Desktop document delivery using portable document format (PDF) files and the Web*

By Jean P. Shipman, M.S.L.S., AHIP
Associate Director
Information Resources Management

William L. Gembala, M.L.
Information Management Librarian

Jeffrey M. Reeder
Web Applications Developer, IAIMS

Brian A. Zick
Library Technician
Document Services

Mary J. Rainwater
Library Specialist Supervisor
Document Services

Health Sciences Libraries
University of Washington
Box 357155
Seattle, Washington 98195-7155

Desktop access to electronic full-text literature was rated one of the most desirable services in a client survey conducted by the University of Washington Libraries. The University of Washington Health Sciences Libraries (UW HSL) conducted a ten-month pilot test from August 1996 to May 1997 to determine the feasibility of delivering electronic journal articles via the Internet to remote faculty. Articles were scanned into Adobe Acrobat Portable Document Format (PDF) files and delivered to individuals using Multipurpose Internet Mail Extensions (MIME) standard e-mail attachments and the Web. Participants retrieved scanned articles and used the Adobe Acrobat Reader software to view and print files. The pilot test required a special programming effort to automate the client notification and file deletion processes. Test participants were satisfied with the pilot test despite some technical difficulties. Desktop delivery is now offered as a routine delivery method from the UW HSL.

INTRODUCTION

With full-text electronic journals and Web-based information, full-text desktop delivery of print information is a frequently requested service by clients of the University of Washington Health Sciences Libraries (UW HSL) in Seattle, Washington. This demand is further

accentuated by the geographic area served by the UW HSL. The UW School of Medicine serves as the medical school for a five-state region including Washington, Wyoming, Alaska, Montana, and Idaho, commonly referred to as the WWAMI educational program. UW faculty and students are located at major universities and numerous clinical sites throughout this WWAMI region. These clients are unable to visit the UW HSL in person to retrieve needed health sciences related information. Yet their information needs are as legitimate as those of clients situated at the main Seattle campus. To meet the needs of distant learners and

others, the UW HSL has conducted a pilot test providing scanned journal article information to clients located in remote clinical settings. This paper describes the techniques used to provide desktop full-text information on demand and the outcomes of the pilot test.

Pilot test rationale

A faculty survey recently rated desktop access to full-text information as one of the most desirable services to be offered by the UW libraries. Desktop document delivery addressed the libraries' strategic plan objectives:

1. Coordinate and develop client education programs and service models that actively promote client self-sufficiency.

2. Develop and implement electronic enhancements of client services.

3. Provide timely and cost-effective access to information resources through study, policy changes, and improvements of library operations.

Library clients now expect not only to identify relevant citations through the plethora of databases offered by the libraries, but also to retrieve full-text material at their office workstations, clinical units, and from home.

Previous Internet delivery

The UW HSL previously employed other methods for delivering needed journal articles via the Internet. Staff tested three different versions of the Lister Hill Center for Biomedical Communication/National Library of Medicine's DocView software and applied the Research Libraries Group's Ariel product as a transmission method with other libraries since 1993. There were advantages and disadvantages with both products. DocView operated only on computers with Microsoft Windows installed and required that the institution sending the scanned files create them using the Ariel software. Clients also had to obtain copies of the client DocView software, which was not released for public use except for testing purposes. The Ariel product originally required that both sending and receiving institutions have the Ariel software in place. In addition, Ariel is available only in DOS or Windows versions, thus precluding use by Macintosh platforms. Neither software program delivered documents to all types of computer platforms in a manner that could be easily distributed to local and distant education sites.

Pilot test goals

The purpose of the pilot test was to explore and refine a methodology for transmitting documents to clients' electronic desktops regardless of computer platform

by applying current systems capabilities that were widespread within the university system and that incorporated industry standards. The methodology used for this pilot test was also applicable to electronic reserves, distance learning support, and document delivery to other institutions' interlibrary loan departments.

LITERATURE REVIEW

As interest in digital libraries has continued to increase, so has the volume of published research relating to electronic document delivery projects. A search of the library, information science, computer science, and medical literature revealed numerous references to electronic document delivery. This literature can be categorized into two groups: (1) delivery technology that deals with issues related to fax, e-mail, and Web delivery, and (2) scanning software that uses mimicry techniques to deliver articles electronically.

Delivery technology

Since the arrival of affordable and adequate fax technologies in the early 1980s, individual libraries have utilized fax machines to transmit articles to other libraries and individuals. A number of studies researched the possibility of the use of fax for document delivery. A large overview of early facsimile technology in libraries was reported by McQueen and Boss in 1983 [1]. The use of fax machines in medical libraries was reported soon thereafter [2–3]. Since that time, more advanced models of fax document delivery have been reported including the use of online public access catalogs (OPAC) to allow fax delivery on demand [4].

Articles on scanning and e-mailing of documents are a phenomenon of the 1990s. Research on this type of document delivery includes the MARCEL project at the University of Northern Colorado, which used Multipurpose Internet Mail Extensions (MIME) to attach scanned articles to e-mail messages [5]. Scanned documents are often converted to ASCII text and delivered via the Internet; research examples include OCLC's First Search [6], and UMI's ProQuest [7]. The problems associated with ASCII documents are largely aesthetic and preclude the delivery of tables and illustrations [8]. A third delivery technology, the Web, has been applied to interlibrary loan request forms [9], but only recently to client desktop document delivery [10]. Although many copyright-free government publications, such as the *Morbidity and Mortality Weekly Report* (*MMWR*), are distributed electronically via the Web, copyright has been a barrier because no formal laws regarding electronic articles have been implemented.

Mimicry techniques

Many recently published document delivery studies describe the use of mimicry software, which produces

Figure 1
Article Retrieval Page

University of Washington
Health Sciences Libraries

PDF Article Retrieval

To retrieve your article, below enter your claim number sent to you via email. Click on "Submit".

[ ] Submit

Contact Document Services at (206)543-3436 or 543-3441 if you have questions or experience difficulties.

If you have never retrieved a pdf article before, please see our Instruction Guide.

URL: http://docser.hslib.washington.edu/docser/uw.html
Last Updated: July 17, 1997

---

a duplicate of the original printed document in electronic format. An overview of mimicry software by Jacso explains the issues involved [11]. Examples of mimicry software in the scanning software group include Ariel and Adobe Acrobat Capture. The Ariel document delivery system sends scanned documents to participant libraries over the Internet and allows the receiving library to print documents that simulate the original [12]. However, the Ariel system is not designed for direct delivery to library clients. Adobe Acrobat Capture is another example of mimicry software. The Adobe portable document format (PDF) is increasing in popularity, particularly because of its usability, freeware status, and widespread use by the United States Government [13].

METHODS

Pilot test participants

Pilot test participants were located within various remote UW clinical and research sites, including the Roosevelt Clinics and the Harborview Medical Center, both located miles from the main health sciences campus. These sites were selected for several reasons. First, these sites delivered patient care, requiring that information be available in a timely fashion. Also, clients in these clinical settings often did not have the time to travel to a library to obtain needed information so they provided a good virtual client model. Individuals were recruited through various campus presentations, contact with the director of residency education, librarian contacts, and word of mouth.

Pilot test phases

The Desktop Document Delivery ten month pilot test was divided into two phases. Phase I involved scanning articles into PDF files, which were sent as MIME standard e-mail attachments. Participants learned how to detach the scanned article from the e-mail message, download the article, and view the article using the free Adobe Acrobat Reader software.

During Phase II of the pilot test, articles were scanned into PDF format and stored on a Web server. Participants were notified by e-mail of article availability and were given a claim number and instructions for retrieving the article from the Web server. Participants entered their claim number on a Web page (Figure 1) and saved the retrieved article. While the scanned articles were archived on the Web server for two weeks, the claim numbers could only be used once. If an error occurred during the downloading process, the participant could request that the file be re-linked to the claim number.

Equipment used

Initial scanning was performed using a workstation that included a Dell Pentium 90 personal computer with 32MB RAM and 1GB hard drive, and a Hewlett Packard Scanjet 2C scanner. Through development efforts with Canon, a scanning workstation consisting of a Canon GP30F photocopier scanner and a Dell Pentium Pro 200 personal computer with 32MB RAM and 2 GB hard drive were later used to provide faster scanning processing.

Figure 2
Operator Page

Document Delivery Operator Section
Mail Notice of Completed Request

Requestor's email address
○ @u.washington.edu
○ @dehpost.sphcm.washington.edu
○ @hermes.hslib.washington.edu
○ @

Affiliation University of Washington

PDF file name

DOCLINE Number

Carbon copy requestor's email notice to*
○ @u.washington.edu
○ @dehpost.sphcm.washington.edu
○ @hermes.hslib.washington.edu
○ @

System design

An in-house system was designed to facilitate test operations. Two major components of the system included a Web server, which stores PDF files; and a database that tracks requests, sends e-mail notification messages, monitors PDF file downloads, and deletes PDF files. A key system feature was easy configuration. Many of the configuration options were stored in the database, such as standard appearance elements for HTML pages, the location of the PDF file directory, and the name of the mailhost computer, for sending e-mail messages. Most Web pages were then generated on the fly, which allows stored parameters to be dynamically incorporated into a page.

Requested articles is scanned using Adobe Acrobat Capture software and saved in a PDF file directory. Staff complete an "operator" form (Figure 2) that indicates the client's e-mail address, the PDF file's name, and the client's affiliation (e.g., University of Washington). To complete the automatic notification, the existence of the PDF file is verified by the database, a unique article claim number is generated, and an e-mail message is sent to the client via a script.

As no legal decision has clarified the copyright status of transmitting documents via the Internet, UW HSL designed its system to comply with current copyright law. Access to PDF articles is restricted to the requesting individual client by employing a claim number. The link between the claim number and the scanned article file is broken after the first time the client attempts to retrieve the file from the server. If the client never retrieves the file, it is automatically deleted after two weeks. Using these techniques, the files are maintained on the server for a limited time and never reused or provided to more than one client. Copyright warning statements are included on all request forms as well as on the notification e-mail and as part of the PDF file.

At the conclusion of the pilot test, the decision to extend this delivery method to non-affiliated libraries (interlibrary loans) and their clients required changes to the original system design because borrowing libraries needed to be notified when article files had been delivered directly to their clients. This was accomplished through modification of the operator form to include multiple client affiliations, the entry of a DOCLINE number, and the ability to send a carbon copy of the e-mail notification to borrowing library staff.

Request method

Test participants sent their requests for PDF articles to an e-mail address that was designated for pilot test requests only. This assured that requests were identified for special handling by staff trained in PDF processing.

Instructions created

Participants often did not have the required software or skills to participate in the two phases of the pilot test. User instructions were developed to assist in ob-

taining the necessary computer software and information needed to obtain the scanned articles. Instructions were created for saving and downloading MIME e-mail attachments; using WS_FTP and Fetch file transfer protocol (FTP) software; installing Netscape; installing the Adobe Acrobat Reader software; and installing WinZip. Separate instruction guides were created for Windows 95, Windows 3.x, and Macintosh clients. Each guide included numerous illustrations and was about five to eight pages long. Abbreviated instructions were also created for more experienced participants. These instructions were provided via a Web page and mailed upon request.

RESULTS AND DISCUSSION

Pilot test participant survey results

At the end of each phase, a survey was distributed to test participants (Appendix A). There were four participants in Phase I, but only two replied to a telephone evaluation survey. Participants received oral instructions before attempting to install the software, and consequently they reported no problems with installation. Downloading of articles took between two to ten minutes and printing took two to three minutes. The PDF documents received were of good quality.

Phase II had twenty-one participants, sixteen of which replied to the mailed survey (76%). Overall, the participants were pleased with the delivery method and with the quality of the documents. Test participants used either Web-based or print instructions to install the software and reported few problems. One participant reported being unable to retrieve or print his documents. Test participants reported print times ranging from one to thirty minutes per page.

Scanning outcomes

During Phase II, a total of 362 articles were scanned and delivered to twenty-one different participants. A staff log was maintained that included scanning times, processing times, file sizes, and problems encountered. The average file consisted of seven print article pages, which translated to an average file size of 1,072 KB. The file size grew to an average of 2,669 KB for dithered documents (dithering provides better graphic resolutions). Different scanning times and file sizes were achieved depending upon computer equipment power and whether hardcopy originals or photocopies were scanned. Photocopying materials and then scanning the photocopies actually reduces staff scanning time because a document feeder can be used. The average scan time was 2.7 minutes per seven-page article. It was determined that 178 articles, or 1,500 pages, can be scanned during an eight-hour day using high-speed computers and scanning equipment.

Printing times

The average reported time required to print a single page was 1.4 minutes for text documents and 2.9 minutes for documents with graphics.

Graphics

The UW HSL staff experimented with scanning images that are often unique to the health sciences field including photomicrographs and gels. Dithering the gels at a resolution of 400 dots per inch (dpi) produced the best reproduced image of the original. The usual scanning resolution employed was 300 dpi.

Costs

The additional cost for scanned article delivery was calculated to be $0.91 per article based on a daily output rate and cost assessment, which included student staff time, hardware and software costs, and photocopy charges. This amount did not include costs associated with pulling articles from shelves, photocopy costs, technical support assistance, problem solving, or reshelving. The costs of operation exceeded expectations. Learning scanning techniques took staff time in addition to time spent performing daily scanning.

Problems encountered

Problems could be classified into three categories: technical, operational, and client participation. Technical problems centered primarily on participant difficulties. Downloading PDF files demanded a fairly powerful computer. Often the PDF files were large enough to take up a substantial portion of the participant's e-mail account in Phase I and hard drive storage capacity. Some participants were bothered by display and printing problems such as older laser printers demanding wider margins resulting in text lost from page edges. Smaller monitors often could not display the full PDF image, requiring scrolling back and forth in order to view the full article.

UW HSL staff also grappled with scanner training issues. Scanning an article is a more difficult and less intuitive procedure than photocopying. Highly trained staff were required to handle even routine problems. Even after several months of experience with scanning to PDF format, staff still ran into problems that required a supervisor's intervention to resolve. For computer literate staff, the minimum training time required for very basic operation of the Adobe Acrobat Capture software was about thirty minutes. However, in order to work with the complexities of the software and to resolve problems, a recommended training time of fifteen hours would be more realistic.

Finally, the number of test participants did not meet expectations. Despite initial enthusiasm and expressed interest, motivating people to participate in the pilot

test was difficult. Potential test participants were either intimidated by the technical requirements of the pilot test (e.g., downloading the Adobe Acrobat Reader software, FTPing files, and manipulating e-mail messages), or did not have the time to participate.

CONCLUSION

The pilot test proved to be a success and our clients agreed that the desktop delivery method should be offered as a routine delivery choice. Electronic request forms were revised to include electronic delivery as an option and availability was announced via the UW HSL's newsletter, *Books & Bytes.* A special flyer was distributed with all filled document delivery requests. An e-mail announcement was sent to departmental mailing lists and to clients who had expressed interest in the service as well as to previous test participants. For the first five months, 148 articles were sent using desktop document delivery for no extra charge. Starting January 1998, requestors will be charged a fee for this delivery service.

What factors should a library take into account when considering offering desktop delivery of documents?

Clients should have basic computer literacy skills, computer equipment with an adequate amount of power and memory, a laser printer, and required software. A suggestion for assuring client technical competency is to establish a test PDF file. Clients should be encouraged to successfully access and print the test file before requesting PDF delivery.

Library staff should decide to whom they will provide the service, how requests can be placed, how clients will be notified of file availability, and how copyright will be addressed. Quality computer and scanning equipment is essential. Staff training time also needs to be provided. Client technical support boundaries also need to be established.

Desktop document delivery will appeal to a wide range of individuals. Remote delivery of articles is of immediate benefit to clientele spread over a large geographic area. Remote clinical clients will also enjoy the speed with which articles can be delivered compared to standard United States mail. Researchers will appreciate the improved resolution quality over that provided by fax technology. Desktop delivery will also be of immediate use to the disabled client for whom a physical visit to the library would be difficult if not impossible. Further, as the technology becomes more streamlined, this method of delivery is likely to become popular with the average client.

REFERENCES

1. McQueen J, Boss RW. High-speed telefacsimile in libraries. Lib Tech Rep 1983 Jan–Feb;19:7–111.
2. Pfaffenberger A, Deagle JH, Pfaffenberger RC. Rapid delivery of medical information to remote hospital sites by telefacsimile. J Am Osteopath Assoc 1984 Apr;83(8):604–13.
3. Hover LM. Telefacsimile service in a hospital consortium. Bull Med Libr Assoc 1987 Jan;75(1):35–6.
4. Whittaker M, Malamud J. UnCover: the article access solution. Bull Med Libr Assoc 1994 Apr;82(2):181–2.
5. Smith J, Delaney T. Marcel: MIME prototype study in electronic information delivery. Interlend and Doc Supply 1996 Jan;24(1):24–27.
6. Tenopir C. A second look at FirstSearch. Libr J 1994 Nov 1;119(18):30–32.
7. Combs, Jr J. ProQuest Direct: remote research and electronic document delivery from UMI. Libr Soft Rev 1997 Mar; 16(1):50–63.
8. Balas J. Online delivery service. Comput Libr 1995 Oct; 15(9):32–34.
9. McCloskey J. Web-based forms for ILL using HTML. J Interlibr Loan Doc Deliv Inf Supply 1996;7(1):79–88.
10. Dzierzak E, Hayes K, Lucas W. Delivering the goods: Internet delivery of printed material. Proc AMIA Annu Fall Symp 1997:934.
11. Jacso P. Mimicry software: a new option for electronic document delivery and transfer. Online 1995 Mar–Apr;19(2): 74–81.
12. Jackson ME. Using ARIEL, RLG's document transmission system to improve document delivery in the United States. Interlend and Doc Supply 1992 Apr;20(2):49–52.
13. Balas, op. cit.

Appendix A
Desktop Document Delivery Project - Phase II
Evaluation Survey Summary Results *

Installation

	Excellent	<------------------------------>		Poor
	5	4	3	2	1
Ease of installation of necessary software was:	6	4	2		1

What problems did you encounter?

* 4 participants reported no problems.
* 2 participants reported minor problems but were ultimately able to install the software.
* 1 participant reported that his student assistant was never able to load the software on the student computer. The participant was not able to troubleshoot his student's problem and he reported HSLIC was unresponsive to email queries.

If you installed the following software, how easy was it to install?	Very Easy	<------------------------------>			Impossible	
	5	4	3	2	1	N/A
Adobe Acrobat Reader 2.1		1	1	1		6
Adobe Acrobat Reader 3.0	6	3	3	1		2
Netscape	4	1				8

User Instructions

Which version of the instructions did you use if you referred to the Health Sciences Libraries user instructions? (Circle all that apply)	Web-based	Print	Nothing
	7	5	3

	Excellent	<------------------------------>		Poor
	5	4	3	2	1
Instructions for installation and use of software were:	6	3		4	

How could the instructions be improved?

* 1 participant suggested that it was not clear that Acrobat be downloaded before attempting to retrieve the document.
* 2 participants reported that the instructions did not apply to their particular situation and would have liked alternative procedures.
* 1 participant suggested that the instructions should assume less computer literacy.

* All participants did not necessarily respond to every survey question.

What improvements do you recommend?

> * 4 participants would like faster service.
> * 1 participant would prefer to retrieve a group of documents with one retrieval code.
> * Several participants complained about slow printing. 1 participant suggested scanning articles at a lower density to facilitate faster printing.
> * 1 participant suggested that we allow access to articles more than one time so if errors were made in saving or printing, he would not have to request the article again.
> * 1 participant would like improvements to print quality (i.e. missing characters along the edges of the page).

General

Please provide information on your computer configuration (IBM-compatible PC or Mac, Printer, Windows 95 or Windows 3.1):

> * Power PC 7100 - 16 Mb RAM, 2 G harddrive
> * Pentium, HP LaserJet Printer. Windows 95 32 Mb RAM, 2 Gb HD
> * Power-Mac System 7.5.5
> * Mac - Power Macintosh 7200/120
> * IBM-compatible PC, Windows 95, Used Internet Explorer to retrieve articles.
> * Mac Quadra 650
> * Power Macintosh 6100/66 Hewlett Packard LaserJet 5si mx
> * Mac, Quadra 650, Apple personal laser writer
> * Mac
> * Work - Mac Power PC 7100/80 486 PC Compatible Printer - Several Lasers avail.
> Home - 486 PC Compatible + Okidata
> * Mac Windows 95
> * Macintosh II SI
> * Mac computer HP Deskjet 850c printer
> * Pentium 133 Windows 95 32 Mb memory
> Hewlett Packard LaserJet 5 si/5 si mx ps
> * PC, Windows 95, HP LaserJet IV

	Computer Guru	<----------------->		Novice
	5	4	3	2	1

How would you rate your level of computer expertise, in general?

	5	4	3	2	1
		4	10	2	

Additional Comments:

> * 3 participants reported great satisfaction with the service.
> * 2 participants reported frustration trying to retrieve articles.
> * 1 participant would like to have a back-up way to retrieve claim numbers.
> * 1 participant suggested that the service should cover journal titles held elsewhere on campus.

Cost

If we were to use Desktop Delivery to deliver documents to you in the future, what would you pay above our base charge ($4 budget, $4.98 cash) to have articles delivered to you via the Internet? (circle one)

$0	$0.01-$1	$1.01-$2	$2.01-$3	$3.01-$5	$5.01 -
5	5	5			

Documents

	Excellent	<--------------------------------->			Poor
The quality of the document(s) you received was:	5	4	3	2	1
	8	5	2		

What problems did you have with retrieving the documents? Were they with the computer image or print quality? Did you request re-scanning of any articles, or parts of articles?

> * 5 participants reported no significant problems in print quality.
> * 5 participants reported margins cut too closely or missing letters along the margins of the page. 1 participant reported that he requested re-scanning a few times.
> * 5 participants reported being unable to retrieve or print their document(s).

Overall

How long did it take to retrieve an article on average?

> Participants reported times ranging from a few seconds to one week.

How long did the article take to print on average?

> Participants reported times ranging from 1 minute per article to up to 30 minutes per page.

In general, when you retrieved your article did you: (circle all that apply)	Save	Print	View
	9	15	8

How important to you is the ability to manipulate text within the document (ability to move text and copy to other applications)?	Very Important		<------------------>		Not Important
	5	4	3	2	1
		1	3	4	9

If Desktop Document Delivery was offered for the delivery of articles on a regular basis, would you use this service?	Yes	No
	14	0

	Excellent	<--------------------------------->			Poor
	5	4	3	2	1
Your satisfaction with the delivery method was:	9	2	4		1