# Whisper: Supporting Knowledge Transfer for Academics

Mark David Slater
Department of Computer Science
University of California, Santa Cruz
Santa Cruz, CA 95064

August 1, 2005

**Abstract**

Today, despite decades of research, computer support for academic knowledge workers is fragmented and poorly integrated. While office automation and other forms of Computer Supported Collaborative Work (CSCW) approaches have benefitted academics, no single environment today integrates the basic activities of the academic knowledge worker, including: individually or collaboratively writing research papers, sharing research papers, reviewing papers for publication in a journal or conference, and persistently sharing comments and observations on existing literature. We have developed an initial model that describes the mechanisms used by academics to transfer knowledge to other individuals, project groups, and communities, as well as mechanisms used to transfer knowledge within project groups and communities. Based on our model, with revisions stemming from the results of our study, we will develop a novel environment called Whisper (for Web Information Sharing Project), that will support the mechanisms of knowledge transfer between academics. Whisper will provide tools for academics, including digital library spaces with shared articles and annotations, versioned file spaces, archived discussion lists, and community journals.

By creating an online workspace for academics of all fields, Whisper can reduce the cost and effort needed to setup shared workspaces between physically distant collaborators, as well as interdisciplinary collaborations. The persistent repository of project information, community discussions, and online journals will provide easier access to past and current research; this helps facilitate knowledge transfer long after the entity that created the knowledge is defunct or becomes unavailable. In addition, our model will be able to identify: the knowledge transfer mechanisms most appropriate for a given entity pair, the point at which formal communities are viable, and at what point mass knowledge transfer mechanisms are viable within a given community.

## 1 Introduction

Today, despite decades of research, computer support for knowledge workers in academic is fragmented and poorly integrated. While office automation and other forms of Computer Supported Collaborative Work (CSCW) approaches have benefitted academics, no single environment today integrates the basic research activities of the academic knowledge worker, including: individually or collaboratively writing research papers, sharing research papers, reviewing papers for publication in a journal or conference and persistently sharing comments and observations on existing literature.

Consider a typical collaborative conference paper. It begins as a file that is exchanged multiple times via email as each author makes contributions. The paper is then uploaded to a web-based review system. Upon acceptance, the connection between the reviews, the paper source, and the final paper version is broken, since the camera-ready paper is submitted to the publisher's own web-based paper manager. The final paper is then sent to an institutional digital library. Researchers download papers from the library, storing them in a folder on their local disk, where they are difficult to search, and are completely dissociated from

their bibliographic metadata. The paper has now crossed five system boundaries, effectively eliminating any possibility of advanced collaboration such as shared annotations, and shared collection management.

There is a research opportunity to uniformly support knowledge transfer mechanisms in academic communities. A model of current mechanisms can be created based on surveys of academics, personal observations, and informal discussions. Using this initial model, an online knowledge workspace for academics can be created to better support the identified mechanisms of knowledge transfer. User studies and analysis of usage metrics will provide additional details and validation of the knowledge transfer model.

We have developed a preliminary model of mechanisms used by academics to transfer knowledge to other individuals, project groups, and communities, as well as mechanisms used to transfer knowledge within project groups and communities. A short study consisting of surveys and interviews with academics, from undergraduates to professors, will provide us with an initial verification and sanity check of our model. Based on our model, with revisions stemming from the results of our study, we will develop a novel environment called Whisper (for Web Information Sharing Project), that will support the mechanisms of knowledge transfer within the model.

Whisper will provide tools for academics at three different levels, including:

- Individual: A personal digital library space with shared articles and annotations, versioned file space, and WebLog
- Project: A project digital library space with shared articles and annotations, versioned file space, archived discussion lists, wiki, and artifact management
- Community: A digital library space with shared articles and annotations, community journals, archived discussion lists, wiki, and conference management

The project will contribute a detailed and validated model of knowledge transfer within the domain of academic research, including a set of supporting tools that are mapped to each mechanism. The Whisper environment will provide academics with a validated, integrated toolset for performing knowledge work, organizing their knowledge, and integrating knowledge creating organizations and people. Whisper itself, via its adoption or lack thereof, will test the hypothesis that the integration of individual, project, and community knowledge flows in a single system will yield substantial improvements in the efficiency of academic knowledge transfer.

By creating an online workspace for academics of all fields, Whisper can reduce the cost and effort needed to create shared workspaces between physically distant collaborators and interdisciplinary collaborations. A persistent repository of project information, community discussions, and online journals will provide easier access to past and current research; this helps facilitate knowledge transfer long after the entity that created the knowledge is defunct or becomes unavailable. In addition, our model will be able to identify: the knowledge transfer mechanisms most appropriate for a given entity pair, the point at which formal communities are viable, and at what point mass knowledge transfer mechanisms are viable within a given community.

## 2 Objectives and expected significance

Knowledge is created at every level of academia, including individuals, project groups, and research communities. For knowledge to have utility, it must be transferred from one mind to another, for only when knowledge is instantiated in someone's mind can it be used as the base for future action and thought. Our project seeks to develop a complete model of the mechanisms employed in academic knowledge transfer and, using that model, to improve the knowledge transfer process.

Ideally, knowledge flow among people is frictionless, akin to drops of water on a teflon pan, where ideas are transferred among people quickly, with minimal effort. In reality, there is friction in the flow of knowledge, causing knowledge flow to be slower and more laborious, sometimes not occurring at all. Most academic knowledge workers make use of the same toolset: e-mail client, web browser, word processor, spreadsheet, and presentation authoring. It is not our purpose to replace or recreate these tools, but rather

to give academics a standard method for collaboratively creating knowledge that cuts across tool vendors and operating systems; once knowledge has been created or captured with these tools, we seek to improve the transfer of this knowledge to its intended audience.

The friction generated during knowledge transfer can only be reduced once the sources of that friction have been identified. When two people collaborate on a paper using email to exchange successive versions, friction takes the form of lost changes and the inability to safely work in parallel. Completed papers are frequently held to a page limit for publication; friction here results in less detail, and possibly little to no information about basic assumptions or failed experiments along the way. The business model of most academic journal publishers limits the number of papers that can be published in any one volume, and friction appears in the delay between a paper's acceptance and its publication, sometimes lasting years. If a new researcher joins a project in mid-stream, friction is the time lost as one or more existing team members acclimate the new member. Likewise, when starting work in a new discipline, friction is the time spent identifying the most salient works in that field and understanding how they fit into the context of the current research. Friction is not limited to knowledge transfer between people; it can also arise in the time and mistakes made when synchronizing working documents between computers, and materials forgotten on a computer inaccessible from the researcher's current location.

The most complex problems facing science and society today are more likely to be solved at the intersection of two or more disciplines rather than any one discipline. In recent years, the depth of knowledge needed to succeed in a chosen field has required practitioners to become more specialized, leading to fragmentation. This fragmentation leads to multiple related disciplines that are less integrated while making it more difficult to pursue cross-cutting research as required knowledge is spread among tens of journals and conferences. Even the best search engines cannot reliably gauge the quality of the documents they return; flawed documents may be cited more often than high-quality research as members of the field rush to correct the errors. Without tools designed to address knowledge flow within and between disciplines, academics will find it increasingly difficult to function effectively in their field; the challenge for knowledge workers should be the creation of new knowledge, not keeping abreast of recent developments.

The hypotheses we will explore in the Whisper project are as follows:

- Explicitly modeling academic information flow pathways and then identifying aspects of friction within these flows is a useful framework both for understanding inefficiency in academic knowledge transfer, and for identifying targets for improvement in these flows.

- A web-based information system we call "Whisper", which integrates substantial portions of the process of academic knowledge creation and knowledge transfer flows will dramatically reduce flow friction for represented pathways and improve the efficiency of knowledge exchange among its users.

To this end, we have constructed a preliminary model of information flow among academic knowledge workers across a broad range of academic disciplines, as shown in Figures 1 and 2. Figure 1 shows the subdivision of mechanisms we will use to map academic knowledge transfer paths. Figure 2 provides a rough demonstration using the life-cycle of an academic paper, including both the knowledge that goes into it and the knowledge produced by the process. For each case of knowledge transfer, we can then identify the tools commonly used by academics in each of the flows, and the associated friction. For example, discussion between project group members frequently takes place via email; these discussions are not explicitly associated with the project, making them difficult to find later unless project-specific searchable keywords are used in each message. The project repository is frequently duplicated across every computer of every project participant, and even in groups of technically proficient participants, there is frequently no master reference for the latest version of the project files; solo projects face this problem as many researchers now work on multiple computers.

A substantial fraction of the effort in this project involves the construction of Whisper, a web-based knowledge workshop intended to integrate substantial aspects of the information flows depicted in Figures 1 and 2. Based on these initial models, we have determined an initial feature set designed to reduce the most significant areas of friction for people, projects, and communities. Whisper will provide each user, project group, and community with a digital library in which relevant papers can be stored with associated notes and

discussions; users and projects will each have a versioned file repository for their data files and current work; communities will be able to create digital journals to provide peer-review and publication of relevant research within the community. The goal of Whisper is a substantial augmentation of our ability as academics to create knowledge, and have more efficient and targeted mechanisms for accessing knowledge relevant to our daily tasks.

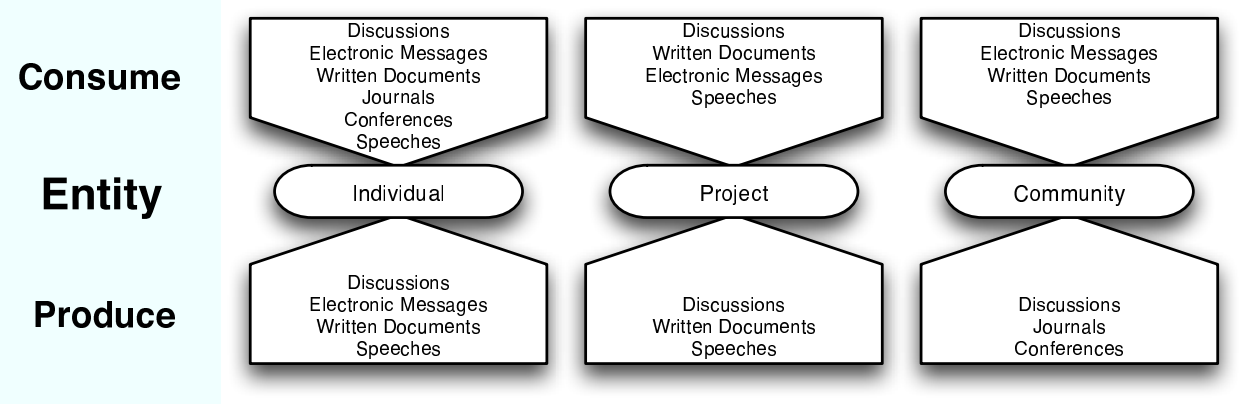


Figure 1: Knowledge Transfer Mechanisms

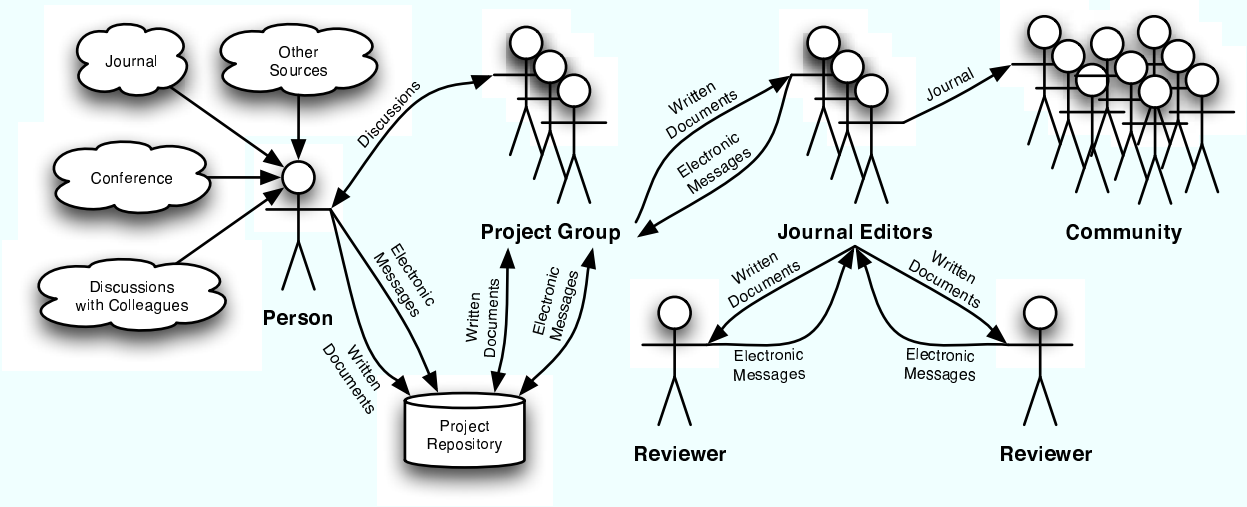


Figure 2: Example Knowledge Transfer Paths

The work promises to make multiple contributions to our knowledge of collaboration and academic knowledge transfer.

- A detailed model of academic knowledge transfer and a characterization of frictional inefficiency in the model. The model will provide the focus points needed to reduce or eliminate the friction affecting knowledge transfer. This model will be validated using surveys of potential users, case studies of those users, and analysis of user activity within Whisper.
- Validated improvements to selected knowledge creation activities. Software engineers have developed version control tools and project management techniques to aid collaboration in software projects;

outside of technical fields, these tools are largely unknown. Even within technical disciplines, many researchers prefer using email to collaborate on papers, even when more sophisticated tools are readily available [29]. Services such as Yahoo! Groups [36] provide shared file space, but have no support for file versioning or status tracking. Our selected toolset will be validated by comparing group projects that use Whisper with those that use more traditional approaches; projects will include both student group projects for a class and longer term research projects at the graduate level.

- The use of networks of relationships to support knowledge exchange. Whisper will allow users to identify those researchers they collaborate with, mentor, or learn from; these relationships can be used to provide greater access to knowledge within each user's personal space, such as their digital library. Users are likely to accumulate close relationships with people that share their research interests; by searching for knowledge within their relationship network, they will quickly find high-quality results, including formal papers, associated notes, and informal discussions related to their search terms. An analysis of user activity within Whisper, along with task-oriented usability studies will help validate this aspect of the project.

- Processes and infrastructure for dramatically reducing the effort required to create online journals and other peer-reviewed information flows. The use of a peer-review to vet new knowledge before publication is rare outside academia, and there are relatively few tools to support the process. Whisper's community tools will enable users to create any number of digital journals and provide discussion forums for every published document. Each journal will support a full peer-review process and allow the publication of papers of any length, as well as links to papers published elsewhere that are relevant to the journal's theme. We anticipate this feature being used to create new online journals, as well as the possibility of supporting the online aspects of existing print journals. The use of the community journals as a knowledge gathering resource will be validated by analyzing user activity within Whisper.

- The use of informal knowledge capture to spur new collaborations and project development. While the focus of most academic research is to produce a formal document that describes new knowledge, informal knowledge, such as a researcher's thoughts on the interdisciplinary impact of a new development or questions that arose while reading a paper, can provide the impetus for new discussions. This will give rise to more collaborations that cross disciplinary and institutional boundaries as users who would otherwise not know of shared interests find each other. An analysis of Whisper's user activity and interviews with users will validate the informal knowledge capture tools provided by Whisper.

Whisper's impact on the academic community has the potential to mirror the impact SourceForge [32] has on the open source community by providing open source projects with the integrated toolset and infrastructure they needed. Students benefit from the personal digital library, where they can keep lecture materials and their associated notes, as well as the project features, such as the versioned file space for both solo and group efforts. In addition, professors can create a course project and place reading material in the digital library, allowing associated discussions about each item to develop within the library. Through the networked relationships, Whisper users will automatically create a high-quality, searchable repository of research relevant to their work by opening their personal digital libraries to their collaborators, colleagues, and students.

Whisper's basic framework will provide administration, user social network management, and security; this framework is already under development. The next stage of development will focus on digital libraries for individuals, then support for projects, including versioned file spaces and message boards, followed by support for communities, including peer-reviewed digital journals. By incorporating existing open source tools, such as Jakarta Slide [24], Subversion [34], Apache Axis [3], and the Spring Framework [33], we will be able to create a compelling system in a relatively short amount of time. Following the release of Whisper's community feature set, we will conduct a series of user studies that will follow students involved in class projects, graduate students performing thesis research, and collaboration between geographically disperse researchers; each study will contrast a set of participants using Whisper for their work with a similar set using more traditional methods.

# 3 Activities to be undertaken

A significant activity for this project involves the creation and deployment of Whisper, a Knowledge Workshop for academics: students, teachers, professors, and researchers. Whisper's deployment will be gradual, beginning with a single lab and approximately ten regular users. As more features are completed and tested in the lab setting, Whisper will be made available to other labs, followed by an entire university department, and eventually labs and institutions around the world. Users will be encouraged to provide feedback related to the existing features, including how well they work (or don't), and prioritizing new feature development.

## 3.1 Whisper Features

Whisper's initial feature set is derived from an analysis of the knowledge flow presented in Figure 2, as well as the personal experiences of its developers. We plan to conduct a small survey within the UC Santa Cruz School of Engineering to determine what, if any, features are missing, and the initial feature prioritization. Whisper's features fall into three categories: those for the individual, those for individual and group projects, and those for communities.

**Initial Surveys.** Before we begin implementing Whisper features, we will conduct a survey of potential users from each of three categories, but focusing on those in a university setting. Whisper's target users fall into three categories: teachers, students, and other researchers. The teacher category includes university professors and lecturers, and industry consultants and trainers. Other researchers include government and industrial researchers and workers, as well as hobbyists. The survey will attempt to ascertain the features most important to each type of user; their answers will help to determine priority for feature implementation. In addition to prioritizing known features, we will ask participants to suggest additional areas of their work that might be augmented by Whisper; this will help ensure that we have not missed a significant source of friction in our planning.

**Individual: Needs.** Just as Engelbart started by augmenting an individual researcher, Whisper also begins at the level of the individual user. The individual's daily needs must be met by the system or it will simply go unused. The primary need of academics is access to research-related documents, papers, articles, and technical manuals, as well as a method for managing notes related to each item. Users also need a general file space for items not directly related to a specific research project, informal methods for receiving and publishing general thoughts and ideas, and the ability to easily manage external access to their work. An overview of current and past work, both within smaller project groups and larger research communities, would also benefit users.

**Individual: Digital Libraries.** As Licklider [27] and Engelbart [15] found, much of a knowledge worker's time is devoted to finding and understanding what is already known or what others have done in the past. Rather than keeping stacks of papers, journals, and other periodicals on a desk or in a cabinet, Whisper will give users a personal digital library implemented using the Jakarta Slide [24] project. The library itself will be accessible indirectly through Whisper's web interface, but also through a direct connection using the HTTP [18] protocol with WebDAV [21] extensions; the direct connection will allow the user to "mount" their library space on a personal computer as a network drive. Each item in the library can have bibliographic metadata and any number of comments attached to it. Comments are threaded and the user can allow other people to both view and add comments associated with each item. The items will also have multiple keywords associated with them. Keyword paths automatically organize documents based on an ordered, user-defined keyword hierarchy; in this way, library items can appear in multiple collections at the same time. At each level of the keyword hierarchy, users will be able to specify their preferred order of the keywords; this order will be used as the default "natural" order when sorting items at that level. There will be no restrictions on the type of data that can be stored by the library. Users will be able to perform textual searches of the (text-based) library documents, comments, and other metadata. As the number of documents within the network of Whisper installations grows, users will also be able to search for related documents within the community spaces; the system will determine how closely documents are related based on common references, citations, keywords, and other metrics.

**Individual: Versioned File Space.** Vogel, et. al. [35] noted, "people usually do not create heavyweight shared workspaces if they do not know that a simple chat or email will grow into a larger, more complex, more formal collaborative work process." Subversion [34], an open source project that implements the DeltaV protocol [10], will be incorporated into Whisper to provide versioned file spaces. These files spaces anticipate the gestation of full-fledged projects from just a few files in one user's file space. Whisper's web-based interface will allow users full access to Subversion's functionality, including checking files in and out, examining older versions, and differences between versions. Standalone Subversion clients, available for all major operating systems, will enable users to keep working copies in a "sandbox" on their local computers; this will allow users synchronize their workspace on multiple computers, as well as continue to work without an available network connection.

**Individual: WebLog.** Many research projects begin life as offhand comments, informal discussions, private musings, or short experiments. Moreover, as the explosion of WebLogs (or blogs) [7] has shown, people like to share their opinions, thoughts, and ideas with the world. Users will be provided with a WebLog space, including access controls on readership.

**Individual: Security.** While the end goal of researchers is to share their results, that does not mean all information in the system should be available to all users. Security should be a primary concern in any groupware application, but it should also be easy for users to understand and manage. Whisper incorporates a security management implementation based on access lists, user roles, and user relationships. Role based security has been previously implemented in other systems [12]; each user is assigned one or more roles, such as project member, project manager, administrator, or guest, and their access to each file is granted or denied based on their role. This simplifies the administration of access rights because, rather than configuring access for each individual, access can be granted (or denied) for all people in a specific role, leaving only exceptions to be handled on a case-by-case basis. Whisper adds to this by capturing relationships between users and allows users to control access based on their relationship to others, as well as their roles.

We have identified four types of relationships: Mentor, Student, Collaborator, and Colleague. These academic relationships are not captured by existing social networking tools such as LinkedIn [28] or Orkut [31] . Mentor and student capture both academic pedigree, such as a professor and graduate student, as well as informal mentoring, such as a teaching assistant and student, or graduate student and undergraduate student. Collaborators are people who work closely together, often on the same projects or within the same laboratory group. Colleagues are people who are familiar with each other but do not have ongoing joint projects, such as professors with different research interests within a department or researchers working in the same field but at different institutions.

**Individual: Dashboard.** As people frequently have a number of ongoing projects, and participate in a variety of research communities, it can be difficult to keep track of everything. The user's home space within Whisper provides a "dashboard" overview of their work. This overview highlights new relationship requests from other users, activity within the user's active projects, and community related events and activities. Project activity might include new messages or replies, new project members, project event changes, and files that are new, modified, or deleted. Communities activity highlighted in the dashboard includes new or changed events, new journals, and community governance changes.

**Project: Needs.** Most academic research is conducted within a project, where one or more individuals have a stated goal to reach or question to answer. Projects provide members with a shared workspace to store project data and files, as well as a project digital library with functionality identical to that of user's personal digital libraries. As noted in the past [11, 17], it can be difficult to capture the small decisions and assumptions made by all project members as work progresses. Whisper provides three methods for recording the history of a project: a message board, wiki, and per-file commentary. Whisper also provides for project administration, events, and access controls.

**Project: Versioned File Space.** The project space revolves around a version controlled file repository managed by the Subversion tool [34]. As with a user's personal file space, the project files can be checked out into a local work areas on their computers. Changes made to the local files are not visible to other project members until the user performs a "check-in", allowing two people to edit the same file at the same time. Subversion detects these conflicts and forces the second user to examine the file in and resolve the differences

before the check-in operation is allowed to proceed. This conflict detection and resolution largely prevents one person's work from being lost or overwritten by another.

Just as digital library documents have commentary associated with them, so do project files. This commentary fills a similar role to source code documentation in programming and web design. While the comments cannot be placed within the file, they remain attached to the file as long as it exists. Whisper will provide an RSS feed that includes file commentary at the same level as project messages; this has the added benefit of allowing users to monitor messages with their existing RSS clients without needing to log into Whisper's web interface.

**Project: Message Boards.** Projects also have message boards where members can discuss project issues and where Whisper will post notification messages related to check-in events and membership changes. The message boards will be implemented so project members can track messages in their projects with their regular email client. Each project will also be assigned an email address or mailing list so users can post messages to the project simply by composing an email message to the project. Messages will also be available for browsing, searching, and replying via the web-based interface.

**Project: Wiki.** The wiki allows project members to document the project from high-level overviews down to individual files. Since wikis can be versioned, the entire project history can be captured and explored from initial questions and answers, to planned schedule, to changes in direction as the project matures. Portions of the wiki can also be made public to communities or the world at large, allowing non-members to learn about the work being performed by the project; it would also be an appropriate method for distributing project results and artifacts to non-members. Whisper will not, however, enforce any kind of workflow or process on project members that would require them to use the wiki for any particular purpose; at most, project wikis will be initialized with a default, but empty, information structure.

**Community: Needs.** Communities provide the framework in which Whisper's users work and develop their projects. As with projects, any user can create a community, provide an initial description, and set the initial scope of eligible members; some communities might be limited in membership to the local Whisper system, but we expect most will be open to the world. Like projects, communities include a wiki, message board, and event management (conferences, symposia, and other community wide meetings). Traditionally, research results have been published in print journals which, until recently, did not provide electronic copies. Whisper communities can create purely electronic journals with an editorial board and reviewers to examine work submitted for publication; these electronic journals are incorporated into each community's digital library.

**Community: Creation.** Communities can be created by any user on a Whisper server. By default, any user from any Whisper site will be allowed to join; the creator can adjust the scope of eligible people by requiring they be from a limited set of Whisper sites, such as those run by a particular university or organization, or already be members in one or more other communities. Communities can be linked together by their relationships forming an interconnected web of research interests. The founder of a community determines these initial relationships, allowing announcements of new communities to related ones.

**Community: Message Boards.** Communities provide a message board and wiki for many of the same reasons projects do, including developing a history and maintaining general information about the community and its goals. However, because communities will generally have a larger membership than projects, a single message board could quickly become overloaded. To prevent this, a community's administrator(s) can create separate message boards for specialized topics within the community. For example, a biology community might create separate message boards for Ecology, Evolutionary Biology, Molecular Biology, and Cellular Biology. Users can monitor either specific topic boards or the entire community board with their email client or through Whisper's web interface; RSS feeds of community messages will also be available.

**Community: Digital Journals.** Publishing results in a journal is frequently the object of research projects. Traditional print journals have a process that can create a significant delay between the date of a paper's submission and its publication. The time and expense of printing a high quality journal creates a significant bottleneck to the overall process of publication, limiting the number and size of the works that can be published in any one volume. Whisper's journals provide the same community review process that print journals employ, but do not have any overhead cost for printing, nor require any kind of page

limitation. In addition, once papers have been reviewed and approved for publication, they are immediately made available to all members of the community, and possibly the world. We realize, however, that many community members will continue to publish their results in traditional print journals and in conference proceedings. For this reason, Whisper supports the submission of links to external papers; after editorial review to ensure the linked paper is appropriate for the Whisper journal, it is published to the community. Communities can have multiple journals, each with a different focus. Ideally, when a project hosted on a Whisper site publishes to a journal hosted on a Whisper site, the publication will include a link back to the project, giving community members and other readers more details and a better understanding of the context in which the research was performed; links to the project would also make it easier for users to find other publications resulting from the project work or by the project's members. The same system will also provide a mechanism for communities to review papers for conferences.

## 3.2 Interaction with External Systems

Whisper installations need to interact with multiple external entities. Some entities support users with lightweight web-based clients and heavier clients that are native applications on the user's local computer. The OAI Harvest Service [25] supports interaction with external data storage systems, including other digital library sites and clients. In addition, other Whisper sites must be considered external systems to protect intellectual property rights of the local users.

**Whisper Clients.** The initial Whisper web client will be a dynamic website, which can be deployed on the same machine as the Whisper server, or on a separate machine. The web client communicates with the Whisper presentation tier using a standard API (application program interface), built on top of the SOAP protocol [22]. Because SOAP is a standard protocol (based in turn on the XML and HTTP standards), most operating systems have built-in libraries that allow native applications to interact with SOAP-based web services. Using these libraries (or custom SOAP implementation when necessary), developers can create standalone client applications that interact directly with the Whisper server. Web clients also allow anonymous access to the public areas of the server.

Native clients can provide full access to the same features offered by the web client. The Desktop Synchronization Manager allows native clients to store local copies of the comments made by and to them, as well as the comment spaces for projects or communities they are involved in. Users can also synchronize their repositories (library collections, personal files, and project files) using either native clients, or standard HTTP, WebDAV, or Subversion clients. Many operating systems already allow WebDAV enabled web spaces to be mounted as network drives.

**Other Whisper Installations.** Whisper servers must be locally installed to satisfy intellectual property requirements; most organizations prohibit their member or employees from storing sensitive data on a third party's system. In addition, the cost of providing an installation large enough to handle the needs of a worldwide user base is prohibitive. To enable users in different locations to interact, the Whisper servers themselves must provide seamless access to resources located in external organizations. Each local server can be configured to allow varying levels of access to users from external sites, ranging from no access to the same access local users have (though projects and communities created by a user are always stored at that user's local server). The Remote Server Access Manager handles authentication of remote hosts, so that users need only authenticate with their local system (a single sign-on) to access remote resources. Once the authentication is complete, the user's client (whether it is a web based client or a native application for their personal computer) will be able to interact directly with the remote server.

**Digital Library Systems.** Whisper servers will provide the OAI Harvest Service [25] to enable other digital library clients and services access to resources within the Whisper server. By default, all public information in a server will be made available; this includes community journals and comment areas, information about local projects, and information about local users. Whisper users will be able to specify how much of their personal information (files, library collections, WebLogs, etc.) will be accessible through the OAI Harvest Service. External users wishing to explore information stored on a Whisper server will be directed to that server's local web client.

# 4 Scope and Timeline

While it will take years to fully implement all the features we have described, a subset of the features can provide a compelling platform for early adopters. Table 4 shows the project timeline, including development milestones for the minimal feature set we believe will allow us to verify the claims presented in Section 2. Other features may fall into place as well; for example, digital libraries for projects and communities might be a simple extension of the same feature for individuals.

Table 1: Phases of development

| Phase | Description | Completion |
|---|---|---|
| 1 | Surveys and interviews for requirements planning; Security mechanisms, including role- and relationship-based access controls; Administration and management for individuals, projects, and communities | Month 2 |
| 2 | **Beta 1** A web-based user interface; Inter-user relationship management for the individual user | Month 4 |
| 3 | **Beta 2** Digital Libraries, with annotations, for the individual user | Month 6 |
| 4 | Inter-Whisper communication, enabling cross-installation interaction | Month 10 |
| 5 | **Beta 3** Versioned file spaces for projects | Month 12 |
| 6 | Interviews with early adopters | Month 12 |
| 7 | Discussion lists for communities | Month 14 |
| 8 | **Beta 4** Electronic journals for communities | Month 18 |
| 9 | Begin metrics analysis | Month 21 |
| 10 | Additional user surveys and interviews to sanity check metrics | Month 23 |
| 10 | Begin user studies (comparison of work with and without Whisper); possible usability studies (analysis of Whisper's UI) | Month 24 |
| 11 | Results from first set of user studies submitted for publication | Month 32 |

Beta versions will be available to select groups beginning with phase 2, but these versions will only gather trial metrics to ensure we are meeting the needs of Whisper's users; beginning with the third beta, we will make the source code of each stable release available to the public for download. The first beta will enable a small group of people to access the system and configure their public information and work relationships with other users. The second beta, incorporating digital libraries for individual users, will open access to the entire UCSC School of Engineering, consisting of mostly graduate students, faculty, and staff, with a few advanced undergraduates. A third beta will add group project spaces, and enable interaction between different Whisper installations at the system level; other departments within UCSC will be invited to participate. The fourth beta enables community interactions; we hope to provide the whisper service to the entire UCSC community at this point. After the fourth beta, we will begin conducting studies that will determine how successful Whisper is in removing friction to knowledge flow; while not detailed here, new feature development will continue, but at a less prodigious rate. If usability studies are conducted, we will do so at one of the earliest sites to adopt Whisper. After the system has been used by the full community for at least three months, we will begin analyzing the metrics gathered by the system; we will invite a sample of Whisper users to participate in an additional round of interviews and surveys.

# 5 Validation

An explosive growth in Whisper installations within academic institutions worldwide, on the order of dozens in the first six months after the fourth beta, would suggest that we have hit a chord within the academic community, and have fielded a valuable service that users feel improves their ability to perform their work. Our user surveys and interviews will also demonstrate how many users feel the Whisper system improves

their work environment, or makes them more efficient. For example, before the fourth beta is made released, we will survey potential users to determine how they accomplish the tasks augmented by Whisper, and how much effort goes into performing those tasks. A few months after Whisper is available to them, we will re-interview the subjects and determine who has not tried using the system, who tried the system and stopped, and who is a continuing user. For those who have not tried Whisper, we will determine what, if anything, is preventing them from using the system. Users who used Whisper for a time and then stopped can be asked about the barriers that prevented them from adopting Whisper in their research process. Active Whisper users will be asked again how they perform the tasks augmented by Whisper (to determine which of Whisper's features they make use of), and how much effort the tasks require; for tasks in which they use Whisper, we will ask how they perceive their effectiveness in the task execution compared to their previous methods.

We believe the vast majority of those who have not tried Whisper, or stopped using it, will cite a lack of research colleagues in their community within the system. Another significant group of these users will be uncomfortable with the technology and the new methods of working; academics can be slow to change in most basic research methods. Beyond this, we expect only technical and social reasons for not trying or adopting Whisper. For example, some user's work may require large files that cannot be easily transferred over network connections. Other users may spend most of their time in the field, without a connection to the internet or a computer. In some cases, granting agencies may have security requirements that cannot be met by any collaboration support software, preventing researchers working on those grants from utilizing Whisper. If a majority of those interviewed that tried Whisper and later stopped using it, did so because they feel Whisper does not provide a compelling service, we will have to reexamine our goals and designs.

Whisper's implementation will enable us to gather metrics on the system's use. The metrics gathered in the system will protect the user's identities and data while allowing us to map knowledge flows within the system. The observed knowledge flows will demonstrate the completeness of our models, to the extent that Whisper is able to capture a particular type of knowledge flow; for example, Whisper will not be able to capture knowledge transfer during telephone conversations. In conjunction with our user surveys, interviews, and case studies, the metrics will allow us to validate our claims related to specific sub-systems within Whisper and their integration and the reduction of friction as knowledge flows from one entity to another.

## 5.1 Digital Libraries

We will utilize Whisper's usage metrics to determine how often users are searching the digital libraries within the system, and the relationship between the person providing the digital library and the person searching it. We expect analysis of the metrics to show increased searches for user pairs with a close working relationship when compared to user pairs that have a distant or no working relationship; for example, we expect a graduate student to make use of their professor's digital library more often than a professor in physics would make use of the digital library provided by a professor in literature. In addition, we will examine the frequency with which users copy a document found by the search into their own digital library; a greater frequency of this event indicates more knowledge being transferred through the digital library system within Whisper. Associating commentary with a digital library document, or reading commentary by other users, are also events signifying knowledge transfer within the Digital Library.

Interviews with the users will determine where the Whisper digital libraries fall when searching for work related to their own. When the number of active digital libraries in a user's working relationship chain is large, we believe that user will search Whisper earlier and more frequently in their research process. Likewise, for users without relationships to others in their field with digital libraries, we expect Whisper to be searched much later in their process, if it is searched at all. User studies will demonstrate how the use of personal digital libraries significantly reduced the friction associated with locating a particular paper and relating that paper to the user's work.

## 5.2 Versioned File Spaces

Usage metrics from Whisper installations will show the number of projects making use of versioned file spaces. If these spaces are not available in individual user areas, we believe a larger proportion of projects will have only one member, as individuals take advantage of projects' versioning capabilities. The metrics will track the number of checkins and checkouts in the repository; for projects, this tracking will be on a per-member basis.

Interviews with users will be used to determine how users view the versioned file spaces in comparison to their previous methods for versioning their work. We will conduct ask some groups of students in a project-related class to user Whisper for their project management, while others in the class use whatever methods they have available to them. At the end of the class, we will ask each group to describe the method they used to keep their work files synchronized, and how much time their selected process required; we will also ask them how often one person's changes were lost or overwritten by another group member, how this issue was resolved, and how much time the resolution took. Based on studies in software development groups, we expect the interviews to show groups using Whisper's versioned file spaces encountered less friction in the form of less time synchronizing their changes, fewer (or no) "lost" revisions, better awareness of the project's status, and better communication between project members.

## 5.3 Community Discussions

To provide enough metrics for validating our claims related to communities, there must be one or more active communities within at least one Whisper installation; in addition, it must include a significant number of members from other institutions — preferably 25% or more. However, other institutions are unlikely to install Whisper until it has been proven in some fashion. This situation can be resolved by allowing external users to have access to the system installed at UCSC after the fourth beta; this system could be maintained by the university, or provided as part of this research project. By providing a peer review process appropriate for reviewing conference papers, we will get a large percentage of the community members to use the system for the conference. We anticipate that some subset of these users will continue to use and adopt the Whisper system afterwards.

Usage metrics from the communities will show how active each community member is, and how often members search or browse the community's discussion boards; we believe more active communities will engender more activity within the discussion areas. We believe there will be a small increase in new projects between geographically disperse users when both users participate in an active community, though it is possible Whisper will not be online long enough for this increase to be statistically significant. Within a single institution, local communities (those restricted to users affiliated with the institution) will have a similar effect on cross-department collaboration as users become more aware of interests and current research outside their own department; similar results were found by Carter, et. al. [9] while studying loosely coupled groups within an organization.

We will conduct user surveys and interviews to examine how much friction due to geographical or departmental boundaries is reduced by examining how close users feel to the research communities they participate in regularly. These surveys should show that users will feel more connected to communities with a presence in the Whisper system than those without, and that more active communities engender greater feelings of connectedness among their members. The surveys will also determine how much more likely users are to enter a joint project with a geographically distant colleague after participating in community discussions with that colleague, demonstrating the potential for these projects to be created in the future, even if they are not being created in significant numbers at the time of the study.

# 6 Related Work

## 6.1 Computer Supported Cooperative Work

Engelbart coined the term "Knowledge Workshop" to describe his Augment system, an environment in which knowledge workers could perform their daily tasks [17]. Since then, some of Augment's features, such as document creation and editing, hyperlinks and hypertext, e-mail, instant messaging, and digital libraries, have become research foci and product lines in their own right. The Computer Supported Cooperative Work (CSCW) domain now encompasses systems like Augment, frequently referred to as groupware, that enable people to work collaboratively on a project.

A traditional goal of CSCW systems has been to provide integration to the tools needed in a domain; this is shown to make information easier to find and reduce unnecessary redundency [14]. We believe this trend is stronger in academia as professors and students have more freedom in the goals they pursue and how they proceed. Early CSCW systems [2, 19] provided users with a linked network of data files, often with a proprietary format. Web sites provided a new option for collaborative spaces, as explored by the Basic Support for Cooperative Work [5, 4] project, which provides users with a shared workspace with locks for individual files as well as threaded discussions for the workspace. BSCW does not provide community support, nor is the system architected to support digital libraries, electronic journals, or shared annotation of works therein.

Content management applications offer support for searching, indexing, a separation of information structure and presentation, multiple types of output (HTML, PDF, or print) from a single input source, access control, workflow management, and life-cycle management [20]. Most commercial content management systems provide support for authoring documents, generally from within a web browser. Some systems, such as Vignette Collaboration [12], go beyond this and provide general project management and collaboration, personal workspaces, and integration with standard e-mail. Content management tools do not generally account for how a document is used outside the organization or group that authored it, a key consideration in academic knowledge flow that Whisper will address.

## 6.2 Digital Libraries

Digital libraries are generally electronic counterparts of traditional libraries; most work to date has focused on this transition [6]. Most are maintained and organized by the organizations that provide them. For example, the ACM and IEEE both provide digital libraries to their members, while businesses and organizations involved in book publishing frequently require payment to download the text of each article or journal. In these cases, there is usually a formal review process used to add new material to the library. The Open Archives Initiative (OAI) [25] is working to develop standards for open protocols that will enable communication between various digital library installations, and Whisper will support these standards.

The Oxymoron system [13] is a groupware tool that targets academics, specifically those in the social sciences. It provides a bibliographic system with shared annotations for each document. While the varied backgrounds and fields of the system's users helps to create new interdisciplinary understanding and foster cooperation in and between communities of users, Oxymoron's primary goal is to provide a community-maintained digital library.

The proposed Collaborative Knowledge Evolution Support System (CKESS) is an enhanced form of digital library [6]. Its purpose is to support communities and the individuals working within them by incorporating workflow and process support tools, decision analysis, dynamic hypermedia, and conceptual knowledge structures into the basic document repository. Process support tools in the system enable the community to manage itself by documenting (and then maintaining) basic processes such as officer elections or new member training. To our knowledge, CKESS remains a proposal and has not yet been implemented.

### 6.3 Organizational Memory

The term "organizational memory" has many definitions in literature due, in part, to the relatively small number of empirical studies [1]. Our use of the term refers to tools which allow an organization to record its knowledge in a persistent state, recall that knowledge at a later time, and make alterations and additions as time progresses. Such tools often differ from digital libraries in the overall quality and type of their content. For example, it would be rarely acceptable for a user to submit a paragraph outlining a potential project, or the maintenance schedule for the organization's building in a digital library; on the other hand, these are good examples of items of interest to an organization that should be remembered over time (especially between predecessors and successors in an organization). More explicitly, digital libraries and document management systems are appropriate for storing formal knowledge in the form of books, documents, and training material; organizational memory tools store informal knowledge such as ideas, facts, assumptions, definitions, unknowns, decisions, and points of view [11].

The Augment Journal served this purpose, even though it is also considered a digital library. Engelbart notes the Journal contains "current and thoroughly used working records of the groups plans, designs, notes, etc." [17]. Internet USENET [23] groups are also a form of organizational memory; while they are greater in user scope (potentially reaching every internet user), the vast number of topics necessitates deletion of older messages from most servers. Wikis [26] are a modern tool which is well suited to maintaining organizational memory. They allow users to edit the content as well as the structure of a web site, can be versioned, and sometimes provide access control (internally or through the web server). Whisper's wiki functionality in conjunction with archived discussion spaces will provide rich support for organizational memory at both the project and community level.

## 7 Conclusion

We believe the tools developed to support academic knowledge transfer can be applied to other types of knowledge work, including the governance of cities and nations, legal proceedings and organizations, commerce, and some parts of everyday life. Once the model of academic knowledge transfer solidifies, we can extend it to these, and other, types of knowledge work, and extend the tools, and provide new ones, to support the greater realm of knowledge work. Some of the recognized benefits to improving knowledge transfer include:

- Citizens better informed of the workings within their government, and representatives more aware of the opinions and concerns of their constituents.
- Legal documents, such as contracts, more easily traced back to earlier versions so weaknesses that have carried are identified and corrected.
- The assumptions and working models developed by an engineer are available for examination and verification long after the engineer has left an organization.

Engelbart noted "we are far short of being able to do a one-pass redesign of any major portion of [our] capability infrastructure if only because of their pervasive, underlying dependence upon human processes" [16]. We feel it is important to develop a high-level understanding of the current state-of-the-art in human augmentation. Tools for knowledge transfer are only one aspect of this greater study, though an important one because such an understanding is unlikely to be develop without greater interdisciplinary interaction between the fields of computer science, computer engineering, psychology, and sociology. Such interaction already exists and provides the foundation for the domains of Human-Computer Interaction (HCI) and Computer Supported Cooperative Work (CSCW); we consider HCI and CSCW important foundations for developing better methods of augmenting knowledge workers.

The result of augmenting knowledge worker's ability to transfer and work with knowledge is increased knowledge synthesis. The complex problems faced by today's societies are unlikely to be solved by any one discipline. Increasingly, new domains of knowledge work are being created at the intersection of more

traditional research domains, such as Bioinformatics (Computer Science and Biology), Bioethics (Philosophy and Biology), and Quantum Computing (Computer Science and Physics). Increasing the knowledge worker's ability to synthesize knowledge from a wider range of disciplines will give rise to better solutions in less time.

# 8 Acknowledgments

I would like to thank Jim Whitehead, Jennifer Bevan, and Laura Triechler for their feedback and assistance in crafting this paper. I would also like to thank Geneva Henry, Rudi Schmiede, and the attendees of the JCDL 2005 Doctoral Consortium for their time and feedback.

# References

[1] Mark S. Ackerman and Christine Halverson. Organizational memory as objects, processes, and trajectories: An examination of organizational memory in use. *Computer Supported Cooperative Work*, 13(2):155–189, April 2004.

[2] Robert M. Akscyn, Donald L. McCracken, and Elise A. Yoder. KMS: a distributed hypermedia system for managing knowledge in organizations. *Communications of the ACM*, pages 820–835, 1988.

[3] Apache Axis. `http://ws.apache.org/axis/`.

[4] Richard Bentley, Wolfgang Appelt, Uwe Busbach, Elke Hinrichs, David Kerr, Klaas Sikkel, Johnathan Trevor, and Gerd Woetzel. Basic support for cooperative work on the world wide web. *International Journal of Human-Computer Studies*, 46(6):827–846, June 1997.

[5] Richard Bentley, Thilo Horstmann, Klaas Sikkel, Johnathan Trevor, and Gerd Woetzel. Supporting collaborative information sharing with the world wide web: The BSCW shared workspace system. *Proceedings of the 4th International World Wide Web Conference*, pages 827–846, 1995.

[6] Michael Bieber, Douglas C. Engelbart, Richard Furuta, Starr Roxanne Hiltz, John Noll, Jennifer Preece, Edward A. Stohr, Murray Turoff, and Bartel van de Walle. Toward virtual community knowledge evolution. *Journal of Management Information Systems*, 18(4):11–35, 2002.

[7] Blogger. `http://www.blogger.com`.

[8] Frank Buschmann, Regine Meunier, Hans Rohnert, Peter Sommerlad, and Michael Stal. *Pattern-Oriented Software Architecture: A System of Patterns*. John Wiley and Sons, 1996.

[9] Scott Carter, Jennifer Mankoff, and Patrick Goddi. Building connections among loosely coupled groups: Hebb's rule at work. *Computer Supported Cooperative Work*, 13(3–4):305–327, August 2004.

[10] Geoffrey Clemmm, Jim Amsden, Tim Ellison, Chris Kaler, and E. James Whitehead. RFC 3253: Versioning extensions to WebDAV (Web Distributed Authoring and Versioning).

[11] Jeff Conklin. Designing organizational memory: Preserving intellectual assets in a knowledge economy. *`http://cognexus.org/dom.pdf`*, 1997.

[12] Vignette Corporation. Vignette collaboration. Technical report, `http://www.vignette.com/Downloads/WP_VignCollaboration.pdf`, May 2004.

[13] Camille Bierens de Haan, Gilles Chabré, Francis Lapique, Gil Regev, and Alain Wegmann. Oxymoron, a non-distance knowledge sharing tool for social science students and researchers. *Proceedings of the international ACM SIGGROUP conference on Supporting Group Work*, pages 219–228, November 1999.

[14] Gunnar Ellingsen and Eric Monteiro. A patchwork planet: Integration and cooperation in hospitals. *Computer Supported Cooperative Work*, 12(1):71–95, February 2003.

[15] Douglas C. Engelbart. Special considerations of the individual as a user, generator, and retriever of information. *American Documentation*, 12(2):121–125, April 1961.

[16] Douglas C. Engelbart. Toward high-performance organizations: A strategic role for groupware. *Proceedings of the Groupware '92 Conference*, August 1992.

[17] Douglas C. Engelbart, Richard W. Watson, and James C. Norton. The augmented knowledge workshop. *AFIPS Conference Proceedings*, 24, June 1973.

[18] Roy Fielding, Jim Gettys, Jeffrey Mogul, Henrik Frystyk Nielsen, Larry Masinter, Paul Leach, and Tim Breners-Lee. RFC 2616: Hypertext transfer protocol — HTTP/1.1.

[19] L. Nancy Garrett, Karen E. Smith, and Norman Meyrowitz. Intermedia: issues, strategies, and tactics in the design of a hypermedia document system. *Proceedings of the 1986 ACM conference on Computer-supported cooperative work*, 1986.

[20] David Gingell. A fifteen minute guide to enterprise content management. Technical report, Documentum, 2003.

[21] Yaron Goland, E. James Whitehead, Asad Faizi, Steve Carter, and Del Jensen. RFC 2518: HTTP extensions for distributed authoring — WebDAV.

[22] Martin Gudgin, Marc Hadley, Noah Mendelsohn, Jean-Jacque Moreau, and Henrik Frystyk Nielsen. Simple object access protocol. `http://www.w3.org/TR/soap12-part1/`.

[23] Mark Horton and Rick Adams. RFC 1036 : Standard for interchange of USENET messages.

[24] Jakarta Slide. `http://jakarta.apache.org/slide/`.

[25] Carl Lagoze, Herbert Van de Sompel, Michael Nelson, and Simeon Warner. The open archives initiative protocol for metadata harvesting. `http://www.openarchives.org/OAI/openarchivesprotocol.html`.

[26] Bo Leuf and Ward Cunningham. Wiki: Wat is wiki. Technical report, Wiki.org, June 2002.

[27] J. C. R. Licklider. Man-computer symbiosis. *IRE Transactions on Human Factors in Electronics*, HFE-1:4–11, March 1960.

[28] LinkedIn. `https://www.linkedin.com/`.

[29] Sylvie Noël and Jean-Marc Robert. Empirical study on collaborative writing: What do co-authors do, use, and like? *Computer Supported Cooperative Work*, 13(1):63–89, January 2004.

[30] The national science digital library. `http://www.ndsl.org`.

[31] Orkut. `http://www.orkut.com/`.

[32] SourceForge. `http://www.sourceforge.net`.

[33] Spring Framework. `http://www.springframework.org`.

[34] Subversion. `http://subversion.tigris.org`.

[35] Jürgen Vogel, Werner Geyer, Li-Te Cheng, and Michael Muller. Consistency control for synchronous and asynchronous collaboration based on shared objects and activities. *Computer Supported Cooperative Work*, 13(5–6):573–602, December 2004.

[36] Yahoo! Groups. `http://groups.yahoo.com/`.