

The Impact of Electronic Bibliographic Databases and Electronic Journal Articles on the Scholar's Information Seeking Behavior and Personal Collection of "Reprints"

Robert B. McGeachin

SUMMARY. This article examines the potential changes to the information seeking behavior of scholars and how they manage their own collection of research article "reprints". With bibliographic databases and electronic journals provided by academic libraries now available at the science scholars' computer desktop, they can now locate and acquire a portion of needed research articles on their own at any time. They also, in some cases, have older paper copies scanned and delivered by libraries as image files at web retrieval locations. Bibliographic citation management software is now in use by many scholars. Personal information management software is available and could also be used. This article reviews possible scenarios scholars can use to manage this new electronic collection of research articles and possible ways libraries can help them in this scholarly activity.

KEYWORDS. Bibliographic citation managers, personal information managers, pim, ocr, reprint collection, scholar's workstation, remote access, library services

Introduction

Most academic scholars have usually amassed personal collections of works relevant to their scholarly research and communication. Books and journal articles have been available and collected for centuries. Published reprint copies and the advent of photocopiers greatly expanded the ability of scholars to collect large numbers of articles. Large collections have brought a need for organization of the collections and creation of access and retrieval mechanisms for them. Today, this is often accomplished by such traditional written and mechanical means as personal card catalogs to a collection of reproduced articles filed by author, title, or subject classified

Robert B. McGeachin, MS, MLIS, PhD is Associate Professor, Head of the West Campus Library and Director for Business and Agriculture Services, West Campus Library, Texas A&M University, 5001 TAMU, College Station, Texas 77843-5001 (e-mail: r-mcgeachin@tamu.edu)
Published in *Science & Technology Libraries*, Vol. 25, No. ½, 2004, pp.127-137.

arrangements. Some scholars currently use bibliographic database software such as Pro-Cite, EndNote, or Reference Manager to replace card catalogs for organizing and gaining access to their personal print collections.

Dream of the Scholar's Workstation Coming to Fruition

The advent of electronic databases and bibliographic software to replace physical card catalogs and the creation of large full-text databases of works have led to the verge of realizing Vannevar Bush's dream of the "memex" portion of a scholar's workstation (Bush 1945). Bush envisioned an increase in a scholar's work efficiency through the capability of full-text and image retrieval of a personal collection of scholarly information at the scholar's desk. Most scholars today deal with a narrowly focused subspecialty and need a large, organized personal collection of articles and information as well as access to acquiring new information in their field. They also need an efficient way to gather, store and retrieve raw research data. The ability to do so electronically has been evolving over the last couple of decades. This is now technically possible, but the best means of acquiring and managing a personal full-text electronic collection needs to be determined by the scholar who now has more available options than ever for personal research information collection management. Ultimately individual scholars must decide what are the most convenient methods to use based on their level of comfort with physical print resources versus electronic resources.

Managing Print Information Resources

Scholars have traditionally collected reprints by writing to the author of articles of interest and requesting a courtesy copy. They have also gone to their institution's library,

browsed relevant journals and photocopied articles from these journals. Articles are then often indexed with some form of personal card catalog and stored according to an indexing scheme of choice. This is usually either alphabetically by author's name or some personal subject classification system devised by the scholar to fit specific research areas. If subject classified, they are often arranged alphabetically by author's name within the subjects. The average time to accomplish this traditional discovery, copying, cataloging and storage is 7.4 minutes per article (note: this assumes multiple articles are retrieved in any session at the library, and this does not include travel time to and from the library which will vary for each scholar) (McGeachin 1998). This may still be a reasonable method for some scholars, especially if they have clerical or graduate assistant help available to help with the cataloging and storage tasks. Scholars also acquire a personal book collection, which may also be arranged alphabetically by author's name or alphabetically by title and cataloged as well. Most scholars are members of a relevant professional society which in most cases produces journal and/or trade publications that are either included in the cost of an annual membership dues package or available at a relatively low additional fee/s. This personal journal collection is not usually indexed or cataloged by the scholar, just arranged in chronological order by title on bookshelves or perhaps in pamphlet boxes on shelves. Most scholars have from one to five personal subscriptions to their most frequently used journals, as a convenience.

With the advent of growing amounts of information resources in electronic format provided by their institution's library and made remotely available at their desktops scholars can find articles and print them on a local printer. This is preferable to trying to read the articles online for many scholars and produces a physical copy to add to their existing collection. Such electronic to print articles can be cataloged and filed into the existing reprint and photocopied

article collection for later reference. These remotely accessible electronic articles are a great convenience for the scholar, saving them large amounts of travel and photocopy time at the library. In many cases, if their local library provides remote electronic access to titles that they have as a personal subscription for which they pay extra above the base cost of their society membership, they are choosing to drop their personal print subscription and just use the equally convenient library supplied electronic version.

Conversion of Legacy Print Collections to Electronic Files

For scholars with an existing collection of print resources that have taken up electronic storage and retrieval of new electronic resources, the question of whether to convert their existing print collection to electronic format is raised. They must decide if the time it would take to do so is worthwhile for them or not. To do so they would have two choices. The first is a conversion process of scanning their existing documents, followed by Optical Character Recognition (OCR) on the scans and cleanup editing on the OCR results before storing the new version of the document in some file format. But this can be a time consuming process. A 1998 investigation found that using PageKeeper version 2.0 software (equivalent to about OmniPage version 6.0 OCR software) to scan and perform OCR took an average of 62 minutes per article to process, which is far too much of an effort to be worthwhile for most scholars (McGeachin 1998). Now, five years later, there have been increases in the OCR abilities of software. For example, OmniPage is currently in version 12.0, and a repeat of this time and effort study should be done to determine how long conversion of the average article would now take.

The second choice is to scan and convert the documents to Portable Document Format (PDF) files for storage. While this is relatively fast and the same file format in which many

electronic resources are now being saved, the resulting files are relatively large and need a lot of storage space. Unless further processed with OCR, they are only image files that cannot be searched. One of the new native file output types of OmniPage version 12.0 is PDF with OCR done on the output files such that these combination image and OCR text PDF files are fully searchable. But the relatively large size of PDF image files may lead some scholars to save a smaller version. Products that can convert PDF files to much smaller file size Microsoft Word documents include OmniPage Pro version 12 (ScanSoft 2003a) and PDF Converter for Microsoft Windows (ScanSoft 2003c).

Electronic Information Resources

Scholars read and most frequently cite in their publications those journals that are most convenient to them in terms of access. A bibliometric analysis of publications by molecular biologists at the University of Chicago showed that, on a campus with multiple libraries with scientific journals, they used and cited most frequently those titles that were in the library closest to them (Hurd, Blečić, and Vishwanatham 1999). So, the convenience of supplying electronic bibliographic databases for resource discovery and electronic journals and books for easy retrieval to the scholar's desktop is extremely important and will directly influence scholarly reading and citing habits and most likely the outcome of research efforts. Many scholars have not used print bibliographic indexes as their primary resource discovery tools. Rather they use easier methods, the most frequent of which are to consult with colleagues and/or just follow and use literature cited in articles they already have. Hallmark (1994) found that two-thirds of all citations in publications by scientists came from these two sources. Brown (1999) also found that scientists' most used means of finding current literature were browsing current journal

issues, conversations with colleagues, and attending scientific conferences; with less than half using bibliographic indexes. So, with the provision of electronic bibliographic databases at their desktop, broader and more complete literature searches and reviews of the literature on an area of research can be performed by scholars, which should result in improved scholarly and scientific results. But more complete knowledge of the literature leads to increased acquisition of personal article copies and the need for more efficient retrieval of articles, personal indexing, and storage capabilities.

Scholars can now access electronic copies of articles at their desktop and have a number of choices of how to then deal with them. If they have confidence that they will have perpetual electronic access to articles, they may just want to maintain a personal index of bibliographic records and corresponding URLs included for easy future access to the articles. One group of software tools that can perform this function are bibliographic citation managers. There are now a growing number of these available for use by scholars. They usually include such features as the ability to:

- Create a database/s of records to a variety of resource types (articles, books, electronic resources, chapters, manuscripts, proceedings, AV materials, etc.)
- Import selected records from electronic bibliographic databases
- Search and retrieve records from bibliographic databases and online catalogs from within the citation management program
- Format the citations in any of hundreds of different journal or association citation styles and produce properly formatted, complete bibliographies of selected records in any of those styles

- Integrate with common word processing software such as Microsoft Word or WordPerfect in the creation of articles with inserted citations and final complete bibliographies (for example, this article was written using Microsoft Word 2000 integrated with EndNote 5.0 to create the citations and bibliography)

Available bibliographic citation-manager products now come as either purchased desktop software applications that operate from the local desktop computer and/or local area network server, or as subscription web-based applications hosted on the vendor's web server. The former do not require Internet access to operate, but the latter have the potential advantage of being accessible to users from anywhere they are located.

Desktop bibliographic citations managers include:

- EndNote version 7 (ISI ResearchSoft 2003a)
- ProCite version 5 (ISI ResearchSoft 2003b)
- Reference Manager version 10 (ISI ResearchSoft 2003c)
- Bibloscape version 5.3 (CG Information 2003)
- Reference Assistant (CrazySquirrel Complete Solutions 2003)
- Citation version 8.2 (askSam Systems 2003)
- Library Master (Balboa Software 2003)
- Scholar's Aid 4 AE (Scholar's Aid Inc. 2003)

Web-based bibliographic citation managers include:

- WriteNote (ISI ResearchSoft 2003d)
- RefWorks (RefWorks 2003)
- NoodleBib (NoodleTools 2003)

Most of these products have educational pricing available to make them more affordable in the academic scholarly environment.

One other option available to scholars that just want to capture an index of URLs to electronic resources is to use their web browser's "bookmark" or "favorites" functionality to capture and store URLs. With the bookmark record editing features and hierarchical ability to create layers of folders, scholars can accumulate very extensive collections of links to both electronic articles and web resources.

Another option now available to scholars, with the availability of accessing electronic versions of articles and books at the desktop, is to download a personal copy of the electronic file and store it locally. The increased availability of inexpensive desktop storage space or local area network storage space is one factor that makes this feasible for scholars. Again, the question of how to index, store and retrieve these local file copies is raised? One simple possibility that might work for some scholars is to save the files with the resource title in the file name, and create a personal subject hierarchy of folders in file storage utilities such as Windows Explorer. A simple level of retrieval can be accomplished with the Windows search utility and word searching within titles or documents, or by browsing through the file names in the Windows Explorer file display.

A more elegant and robust solution is the use of Personal Information Management (PIM) software to store, index and retrieve the electronic resource files. PIM is currently used to label and describe a wide range of software applications. These range from those focused on very specific types of information management such as Troopmaster which records and organizes information for Boy Scout Troops; to managing sales contact information; to managing personal calendars, address and phone books, e-mail and transfer of these to and from

Personal Digital Assistant hardware; to the one this article focuses on, which is broad and comprehensive storage, indexing and retrieval of almost all files and file types in the personal computer. This class of software is designed to be a more holistic approach to managing many more types of personal files than the bibliographic citation managers that just focus on records about articles, books and web sources. Common types of files PIMs are designed to store, index and retrieve include: image files of many formats, PDF files, text files of many formats (sometimes including e-mail), and web files. PIMs also usually allow existing files to be further annotated by the user. The user can also define the storage subjects/categories and create multiple cross-references among items.

Personal Information Manager programs include:

- PaperPort Pro 9 Office (ScanSoft 2003b)
- Enfish (Enfish Corp. 2003)
- TreePad (Freebyte.com 2003)
- CatClip version 2.5 (CatRunner LLC 2003)

Using PaperPort as an example, the scholar can use bibliographic indexes to identify relevant resources and link from them to the actual resources to examine online. If they choose to save an electronic copy they can “save as”, copy and paste, or drag and drop the file to a PaperPort folder. Scholars can create and use a hierarchical file folder structure in PaperPort to store all their downloaded and saved information resources. As this collection grows over time the extensive search capabilities of PaperPort allow scholars to locate and retrieve resources from their collection.

The scholar can use PaperPort to convert files from one type to another as it has a number of converters in its program functions, including the ability to do OCR processing on image files.

For example, a scanned image can be moved with drag and drop from a PaperPort folder on to the Word icon in the “Send To” bar, and PaperPort will perform OCR and convert the results into a Word document. Resource item files can also be sent to other programs such as Internet e-mail, word processing, spreadsheet, fax, graphics and online service programs. Other PaperPort functions include capturing web pages as image files, the ability to add electronic annotations, as either text or yellow sticky notes, and the ability to add subjects, authors and keywords to item properties to aid in search and retrieval.

The Role of Librarians in Assisting Scholars

A primary role for librarians is to continue to provide information resources to their client scholars in both print and electronic formats. But especially, providing remote access to electronic resources at the scholar’s desktop computer should be facilitated for their client scholars to the degree that libraries are fiscally able. This includes providing electronic bibliographic indexes and databases, electronic journals and periodicals, electronic books, and access to the Internet. Since all literature that scholars need does not exist in electronic form yet (or cannot be afforded as an electronic option yet for what does exist) librarians can also create local electronic delivery systems to take requests electronically for their print journal items; retrieve these from their collections, scan and convert them to into PDF documents, and place them in private web-based storage for only the requesting user to retrieve. In this case, an e-mail is sent to the customer with a URL for the location of the PDF file to use for retrieval. The PDF file of the article would be removed after being accessible to the customer for one month. So scholars must either print it off or save the PDF file themselves if they need long-term access. (This service scheme ensures that scholarly “fair use” of the original material has been observed

by only providing one copy on request to a single customer.) This makes essentially the entire print journal collection of the library accessible from the scholar's desktop.

Librarians can also play a role in assisting their customers manage electronic information once they have acquired it. They can assist in the usage of bibliographic citation managers either directly by providing access to them for library users, or indirectly by encouraging other computing related entities at their institution to provide them. For example, the author of this paper is on a campus committee that has oversight of student open-access computing laboratories at the Texas A&M University campus, and he successfully made the case to the rest of the committee to include bibliographic citation management software as one of the applications made available in the student computing labs. Another librarian at Texas A&M University created a "connection file" for this software package that allows users to search the Texas A&M University Libraries' online catalog from within the bibliographic citation manager and directly import selected catalog records into their citation manager (Highsmith 2002). Librarians can be knowledgeable in the use of bibliographic citation management and personal information management applications and offer instruction to customers in one-on-one consultations, group classes, and by web-based tutorials. Since the acquisition, classification, storage and retrieval of information are at the heart of library science, librarians are a very appropriate group to advise scholars on how to manage their personal information collections.

Conclusion

Due to the efforts of their libraries, scholars now have many information resources available at their desktop in electronic format. Scholars now have a wider range of options in how to deal with personal information resources than ever before. These include a paper

“reprint” article, journal and book collections with some form of cataloging and physical storage; using document scanning and OCR software to convert their existing print collections to electronic format; using bibliographic citation management software to catalog paper and/or electronic collections; using personal information management software to index and store electronic collections; or some combination of all of these methods. The degree to which scholars employ these various methods is a needed area of study by librarians.

Author Note

Mention or use of commercial software products does not imply official institutional endorsement of them.

REFERENCES

- askSam Systems. 2003. *Citation Bibliographic and Research Note Software*, 15 February 2003, <http://www.citationonline.net/> (accessed 20 September 2003).
- Balboa Software. 2003. *Library Master for Windows 2003*, <http://www.balboa-software.com/lmw.html> (accessed 16 September 2003).
- Brown, C. M. 1999. "Information Seeking Behavior of Scientists in the Electronic Information Age: Astronomers, Chemists, Mathematicians, and Physicists." *Journal of the American Society for Information Science* 50(10): 929-943.
- Bush, Vannevar. 1945. As We May Think. *Atlantic Monthly* 176:101-108.
- CatRunner LLC. 2003. *CatClip 2.5 - your personal information organization solution,2003*, <http://www.catrunner.com/> (accessed 27 September 2003).

CG Information. 2003. *Biblioscape - Bibliographic Software for Citations and Bibliography Generation in APA, MLA, etc. The Best Reference Manager*, 13 August 2003,

<http://www.biblioscape.com/> (accessed 16 September 2003).

CrazySquirrel Complete Solutions. 2003. *Refas - The Reference Assistant*, 2002,

<http://www.crazysquirrel.com/refas/refas.php> (accessed 16 September 2003).

Enfish Corp. 2003. *Enfish - Find Anything, Anywhere, Quickly*, 2003, <http://www.enfish.com/>

(accessed 10 September 2003).

Freebyte.com. 2003. *TreePad: Personal Information Manager, Notes Organizer, Word*

Processor, PIM, Database and More!, 2003, <http://www.treepad.com/> (accessed 27 September 2003).

Hallmark, Julie. 1994. Scientists' Access and Retrieval of References Cited in Their Recent

Journal Articles. *College & Research Libraries* 55 (3):199-208.

Highsmith, Anne L. 2002. [Personal Communications] amdb.enz EndNote connection file for

LibCat, Texas A&M University.

Hurd, Julie M., Deborah D. Blecic, and Rama Vishwanatham. 1999. Information Use by

Molecular Biologists: Implications for Library Collections and Services. *College & Research Libraries* 60 (1):31-43.

ISI ResearchSoft. 2003a. *EndNote -- Product Information*, 5 August 2003,

<http://www.endnote.com/eninfo.asp> (accessed 20 September 2003).

———. 2003b. *ProCite for Windows and Macintosh Product Information*, 19 August 2003,

<http://www.procite.com/pcinfo.asp> (accessed 20 September 2003).

———. 2003c. *Reference Manager for Windows Product Information*, 20 February 2003,

<http://www.refman.com/rminfo.asp> (accessed 20 September 2003).

- . 2003d. *WriteNote -- Product Information*, 21 February 2003,
<http://www.writenote.com/wninfo.asp> (accessed 20 September 2003).
- McGeachin, Robert B. 1998. The Feasibility of Using Document Management Software in the Scholar's Workstation. *Journal of Academic Librarianship* 24 (6):449-553.
- NoodleTools. 2003. *NoodleBib - The MLA and APA Bibliography Composer*, 2003,
<http://www.noodletools.com/noodlebib/index.php> (accessed 16 September 2003).
- RefWorks. 2003. *RefWorks Web Based Bibliographic Management Software*, 2003,
<http://www.refworks.com/productinfo.shtml> (accessed 16 September 2003).
- ScanSoft. 2003a. *ScanSoft - OmniPage Pro 12 Office*, 2003, <http://www.scansoft.com/omnipage/>
(accessed 20 September 2003).
- . 2003b. *ScanSoft - PaperPort Pro 9 Office*, 2003,
<http://www.scansoft.com/paperport/pro/> (accessed 20 September 2003).
- . 2003c. *ScanSoft - PDF Converter*, 2003, <http://www.scansoft.com/pdfconverter/>
(accessed 28 September 2003).
- Scholar's Aid Inc. 2003. *Is Scholar's Aid Right For You?*, 27 August 2003,
http://scholarsaid.com/right4you_research.html (accessed 20 September 2003).