PART I

FROM CONCEPTS TO USING WIKI

CHAPTER 1

Introduction to Discussion and Collaboration Servers

Before getting into the heart of "the wiki way", you might want to know more about what "discussion and collaboration servers" are. We refer to this term frequently. The *Wiki* concept, to which we devote most of this book, belongs to this general class of tools in the multiuser context, in addition to being a tool to collect and cross-reference information.

There are of course different ways to collaborate, depending on the work styles and locations of the participants. For the purposes of this book, our assumption is that the use of a computer network is a natural mode of communication for the reader.

We contemplated having this as an appendix, but an introduction does serve a purpose and can help you later determine whether a wiki is appropriate to your intended purpose. However, if you're the impatient kind, you can for now skip this chapter and go ahead to the next. As proponents of Wiki and the Web, we are fully aware that life, learning, and a given reader do not have to progress in a neat linear fashion. There is a progression from cover to cover, and we hope an entertaining read, but it's not essential that you slavishly follow it. Hence the detailed table of contents, list of tips, and other entry points into the content.

IN THIS CHAPTER

This is an overview chapter designed to give a background in broad strokes on the subject of computer-mediated discussion and collaboration.

- Collaboration Models introduces the basic conceptual models for collaboration, such as e-mail exchange, shared access, and interactive pages. The characteristics and differences are illustrated. Emerging collaboration models are also touched on.

- Who Uses Collaborative Discussion Servers? analyzes and groups the different types of users likely to be involved in collaboration or sharing situations. The companion section Whatever For? makes a case for why users would want to consider "yet another tool" instead of just making the best of it with existing exchange methods such as e-mail.
- Features of a Web-Based Collaboration shows why Web-based collaboration services are attractive and notes that the primary requirement of any such solution is that it be easy to use.
- On the Horizon: WebDAV mentions an emerging Internet protocol enhancement that could make interactive collaboration over the Web as natural as browsing.
- Comparing Wiki with Other Collaboration Tools concludes the chapter by explaining the special attractions of the wiki solution—mainly that it works here and now with few resource investments.

COLLABORATION AND DISCUSSION TOOLS

The *Oxford English Reference Dictionary* gives us these baseline definitions:

1. *Discussion, n. A conversation, especially on specific subjects; a debate.*
2. *Collaborate, v. To work jointly.*
3. *Server, n. (Computing) A program which manages shared access to a centralized resource or service in a network.*

From this you can correctly draw the conclusion that here we have some kind of software tool that promotes and mediates discussions and joint working between different users. Such a tool can often form an important resource for collaboration projects—this book is but one example.

Computer-served collaboration and discussion can be set up in many ways, and more are being devised as we get new resources, develop new ideas, and become more accustomed to working over a network.

We first look at some of the existing generic models, to set the stage for later discussions in this book.

COLLABORATION MODELS

In principle, you find three collaboration models over a network:

- E-mail exchange (includes the mailing list)
- Shared folder/file access
- Interactive content update/access

Other (emerging) technologies go further than these simple models in terms of sharing data and application functionality over a network (the Internet), but at present it is unclear how pervasive, cost-effective, and easy to use they might become for the individual user.

E-mail Exchange

E-mail exchange provides direct exchanges between the members of a collaborative group. This is "simple" and requires only that members have e-mail capability. To keep up to speed, either all members receive copies of relevant messages, or the messages are broadcast as a mailing or distribution list—again, everyone receives a copy.

The possible "server" software here is whatever manages the e-mail exchanges and possibly the mailing list of members. This software can be either part of the e-mail client or a dedicated "list-bot" on some Internet server. The technology is pure "push", and it is up to the recipient to sort, archive, and make order of the mail flow. In some contexts, the "interruptive" nature of this push is considered an advantage, although other models can include various notification mechanisms.

The ultimate public form of e-mail is the Internet newsgroup, where postings are distributed globally to all participating newsgroup servers and kept there for access during an arbitrary time. Newsgroup readers fortunately don't see the scale of this massive storage, because the client software shields them by allowing them to browse message headers from a designated server. Then they load selected messages to their local systems to read and archive as they see fit—the last step is thus just a little like shared access from the user point of view. The newsgroup differs from the mailing list in the lack of "interruptiveness", the default storage model, and the fact that messages expire on the server.

In schematic form, the user relationships can be as shown in Figure 1-1. Characteristic of the e-mail exchange model is that each user must sort and keep personal copies of all the messages posted to the discussion that might be relevant to access later.

More important, the postings cannot be edited or easily cross-linked in any way useful to the group unless they are somehow collected with annotation features into a central archive. If such archival access is required, you might well prefer postings directly to the database, similar to the later models.

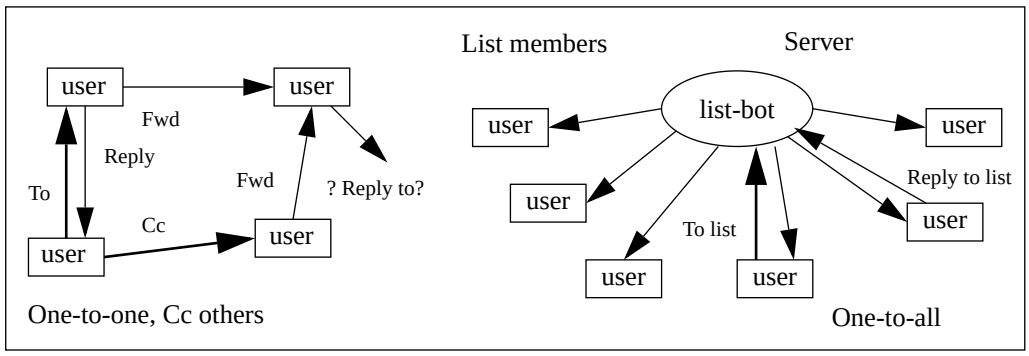

FIGURE 1-1. *Flow model of e-mail exchanges and a mailing list model (right). For newsgroups, the "user" boxes in the latter would be the "NNTP servers" that a user polls to read messages.*

Shared Access

Shared access in its simplest form simply means that the members can directly access the same files in a common repository on a particular server. This is perhaps the most common model for corporate network collaboration, where members have extensive file access in a relatively transparent way across the corporate network. The problem here is one of coordinating edits and updates, so most often we still see a lot of individual, Cc, and broadcast e-mail keeping the members up to speed.

Some share solutions specific to software clients allow more general Internet shares (such as Microsoft Outlook's more corporate-mode folder and file "sharing" components). In practice these often become a kind of push technology akin to mailing lists, because they send automatic e-mail "updates" to subscribing members whenever the shared folder is modified, and they conduct behind-the-scenes "chatter" with synchronization messages to determine when updates are required. Usually (copies of) any posted files must also be distributed to member clients.

Shared access, whether by copy or by actual access to the original files, is almost always combined with some form of graduated access control. Different members, or groups of members, have varying degrees of freedom to contribute and edit the shared material, ranging from owner or author down to contributors, reviewers, and readers.

Figure 1-2 shows a schematic of shared access exchange organization. Note that the "shared" focus is mainly on access to a "common" repository of files or postings, either directly or via synchronized local copies based on some share owner's repository folder. Discussions between members still occur as threaded postings or, as is often the case, as regular e-mail exchanges.

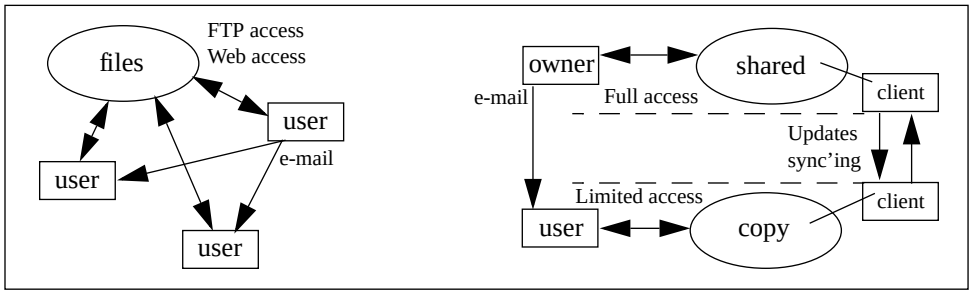

FIGURE 1-2. *Shared files and folders model. Clients such as Microsoft Outlook generally require update and synchronization messages between local copies.*

To find particular items in the store, members need some form of index and prior knowledge of what is there, because many of the files will be in application-specific formats and not immediately "browsable".

Interactive Pages

Interactive page access occurs when the members of the group can collectively edit the same material. Again, we see a number of proprietary solutions that use various metaphors—conference calls, whiteboard collaboration, document review comments, or linked notes. Wiki servers belong to this third group, despite the fact that some hold that "interactive" means simultaneous multiuser edits of the same content.

The point of interactive exchange is that members can collaborate on content, either in real time or asynchronously, by editing the same document (or documents). In this way it more closely emulates a real verbal discussion, with the added feature of being persistent.

The interactive sharing model is illustrated in Figure 1-3. This last figure may seem very simplistic in that the diagram ignores the server-client components, but the point is that the original "document" is available at all times, always in its latest version. Although this figure is superficially similar to the mailing list model (Figure 1-1), the arrows here represent *actual (editing) access to the data repository*, not just the flow of discrete message items being replicated between systems.

Some server models, such as Zope (see its home site at www.zope.org), go even further by making the entire server infrastructure accessible to, for example, collaborative Web-publishing efforts. This means that the (member) users can individually or collaboratively modify not just content but also server behavior by editing actual components.

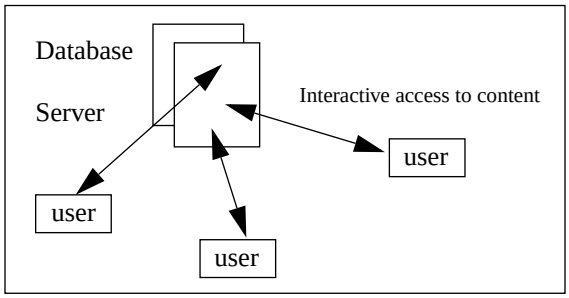

FIGURE 1-3. *Interactive server model with collaborative content*

Other Modes

There is an overall trend lately to broaden sharing and collaborative work from simple (text) document models to more complex multimedia and arbitrary media content. Using extensions based on the standard or evolving Web protocols, eXtensible Markup Language (XML) to complement HTML, or Zope-style application server technology are all approaches that refrain from building in any inherent server-client dependencies.

A second trend suggests the overall shift of the Web from just a huge collection of static view-only pages of text and graphics to a more interactive model, where users can share and work together in a wide variety of media types as a matter of course. The vision is that our applications will no longer care where the documents are but will transparently manage access and sharing, even across the Internet. The recent move by Microsoft into the proprietary "*dotNET*" technology is so far a Windows-based example, even though this builds on the emerging open standards of XML and Simple Object Access Protocol (SOAP). The dotNet strategy is just as much about licensing and control of software as it is about sharing and collaboration, however.

WHO USES COLLABORATIVE DISCUSSION SERVERS?

We find several different groups that use collaborative discussion server technology. Although to some extent overlapping, this grouping is useful to define different application areas. In all cases, whether private or public, remote access can be over local networks, intranets, or the Internet.

- *Individuals*, who can use the server technology to create, organize, and store content for their own use. Although this suggests exclusive use on a

local machine, remote network access can provide off-machine or off-site storage/backup.

- *Special (temporary) collaboration groups,* which implement (nonpublic) server-based collaboration for their specific projects.
- *Special-interest groups,* which set up a server to mediate and archive discussions and joint projects among members and often the general Internet community at large.
- *Academic groups,* which set up this kind of server to complement or even define class or faculty projects, collect and post information, and mediate project review.
- *Corporate groups,* which can use these servers to plan, execute, document, and follow up various projects. The servers may be team-oriented or work for the entire company or division.

These groups cover a lot of territory. So if almost any field is a candidate for the use of this technology, when should you consider using it? For that matter, why would you even want to?

WHATEVER FOR?

"Yet another obscure technology", some might complain when introduced to this kind of tool. "What makes this worth my attention"? We'll return to this question in the specific context of WikiWikiWeb later, but for now we just examine the more general field.

Collaborative discussion servers can be used whenever you want to centralize sharing of discussions or resources. In a broader sense, sharing can also encompass the individual who wants to share data access between different machines or locations or wants to form relationships between different documents on the same system by letting the server organize and link relationships.

Most server deployments are dictated by a particular task or goal. This then defines the nature of the discussion or collaboration needs, at least in a general sense. Sometimes, one is simply looking for an alternative, to replace an existing technology or way of working that isn't good enough.

Especially in the corporate world, there is widespread appreciation that the use of e-mail and mailing lists very quickly becomes unmanageable for some project work, even when using "industrial strength" clients that boast many organizational features.

The concept simply doesn't scale well. An e-mail-based collaboration of any size also requires considerable "attention bandwidth" investment by each individual to sort out, manage, and keep up to date on the important developments. Thus centralized serving cuts down on both wasteful duplication and update difficulties.

It's not unusual, however, to be unclear about the benefits or problems in any given situation until a particular server solution has been tried. This is one good reason to seek solutions that require low initial investments. Common too is not to see initially what structures are relevant to set up; hence, flexible solutions are also desirable.

One low-cost model is the collaborative Web server, especially when a network and Web site platform to build on already exist.

Features of a Web-Based Collaboration

Web-based collaboration has a number of interesting features in today's world of high connectivity and ubiquity of Web clients.

- *Free accessibility of the material.* Given basic connectivity to the Internet and a Web browser, you can access the material from anywhere at any time. Location-independent accessibility is also a key factor in the corporate or institutional intranet.
- *Up-to-date versions.* With centrally updated or interactive pages, you know that you get the current version at all times.
- *Hyperlinking.* The posted material can exist in a rich context of links—to collaborators, to resources, to comments, to older versions, and so on.
- *Independence of platform and application.* This is an ideal, yet reality is surprisingly close to it in a well-designed (or simple) Web-based system that uses standard protocols and content markup.
- *Content markup.* When used correctly, content markup makes content machine-searchable and more easily converted to other media. The prevailing markup standard for the Web is HTML, which provides a minimal but workable content tagging. This is a subset of Standard Generalized Markup Language (SGML), an international standard for the definition of device-independent, system-independent methods of representing texts in electronic form. A related, more powerful markup standard fast gaining in use is XML, because of the way it allows context-specific extensions to convey added embedded information about the data transferred.

There are disadvantages to Web collaboration, to be sure—mainly the impoverished editing interface provided by standard Web clients. Ideally, for broadest use, no special applications or client add-ons or plug-ins should be required to access or collaborate. Expect to see this situation change, however, as more and more Web-aware applications integrate Web functionality and traditional user interfaces in standardized ways.

On the other hand, many of the ways in which collaborative servers are used simply don't *need* elaborate interfaces. People clearly communicate successfully using just voice, writing, and pictures. All these basic modes are adequately supported by past and present Web user interfaces. Anyone doubting this statement need only look at the phenomenal growth of the Internet during the 1990s or for that matter paper publishing throughout the past century or so. This is especially true in the pure-discussion areas of Usenet and e-mail, which until recently have been almost exclusively plain-text with very kludgy interfaces.

Our conclusion from all this? A successful discussion server must be easy for its participants to use. All other potential powerhouse features are likely to be *orders of magnitude* less important, especially if they require extra software to be installed. A second critical factor must be the ability to easily refer to (link to) other "items", of whatever kind.

ON THE HORIZON: WEBDAV

The attraction of collaboration sites on the Web has also spurred further development of Internet infrastructure; in particular, extensions to the underlying protocol that allows users to interact with Web sites and other Internet resources. Any of these developments can greatly influence tools such as Wiki by providing new, ubiquitous editing options and better support for collaborative sharing of media other than text.

Of particular interest is WebDAV, or the Web Distributed Authoring and Versioning project, which is an attempt to extend the current Web HTTP 1.1 transfer protocol to include methods for creative collaboration in arbitrary media formats, not just text. The name clearly spells out the intent. When fully supported by both server and client, this would make collaboration over the Internet as natural as browsing. WebDAV appears based on some of the original Web-editing ideas once taken up by Tim Berners-Lee, considered by most to be the conceptual "inventor" of the World Wide Web.

The main resource for anyone interested in keeping an eye on this emerging technology is www.webdav.org, which has further links to other sites of interest.

COMPARING WIKI WITH OTHER COLLABORATION TOOLS

The rest of the book focuses on Wiki and wiki-related technologies. Wiki is a collaborative open-source tool that works *now,* within the context of existing servers, clients, protocols, and standards. It is also an evolving technology that will assuredly keep apace of infrastructure developments as these are deployed.

The main difference in comparison with most other collaborative tools is that a collaborative wiki is very informal and easy to use. Some sites may for special reasons require logins and passwords, but on the whole, the basic wiki interactivity consists of people "dropping by", browsing and reading, and when so inclined freely adding comments or new content.

As this book shows, setting up, customizing, and running wiki servers is very easy. In many situations, it is literally drag-and-drop and run. Furthermore, you are likely to already have most of the required components. Most computer users today, whatever their work or interests, whatever make and model of platform and operating system, have some kind of Web-browsing capability. Even older or budget systems with little extra resources can acquire a Web browser for free that needs only modest hard disk space. Web hosts are almost ubiquitous, and this book provides a script-based simulation of a Web server for quick home use. Given that, using a wiki is simply set up, browse, and write.

This approach is in contrast to other solutions that can require significant investments in software and time and sometimes new hardware as well. Packages like Lotus Notes or Outlook are admittedly industrial-strength solutions with many collaborative features for corporate needs, but they cost. They cost money, memory, hard disk space, upgrades, and above all much time to learn how to use effectively. Often and less obviously, the cost also involves other tools—for Web space creation and maintenance, for net-conferencing connectivity, for document creation. Thus you see the term "productivity suite" turn up a lot in these contexts.

Wiki is a "light" solution; simplistic perhaps, but capable of packing a surprising amount of functionality for the size of its code, even including the overhead for the perl package and your browser. (In this book, we adhere to the usage that "Perl" is the language as such, while lowercase "perl" is the implementation you install.)

Wiki is in addition a free-form solution, not setting any artificial limits on content as such. Significantly, wiki is a simple, open, and *nonproprietary* storage solution, so that your content is never format-locked to "this year's version".

Finally, wiki is an open-source solution—try it, use it, customize it, recode it. It's up to you. This book will guide you.

Next stop: "So what *is* a wiki, anyway?"