CPMCnet: An Integrated Information and Internet Resource

Robert M. Kahn, Ph.D., Pat Molholt, M.L.S., Jeff Zucker, M.A.
Office of Scholarly Resources, Health Sciences Division
Columbia University, New York, NY

The Network and Computer Systems department of the Health Sciences Library developed CPMCnet, an UNIX-based information and Internet server at Columbia-Presbyterian Medical Center. The project linked Gopher and World-Wide Web protocols as well as clients into an integrated application, providing the advantages of both. Development and use of CPMCnet has opened new channels of communication among information providers and end-users.

"...Strategic advantage is often provided by the ability to restructure through the power of electronic integration. This integration comes in several forms, integration of formerly distinct transaction processes, integration of multiple forms of data representation and knowledge, and integration of groups through communication. Such integration is enabled by two dominant properties, improved interconnection and improved, shared data access. Technology creates new integration possibilities through interconnection and shared data" [1].

Many universities and health sciences centers have embraced the concept of a campus-wide information system (CWIS) to provide "one-stop shopping" for information, and the information is often full-text as opposed to fielded data [2]. At CPMC, mission-critical full-text files were mounted in a DOS application called Folio Views. Folio Views allowed a great deal of control over hypertext links, indexing, full-text searching, and printing, but these features were only available to DOS clients [3]. While a Macintosh client was promised "down the road", the need for wider accessibility to the data prompted an assessment of other technologies.

Columbia University's Academic Information Systems unit had developed a UNIX-based CWIS called ColumbiaNet [4] that incorporated advanced features such as Kerberos-based access control and line-mode cut-and-paste. Mounting files on ColumbiaNet was considered, but having local control over the Medical Center's applications promised greater responsiveness to local needs.

Thus, the Network and Computer Systems department of the Health Sciences Library, part of the Office of Scholarly Resources at CPMC, developed a UNIX-based CWIS called CPMCnet, designed to complement ColumbiaNet and focusing on the needs of the Medical Center. CPMCnet is both a physical and a virtual repository, providing local storage for CPMC full-text data files as well as pointers to biomedical resources on the Internet. The development of CPMCnet has been an integrative effort in both the technical and the socio-political realms, incorporating multiple retrieval technologies, creating new links between departments, and merging many functions into an integrated whole.

TECHNICAL INTEGRATION

Clients and Servers

It has become clear in the last year that the client-server/distributed information model is well suited for providing both local (campus) and wide-area (Internet) information services [4]. Less clear was which server and which client to implement.

On the server side, Gopher [5] and the World-Wide Web [6] were both well established in global academic computing, both presented standard interfaces to information sources, both had powerful mechanisms for searching the Internet, and both could be combined with other important tools such as FTP, Telnet and WAIS. Gopher allowed quick and easy display of text files which could be mounted without major text processing. It also provided a very simple hierarchical menu structure that was well suited to the needs of beginning computer users. The World-Wide Web (WWW) required more effort in the realm of text processing and maintaining links between documents, but allowed greater flexibility of cross-referencing and was better suited for multimedia displays.

On the client side, line-mode clients allowed text access to users regardless of their computer brand or method of connecting to the Internet. Graphics

0195-4210/94/$5.00 © 1994 AMIA, Inc.

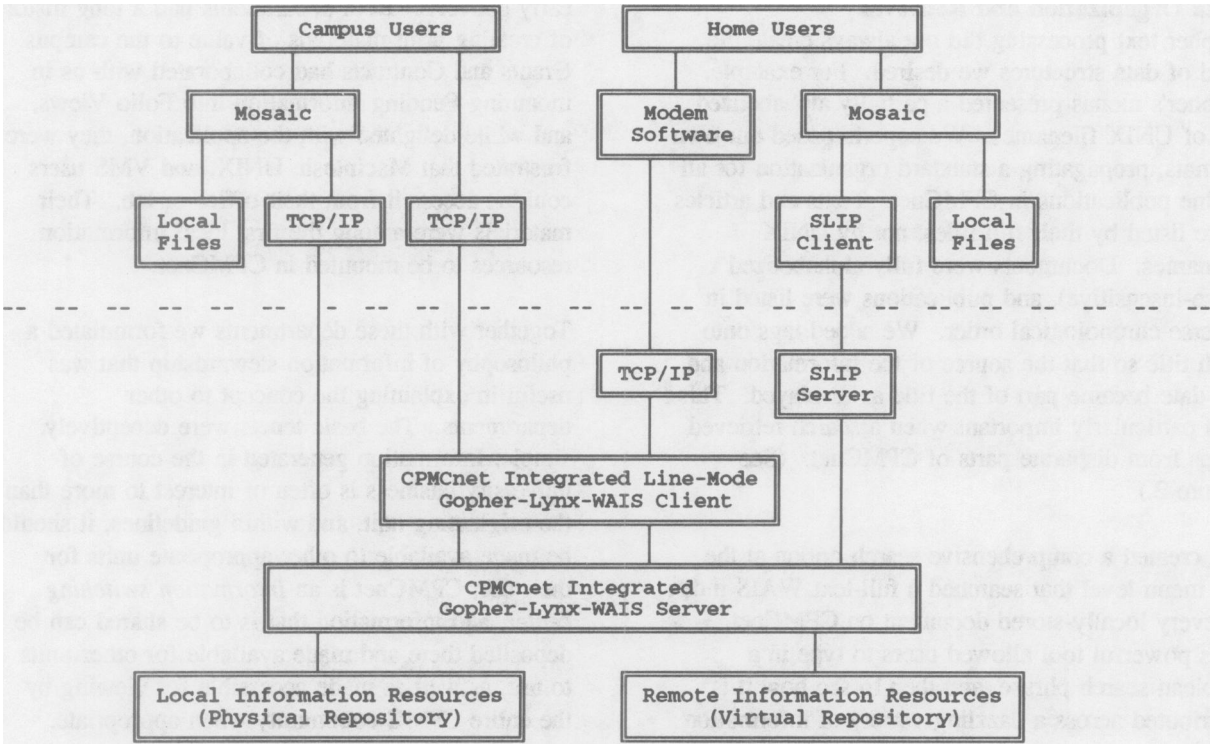

Figure 1: CPMCnet connections, clients, and servers

clients, on the other hand, required direct Internet connections, and depended on the capabilities of each individual computer: the local computer must be on the Internet, have graphics capabilities, and be among the computers for which stable clients have been released.

We decided that the richest computing environment would be obtained by combining the best of each of the available tools. Several products were introduced in 1993 that made an integrated approach viable. The University of Minnesota, developers of the original gopher protocol, introduced Gopher+ clients and servers which, among other improvements, had the ability to indirectly connect with Web clients. Gopher+ clients could recognize documents in HTML (Hypertext Markup Language) and hand these over to say, Lynx, a WWW client/server application developed at the University of Kansas. This meant that a single document directory structure could carry both Gopher-style and Web-style documents; in addition, switching between the two protocols could be carried out automatically by the software.

A second development was the introduction in 1993 of Mosaic clients for Macintosh and Windows computers. Mosaic, the World-Wide Web graphic client developed by the National Center for Supercomputer Applications (NCSA) at the University of Illinois at Urbana/Champaign, had previously only been available for X-Windows computers. Windows and Macintosh clients meant that we could extend UNIX-based multimedia capabilities to much of the campus based on the HTML protocol.

Thus, our goal for the CPMCnet client was to allow users the ability to seamlessly access hierarchical Gopher menus, a mode familiar to nearly everyone; to navigate WWW hypertext links; and to carry out robust full-text searching in both domains. To accomplish this, we customized the source code of the Gopher+ and Lynx clients. We made the keystrokes for navigation, searching, printing, and help identical in the two clients. Additionally, we incorporated WAIS full-text searching which is accessed from both the Gopher+ and Lynx portions of the integrated client. Finally, users with Mosaic or other graphics clients could use CPMCnet as a starting point to navigate multimedia applications. This will be increasingly important as we develop a graphics-rich electronic curriculum environment at CPMC [7]. (See Figure 1.)

Data Organization and Retrieval
Gopher text processing did not always create the kind of data structures we desired. For example, Gopher's menus presented a partially alphabetized list of UNIX filenames. We superimposed our own formats, propagating a standard organization for all online publications in CPMCnet. Texts and articles were listed by their full titles, not by UNIX filenames. Documents were fully alphabetized (case-insensitive), and publications were listed in reverse chronological order. We added tags onto each title so that the source of the information and the date became part of the title as displayed. This was particularly important when a search retrieved items from disparate parts of CPMCnet. (See Figure 2.)

We created a comprehensive search option at the top menu level that searched a full-text WAIS index of every locally-stored document on CPMCnet. This powerful tool allowed users to type in a Boolean search phrase, and then to see how it is distributed across a dazzling variety of information sources.

As an example, a WAIS search for the word 'renal' brought up i) the full-text of an article from P&S Medical Review on sickle cell disease; ii) abstracts of several SCAMC papers presented by Columbia University affiliates; iii) a Medical Logic Module on renal failure; and iv) sections of a CPMC Campus Profile text that reflected which departments on campus were involved in renal-related research. Because of the tags in each title mentioned earlier, a WAIS search that brought up the 'renal' articles showed which journal the articles were from as well as the date of publication.

Problems
Some problems have proved intractable for a small development group. For example, for resources that are restricted to Columbia and/or CPMC affiliates, Medical Center users still must telnet to ColumbiaNet, which has robust Kerberos-based authentication capabilities. A second problem is that widespread installation of Mosaic awaits an upgraded hardware base at the Medical Center.

SOCIO-POLITICAL INTEGRATION

Having a unified client and effective presentation format, all we needed was data! The Offices of Grants and Contracts and External Relations were early converts. Both departments had a long history of creating print materials of value to the campus. Grants and Contracts had collaborated with us in mounting Funding Information into Folio Views, and while delighted with the application, they were frustrated that Macintosh, UNIX, and VMS users couldn't access it from their office or lab. Their materials were among the first local information resources to be mounted in CPMCnet.

Together with these departments we formulated a philosophy of information stewardship that was useful in explaining the concept to other departments. The basic tenets were deceptively simple: Information generated in the course of university business is often of interest to more than the originating unit, and within guidelines, it should be made available to other appropriate units for their use; CPMCnet is an *information switching center*, and information that is to be shared can be deposited there and made available for other units to use, as well as made accessible for viewing by the entire CPMC community when appropriate.

Information Provider Agreements
Other departments came on board, and we were soon faced with telling them that their information was "in the queue". The large number of departments and information sources that provided data to CPMCnet meant that the task of processing text became complicated and time consuming; in order to handle this flow of information with limited staff, it was vital to streamline both the processing and the delivery of the text.

The first problem encountered was the diversity of software used by information providers. Materials for CPMCnet were received in a variety of different word processing packages, in desktop publishing format, in straight ASCII, in database and other software formats on Macintosh, UNIX and MS-DOS platforms. A second problem arose with the page layout and markup of each text. There was, to start, no consistency in the way chapters and sections were marked and virtually every document had a different font size and margin setup. A third problem arose with the physical delivery of information; documents were delivered via UNIX email, MS-DOS email, FTP, hand-delivered on disks, and, in a few cases, via hard-copy. To foster participation as we got started, we took information however we could get it.

```
C P M C n e t

CPMC Publications

1. Biomedical Frontiers/
2. The Reporter/
3. P&S Medical Review/
4. Information News Newsletter/
5. Health Sciences Guide/
6. University & Health Sciences Campus Profile/
```

```
>>
C P M C n e t

P&S Medical Review

1. ABOUT the P&S Medical Review.
2. SEARCH full text of all issues of P&S Medical Review [PR] <?>
3. Current Issue of P&S Medical Review/
4. Past Issues of P&S Review/
```

```
>>
C P M C n e t

Current Issue of P&S Medical Review [PR]

1. [PR:Nov:93] Letter From the Editor.
2. [PR:Nov:93] Antibiotic Resistance--Its Impact on a Great Medical C..
3. [PR:Nov:93] Acute Injury to the Adult Spleen: Evolution in Diagno..
4. [PR:Nov:93] Treatment of Gastric Lymphoma.
5. [PR:Nov:93] A 28-year-old Black Man with Painful Sickle Cell Crisi..
6. [PR:Nov:93] The Whipple Procedure - 1935 to 1993.
7. [PR:Nov:93] P&S Medical Review: Information for Authors.
8. [PR:Nov:93] P&S Medical Review: Editors.
```

Figure 2: Integrated interface for publications on CPMCnet

In order to manage the flow of information, we developed a two-way "Information Provider Agreement" which specified the respective responsibilities of the departments that provide information and of the Office of Scholarly Resources, which manages the information. The agreement was adapted for each information source so that it was as specific as possible for each source. The finished agreement specified the software source of the information, the page layout and markup of the documents, and the timeline and method of delivery of updates. The agreements provided clear instructions for the providing department about how to prepare and deliver documents. Since each information source is slightly different, we have not created a single monolithic agreement which all must abide by, rather we customize the agreement for each source. This does mean that each source involves a development process unique to itself; however, once the initial development has taken place, future updates can remain consistent and require less development effort.

New Lines of Communication

Another kind of integration was provided by feedback options distributed throughout the CPMCnet menus. Users could quickly send comments and questions to the CPMCnet developers or suggest new information sources. More importantly, patrons could quickly send questions or comments to the diverse and geographically dispersed providers of the information, crossing barriers of buildings, elevators, streets, and departments. To facilitate this, all full-text sources have an 'About' file that describes who is responsible for maintaining the information. Several departments asked us to incorporate "feedback" options in their menus. (See Figure 3.)

When navigating resources beyond CPMC, the user was often unaware of where the information resided. We decided that references to the origin of a data file should be provided on the menus in order to clarify the notion of responsibility for information and/or for its presentation or format.

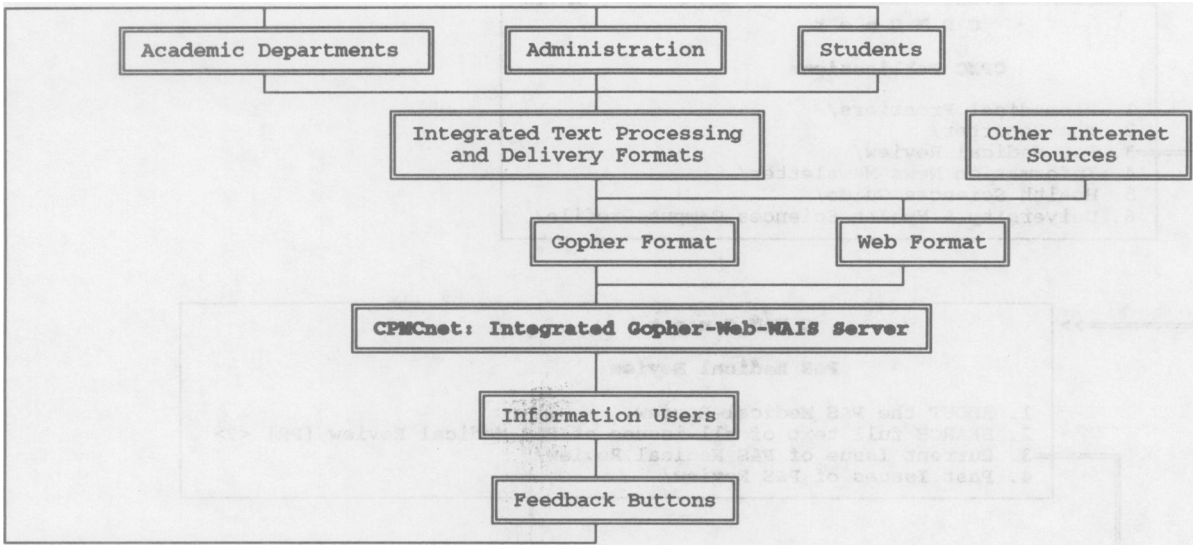

Figure 3: The information integration loop

Library Expertise

Another kind of cross-department integration can be anticipated as the Library brings to bear its expertise in end-user searching and cataloging, helping departments make their information more usable and retrievable. The Library's role in collection development and cataloging is seen as key to providing sensible and navigable pathways to the mass of information on the information superhighway. It is possible to organize the menus by subject area, by geographic location, and so on. Librarians, who routinely work with end users, already understand how information is sought and used by CPMC faculty and students. Translating this into the provision of electronic pathways simply uses the skills of librarians in new ways.

CONCLUSION

"The link between information and influence has often been noted; data are a political resource and building a database is often a political act" [8].

This rather telling quote makes clear that multiple agendas sometimes conflict. As we work to integrate information across administrative departments and academic units, to form new collaborative relationships between the library, its network and computer division, and the rest of the campus, and to move the academic enterprise forward in its rational and productive use of information, the road is not totally smooth. At CPMC we have used the notion of integration to provide some overarching principles and a degree of logic to the sharing of information resources.

References

[1]. R.I. Benjamin, M.S.S. Morton. Information technology. *INTERFACES*, 18:3, May-Jun 1988, 97.
[2]. J.J. Branin, G. D'Elia, T.W. Shaughnessy. Implementing an integrated information center in an academic setting: Research in progress. *Proceedings of the 56th Annual Meeting of the American Society for Information Science*, 30, Oct 1993, 217-220.
[3]. J. Zucker, R.M. Kahn, N. Narayanan. Clinical, scholarly, and campus information hypertext tools at Columbia-Presbyterian Medical Center. *Proceedings of the 17th Annual Symposium on Computer Applications in Medical Care*, Oct 1993, 539-543.
[4]. Campus computing environment: Columbia Univ. *CAUSE/EFFECT*, 16:2, Summer 1993, 34-37.
[5]. F. Anklesaria, et. al. The Internet Gopher protocol. *Internet RFC 1436*, March 1993.
[6]. T.J. Berners-Lee, R. Cailliau, J-F Groff, B. Pollermann. World-Wide Web: The information universe. *Electronic Networking: Research, Applications and Policy*, Spring 1992, 52-58.
[7]. P. Molholt. Integrating libraries into the curriculum: The CHIPS project. *Emerging Communities: Integrating Networked Information into Libry. Svcs.*, Ann P. Bishop (ed), 1994, 275-84.
[8]. P.G.W. Keen. Communications in the 21st century: Telecommunications and business policy. *Organizational Dynamics*, 10:2, Autumn 1981, 65`.