance. Conference calls spanning several time zones are virtually impossible to arrange without major inconvenience to at least one party.

Travel and communication expense are avoided if the collaborators communicate over the Internet. The time zone differentials are defeated if the bulk of communication occurs asynchronously. If collaborators are willing to learn a few new methods, distributed asynchronous collaboration over the Internet through a Research Web can support most tasks of communication and collaboration.

1.4.3 The Influence of Geographic Differentiation on Collaboration

Collaboration, the creation of new shared knowledge, is profoundly influenced by the heterogeneous nature of culture, politics and economic well-being. The inclination to collaborate is profoundly influenced by the culture of the society. Within a given pluralistic nation, religious and ethnic communities, and class differences create a mosaic of attitudes affecting collaboration. For instance, in British Columbia, some native North Americans practice "information bartering"[40]. Asking a question is likely to be met with evasion unless some information is offered. The pervasive influence of Confucianism in Asia has been cited as an obstacle to the practice of science[41,42]. National character is yet another layer of attitudes[43]. For example, in the United States individualism is instilled into the children, especially males, of the society at an early age and reinforced by the national mythology.

Language is not as great a difficulty (for anglophones) as it was in the early 20th century. English is now the de facto language of science[44] and most international conferences use English as the working language. Of course in many specialized areas such as regional geography knowledge of the local language is a powerful requirement. English is not the native language for most of the scholars of the world, thus when non-native speakers communicate written English is often preferred to spoken English[45]. Even a small time lag allows the writer to reflect before sending and the reader time to study the message reflectively[46].

The political attitudes toward communication technology have recently been in turmoil. Religious fundamentalism has caused legislation restricting Internet access in many Muslim countries. In the United States, "Computer Decency Acts" have been placed into law, and then overturned by more liberal courts. In Germany and France, the presence of Ultra-Nationalist and Neo-Nazi WWW sites has generated a great deal of political debate. The People's Republic of China has many restrictions on Internet activity and no stable policy[47].

The distribution of wealth between nations and among their citizens has a profound influence on the ability to collaborate. Very poor nations do not have the communication infrastructure to support access to the Internet, even assuming that the prospective collaborators have the necessary equipment. International consortia sometime fail to recognize the needs of their poorer partners[48]. Though computers capable of accessing the Internet are routinely discarded in the developed world, the means to distribute them to the needy are not well developed.

National telephone monopolies sometime charge exorbitant prices for Internet access, leading to "internet strikes" in Brazil, China, France, Germany and elsewhere[49]. Even within the United States, Internet access costs are extremely variable. All aspects of the access situation are extremely fluid. For instance, the Village Internet Program of the Grameen Communications and Grameen Foundation USA is beginning to provide "Cyber Kiosks" to village entrepreneurs in Bangladesh[50]. Wireless communication technology may alter the economics of telephony to the advantage of the currently disadvantaged.

1.5 Contribution to Knowledge

This dissertation describes the collision between the new information technology and the established methods of collaborative scientific research. The opportunities presented by the Internet and associated technologies offer research teams powerful new methods of managing collaborations. These methods have a profound impact on every phase of research activity. It is not technology alone that disturbs the equilibrium of the past; management of the collaborative effort is confronted with new challenges. The behavior of collaborators is also in need of reconstruction as the old methods are replaced with the new.

The impacts of the Internet and the World Wide Web are so new that software is just now emerging that will capitalize on the capabilities of the technology to solve old problems or to facilitate collaborations. I present a coordinated suite of web-based tools that were designed and implemented to facilitate research collaboration in the new information environment.

A concept called the Research Web (RW) is presented. The Research Web is a generalized environment for asynchronous scientific collaboration in the new information environment. The RW has a baseline configuration of tools useful to any research team. The nature of the research topic will determine any additional specialized tools that the collaborative team needs. The concept is flexible and can meet almost any needs. The unifying technology is the WWW -- every document is accessible as a web page, and web-based forms guide most interaction.

A new research communication genre is presented. This genre, the Research Web Essay, makes research papers interactive. The reader can annotate the paper publicly, thus turning the paper into a means for collaborative dialog within the research team or the "invisible college." Marginal notes can once more be used by authors, rather than footnotes (which have no meaning in a scroll document) and endnotes (very inconveniently placed). Citations can be clicked to present bibliographic data, including abstracts. The bibliographic data window contains paths to allow the user to annotate the reference, and to display full text if available. Definitions of terms used in the Essays can be displayed from a glossary at a click. The user may annotate the definition. Ancillary documentation such as images of proper nouns and notes to team members may be popped up at a click. References and footnotes are included at the end of the document and include return links to all places cited. The reader can then quickly see the context in which a work is cited.

The critical apparatus of the RW Essay is presented in DocReview[51] (see §4.3). DocReview is an annotation program that allows the reader to attach an annotation to a predetermined segment of text or to images by clicking an icon, an orange colored W. Clicking an orange colored R allows reading existing annotations. This annotation process is the principal means the research team has for refining working documents. Any document may be DocReviewed and, in a collaboration, most should be. Meeting minutes, position papers, agendas, proposals, specifications, and even spreadsheets have been DocReviewed with success.

The tools and concepts presented herein were applied to several collaborations with mixed results. The collaborations are described and analyzed and the results are used to develop a series of recommended practices. The nature of the collaborative team is investigated. The roles that members need to play (collaborator, facilitator, scientific lead, research assistant, etc.) are discussed.

While the target activity of this dissertation is scientific research, the arguments presented and the tools described will apply to a large variety of group activities. With modifications and incorporation of tools that match the problem domain, activities such as public participation in decision-making can be supported. Problem-solving activities such as planning in engineering and architecture can benefit from the Research Web environment. The constants in the arguments presented herein are:

- The collaborator is given the freedom to schedule personal participation.
- The asynchronous format allows the collaborator to reflect while reading, to reflect on a response, and to reflect on the phrasing of the response.
- All documentation and dialog is permanently recorded.
- All documents are presented as WWW pages, a universally understood genre.
- The entire web site, documents or e-mail, is capable of being searched on line.
- Access to the collaboration can be made from any Internet enabled computer anywhere in the world.

1.6 Organization of Dissertation

The conceptual framework for this work, a triad of foundations, is presented in the following Chapter. The proposed new research environment, the Research Web, is presented in Chapter 3. Chapter 4 discusses the tools developed for or planned for inclusion in the Research Web. Empirical field studies of three Research webs and of 100 document reviews using DocReview, the critical apparatus of the Research Web, are found in Chapter 5. Conclusions and directions for future research are presented in Chapter 6.

Notes to Chapter 1

[1] Hendricksen 1994

[2] Price 1963

[3] Merton 1973

[4] Thagard 1997

[5] William Bevan 1989, 5

[6] Niiniluoto 1995

[7] Walsh and Bayma 1996a, 667

[8] Jerome J. McGann 1997, 40

[9] Funtowicz and Ravetz 1993

[10] Ziman 1976, Chapter 5

[11] Weimer 1979, 78

[12] Latour 1987

[13] Etzioni and Etzioni 1997

[14] Sudweeks and Rafaeli 1996

[15] Etzioni and Etzioni 1999

[16] Etzioni and Etzioni 1999, 247

[17] Aronson, Harrè and Way 1995

[18] Sayer 1992

[19] Ludwig von Bertalanffy 1968, 195

[20] Bales 1950

[21] Meyers, Seibold and Brashers 1991

[22] Mowshowitz 1981

[23] Schade and zu Putlitz 1996

[24] Kahn 1993

[25] Goodchild *et.al.* 1997

[26] Bero and Rennie 1995, 1935

[27] Clarke and Oxman 2000

[28] Bero and Rennie 1995, 1937

[29] MacArthur Foundation 2000

[30] Kahn 1993, *passim*

[31] ibid., 17

[32] Anon 1998

[33] Schade and zu Putlitz 1996, 279

[34] Morrill 1987

[35] Thagard 1997

[36] Endersby 1996

[37] Walsh and Bayma 1996b, 346

[38] Koku, Nazer and Wellman 2000

[39] Pickering and King 1992, 359

[40] Hébert 1986

[41] Tsou 1998

[42] Hsü 1992, xxii

[43] Hofstede 1991

[44] Garfield 1967, 19

[45] Sanderson 1996, 107

[46] Friermuth 2001, 176

[47] Xiaoming, Zhang and Yu 1996

[48] Cohen 2000, 2157

[49] New York Times, Circuits, 14 Jan 1999

[50] Yunus 1998

[51] Hendricksen 1998