The Scholar and His Library in the Computer Age

ABSTRACT

The advent of powerful high-speed computers and the development of networked information resources have freed the scholar from the limitations of his private library, but new problems have arisen. Lack of standardization in both hardware and software, reluctance on the part of many scholars to master the new technology and resources, and the overwhelming choices facing the adventurous modern scholar present barriers to optimal information retrieval. The library must help resolve many of these problems and must utilize the new technology to store, catalog, retrieve, and deliver information regardless of its format.

THE PRE-COMPUTER SCHOLAR

I am by trade a medievalist. This means that most of the information I use is in the form of the written word, stored for most of my career in books and codices found in libraries and archives. Thus, I have spent a good deal of my scholarly time and energy in and dealing with libraries. Like all of us, including the librarians among us, I have always found the libraries I worked in, from the smallest to the largest, to be inadequate and have cursed the various number systems used (see Lehnus 1980), the lack of analysis of serials, the inadequacy and inconsistency of subject headings, the physical location of collections, in short, all of those things we all complain about in our libraries. To overcome these inadequacies, I did all the things a scholar did in the pre-computer age. I slowly
acquired a large library. I bought a Mudge (1917-36), then a Winchell (1967), then a Sheehy (1976) (not to mention Totok and Weitzel [1984], Malclès [1950-58], Kaufman, et al. [1955]), perused them thoroughly, and filled them with marginal notes full of Library of Congress, Dewey Decimal, and Harris numbers. I assiduously looked through the various catalogs of the Library of Congress, the Bibliothèque Nationale, the British Museum, the New York Public, etc. etc. As time went on and things got better, I could depend on the various bibliographies listed in my trusty Besterman (1965-66). I ordered microfilms for myself and acquired such things as Keil's *Grammatici Latini* on microfilm. When microfiche came along, I was able to get Migne's *Patrologia Latina* on microfiche. Already in the dark ages before photocopying, I was able to photograph, to copy by holograph, and to use one of the various peel-aparts available then. Even before one could buy a personal Apeco or Xerox machine, I could send microfilm to Ann Arbor and obtain strip xeroxes by Copy-Flow process; for example, I have all of the manuscripts of Wolfram's *Parzival* photocopied and bound, so that I can overcome the problem of sequential searching, which makes microfilm so hard to use for manuscript comparison. I even have several volumes of articles and excerpts I typed in by hand. I was always trying to lessen the burden of library work, figuring out how to use edge-punched cards, how to make sure I always had a supply of three-by-five cards and a pen at all times in the library.

Fonts were a particular problem for the scholar before 1980. One had to learn to use a quill or one of the metal nibs available commercially (e.g., Speedball) to imitate the hands of the manuscripts (as I did in Marchand 1969). Technology afforded the possibility of typing scripts with the Vari-Typer and later with the IBM Selectric, both with replaceable elements. One could have one's typewriter modified by having font faces soldered on, and there were even people like William Bennett, who had rubber stamps made, or Norman Willie, who was reported to have had a Gothic typewriter made. One could get kits that enabled one to replace keys (Type-It). With the advent of the mainframe, the chain printer afforded the possibility of making some fonts, thus creating characters such as thorn and edh.

For large bibliographical projects, one could make use of edge-punched cards or even of Hollerith cards sorted by machine (see Reichmann 1961). One could shingle old note cards and photocopy them off, making bibliographies sorted as to author, date, subject.

Enough of this litany on the scholar of yesteryear. I was still bound to a great extent to an individual library, to indexes of books, such as the impossible-to-use-but-indispensable indexes to Migne, to an occasional trip to another library. The advent of the high-speed computer
with large storage capacity has changed all that, especially the advent
in the 1980s of the desktop (then laptop, notebook, hand-held) with
hard disk, floppies, large amounts of RAM, incredible input (scanner,
voice input) and output (CGA, then EGA, VGA, SVGA; voice; laser
printers; color printers) devices, and storage and retrieval devices such
as CD-ROM. The present-day scholar is freed from the tyranny of the
mainframe, has available to him at his desk or under his tree riches
which no Croesus of yesterday could have afforded; we have paradise
enow at our fingertips. What is wrong?

As I talk to colleagues around the world, a sense of frustration
seems to be prevalent. We know that there are untold riches and programs
that would solve our problem, but we do not know how to find or
use them. Without waxing too philosophical, I remember an old paper
of Heinz von Foerster's in which he pointed out that the metaphors
we use in speaking of computers are often awry (von Foerster 1970).
The network is not an information superhighway, it is an ocean of
pathless pieces of information, a self-organizing system where no one
knows his way, and the scholar is awash in this sea of information.
All one needs to do is to look at the messages on any list, to talk to
any colleague, to look at any of the FAQs available. We need help,
and the rest of this paper concerns some suggestions I have. Help will
have to come from one of the players in the game. Let us look at them.

THE COMPUTER INDUSTRY

The computer industry itself, and this includes manufacturers of
hardware, software, etc., needs to think about the user. Each and every
interface has to be learned, and protocols differ widely. If one just asks,
for example, how to escape from a program, one will notice that there
is a bewildering number of ways. No one uses <Escape>, <Break>,
or <Ctrl-C>, as recommended in the programmers' manuals. We find
<F1> (WordPerfect), <x> (List), <Ctrl-Z> (FancyFont), <Q> (nn),
<q> (XT-Gold), <bye> (FTP), etc. etc. Dana Noonan has complained
a number of times concerning the lack of any break key at all, for example
when using File Transfer Protocol (FTP).

The lack of standardization in such a small matter as an escape
key is symptomatic for the whole industry. We have several platforms—
DOS, OS2, Windows, Macintosh, UNIX—each incompatible with the
other. One of the greatest problems in an academic office occurs when
one of the parts uses a Mac and the other uses a DOS-based machine,
and everyone knows the problem of receiving a PostScript document
when he has only an HP printer, not to mention such things as Rich
Text Format. The industry needs to standardize, but it also must consider
the needs of the users more. Recently, we have seen a proliferation of tsr and high-memory programs, programs that require increasing amounts of RAM and memory handlers, and the industry has been slow to handle the situation, so that one receives daily calls from colleagues who are finding that they cannot install this or that program, e.g., the CETEDOC CD-ROM. Those who feel that they cannot do without graphical user interfaces and use Windows, often find their machines locked up, always for no reason. Of course, we do not want to go back to 640K, a 6.7Mhz CPU, and a 20M hard disk, but we do need some consideration.

THE MODERN SCHOLAR

The poorest player in the game is the scholar himself, and the Lord helps those who help themselves. There are many closet Luddites lurking in our midst, as became painfully clear in a recent New Yorker bite (Baker 1994). Many scholars seemingly just do not want to give themselves the time to learn about the computer, and this seems only reasonable, if they would just leave it at that. Everyone, however, is in search of a guru, and woe unto you if you get to be known as one; the guru's life is filled with nocturnal and Sunday afternoon phone calls. His knowledge of the field is at best haphazard, frequently erroneous, and always difficult to convey. The fact that yesterday's tools were plug-and-play, right-out-of-the-box, a-child-could-set-it-up devices has led to misunderstanding and mistrust of the computer, where you need an interface and fine tuning for almost everything. Even if, horresco referens, the scholar uses an electric typewriter, one just plugs it in and it runs.

If we return to the picture of the pre-computer scholar drawn in the first part of this paper, we will notice that everything has changed for him. Many of his research tools are now available electronically, some online, some on CD-ROM, and the prices are coming down. A scholar with a large RAM and Windows, for example, can purchase for less that $20.00 a CD-ROM called Languages of the World and can have available on a hot-key a dictionary of twelve languages to be used, for example, within his word processor (see Marchand 1994). Large dictionaries, such as the Random House Unabridged, not to mention the expensive Oxford English Dictionary, are available (our mainframe, as do many, has this one online, with a poor interface, but not to complain). For those who have the money, all the Wilson indexes (e.g., Art Index, MLA) are available on CD-ROM, as are the Bowker indexes (e.g., Ulrich, Books in Print). You can have the Hebrew, Greek, German (Elberfelder), French, Spanish, and Dutch Bibles for
less than $20.00, with a search engine and display on the proper screen in the original scripts (Straight from Heaven, n.d.).

The modern scholar no longer needs his trusty three-by-five cards. When he goes to the library, he takes his notebook computer and a battery-driven scanner. The results, though not always perfect, are better than my handwritten three-by-fives. If he has larger texts to copy, he can take them to his office and scan them using an optical character recognition (OCR) program. I scarcely ever use a pen or pencil anymore.

The modern scholar has little need to travel to other collections to see what they have or to pore through catalogs. He can get online and search through multiple online public access catalogs (OPACs) to find what is available. I used to copy off tables of contents, and there used to be journals, for example, Leuwense Bijdragen and Lychnos, that published tables of contents of journals, for those who worked at less-favored libraries to use. With the new CARL UnCover service, I can have sent to me by electronic mail the tables of contents of whatever journals I customarily use and, as usual, get a fax of one if I need to (for a fee). Even at my own library, I can download lists of books on a particular subject, make myself a shelf list, etc., without leaving my desk.

The modern scholar is at no loss for fonts. Even in the days before vector fonts, I was able to clip fonts from manuscripts, generate printed fonts, and place all kinds of fonts on my screen (for examples, see Marchand 1987). Nowadays, I can create or borrow fonts in TrueType format, have them appear on my screen in any Windows application, and print on my trusty old HP Series II. If I do not have a character, I can modify one, borrow one, or create one.

The index problem has been solved. If I want to find a passage in Shakespeare, for example, I do not go to my Bartlett’s Concordance, look for the multiple occurrences, then go to my shelf, then copy off the passage. I simply interrogate my CD-ROM, using Findtext, call it up with List, and copy it into my document. If I am ever able to afford the new Patrologia Latina on CD-ROM, I hope the search engine is fast enough for me to search it instead of using the indexes I cursed a moment ago.

I spoke of acquiring a large library, especially of reference works. As time goes on, I am able to get rid of many of these, having acquired electronic texts either from the Internet or on CD-ROM. The value of CD-ROMs is that they are (1) small, (2) durable, and (3) easily searched. With the advent of the new SCSI CD-ROM readers, I can now actually go out under a tree and take a library of works with me. My scholarly work is made immensely easier by the computer, and I have not even begun to name the ways. It could be easier still, however, and that is where the last player in the game comes in.
THE LIBRARY

We have been accustomed to looking to the library for help in the storage and retrieval, cataloging, and delivery of information. Using the term library in the widest sense, I think we must do so now. The dysfunction between the library and the scholar is proverbial. One of my favorite stories to illustrate this is ancient, but it was told to me by a librarian, David Kaser: The Harvard University librarian about 1900 encounters a friend on the street. The friend asks: "How are things?" "Fine," says the librarian, "There are only two books out of the library and I am out after one of those right now."12 The computer age has brought us tools which increase the dysfunction (such as the NOTIS system and the "problem patron") and tools which can aid in overcoming it. In this section, I shall seek to identify trouble spots and to make suggestions as to what can be done about them; however, given space constraints, I can only name some of the most acute.

How shall we store our information? Shall we use plain ASCII text or shall we use one of the markup languages recommended? If markup, which one and how much? Note that this decision will determine to a great extent how we will retrieve our information. Note that any intervention on the part of an editor will inevitably distort the information and create noise in the system, which may be difficult to remove. For many texts, plain ASCII will be the first choice.

As just pointed out, the packaging of information contaminates it; we just need to choose that method which distorts it the least. The indexing of information or the use of subject headings, etc., will also inevitably form a barrier between the user and the information; even concordancing will do this. Perhaps, again, the best solution at first will be raw ASCII for texts, with a look-up engine that will find strings. Where codicology is important, as in the study of medieval texts, we will have to resort to other means.

In the handling of visual information, we are just beginning. Of course, nothing can replace the original, be it runestone or codex, but remarkable advances are being made (see Marchand 1992a, 1992b, 1992c).13 However images are to be accessed by the patron, the library should store them in as rich a format as possible, say TIFF, though this may require an enormous amount of storage. Libraries and archives ought to consider three-dimensional representation and reconstitution of objects, for example, by sintering laser. All the manuscripts in the world ought to be stored in the best digitized manner possible, so that their information is at least partially registered. If images are to be manipulated in any manner, even algorithmically, a TIFF file of the first "photograph" ought to be kept. I cannot go into methods here, but it seems to me that the digital camera has advanced far enough
to permit real-time registration and filtration and thus better pictures than have been made so far. It should be obvious also that what has been said up to now applies to audio as well.

Even the best of libraries is of little use if the patrons cannot retrieve the information stored. One of the greatest problems of retrieving electronic texts is knowledge of their availability and accessibility. There are great treasure troves of electronic texts out there whose existence is known to me, but which I cannot access, for example, the Rutgers collection and the Kiel collection. It is to be hoped that projects such as Project Gutenberg, Lysator, and the Online Book Initiative (OBI) will continue to provide access to important electronic texts. God bless the Oxford Text Archive; would that others would follow its lead.

The cataloging of electronic texts is of great importance, and we ought all to have an input into the methods of doing so. Before we can catalog them, we need to know of their existence. I patrol World Wide Web and Gopher-space continuously on the lookout for such texts; it is both disheartening and exhilarating to run into a repository of texts of whose existence you were not aware. The Georgetown initiative, the Center for Electronic Texts in the Humanities (CETH), and lists such as ANSAX-L and MEDTEXTL are grand resources, but we need a more concerted effort. When I see lists of available electronic texts such as that of Wiretap, I am really disheartened.

As I look through OPACs, even using HYTELNET, I find the fact that there is no uniform interface for OPACs an almost intolerable burden. I use OPACs to discover works, to confirm citations, and to make bibliographies. The Committee on Institutional Cooperation (CIC) is to be praised for its efforts in the direction of a uniform interface for its member institutions, but this is only a drop in the bucket. As for capturing a session, I have found that I have to have recourse to a tsr program even in my own library.

Once we have identified sources for electronic texts, what shall we do with them? There is no need, for example, for each library to mirror Gutenberg, OBI, Lysator, and such, since these are available on the net, but each library must inform its patrons of their existence and provide an interface for interrogating and/or downloading them. Here, of course, we must also depend on the scholar to know something about the texts involved, such as the fact that "The Awful German Language," by Mark Twain, is found as an appendix to A Tramp Abroad, so that one does not have to search for it, or that Schiller's An die Freude, found in many different places on the net, is incomplete, lacks umlauts, and is frequently garbled.

There are several problems involving CD-ROMs (Budd and Williams 1993). The first is their cataloging, and I know of no source
that lists all CD-ROMs, much less one that evaluates them properly. In order to find which electronic texts are available on CD-ROM, we need to analyze each one, if we can find out about its existence. For example, it is of some importance to the patron to know that *Desktop Library CDROM*, 1st Edition, August, 1992 (Walnut Creek), *DeskTop Bookshop*, (Unica Ltd.), *Reader's Library* (Micro-Mart Computer), and *Library of the Future Series*, second edition (World Library), contain almost entirely the same texts, but this is mentioned nowhere.

One of the most vexing problems for the user of CD-ROMs is the up-front software. That provided with the *St. Thomas Aquinas CD-ROM*, for example, is not of any great use for my purposes. Some CD-ROMs, for example, CETEDOC, are hard to install, and recently the list MEDIBER witnessed a rather sharp exchange between a user of *Admyte*, who could not get it to work, and one of its authors. Providers of CD-ROMs should adopt either a uniform interface or no interface and should take into consideration such things as conflict in memory space. The Apple/PC conflict frequently extends also to CD-ROMs.

Now I come to what is probably the most important point I wish to make, and this concerns the mounting of CD-ROMs and other databases for remote interrogation. As an individual scholar, I own the *OED2, Thomas Aquinas*, CETEDOC, several Wilsondiscs, many, many electronic texts, *Languages of the World, ICAME, Computer Select*, etc. etc. on CD-ROM, so that I can use them sitting at my desk. Some of these my library doesn't even own. This is not right; it is obvious that one cannot expect each scholar to provide himself with such tools. The library should mount for remote access all of the above and more. We have jukeboxes and towers that will hold as many as 100 CD-ROMs, and many come with the software necessary to network them for remote access (Breeding 1994). You should be able to sit at your desk and access the *MLA Bibliography, Books in Print*, the Verzeichnis Lieferbarer Bucher, FRANCIS, and Livres disponibles, etc. without having to get up. Both CETEDOC and Migne ought to be available in the same manner. I do not know what to do about the Greek texts of the Thesaurus Linguae Graecae (TLG) or the Latin texts of the Packard Humanities Institute (PHI), which must at present be rented, but we have to make a start. Nor do I know how we can persuade those holding private collections to release them.

These are just some of the problems we face. We need to agree upon some sort of uniformity: (1) of platform, (2) of software, (3) of entry (how about MARC?) (see Caplan 1993), (4) of bibliographical entry (there are over 100 formats at present in use, not to mention the idiosyncracies of individual libraries) (see Howell 1983, Stigleman 1991), (5) of basic elements (ASCII, extended ASCII), (6) of fonts.
At present, the individual scholar who wishes to make use of the tremendous possibilities the computer offers him must collect his own base of CD-ROMs, electronic texts, bibliographical software, presentation software and hardware, font software, and OCR software. All of this is managed at present at most universities by a system of unorganized gurus. It ought to be done by the library.

Here is the crux of the matter; we cannot expect our overworked librarians to do this alone, and I do not mean this in a patronizing manner (I have two librarians in my close family, so that I am not inclined to denigrate librarians). Every library needs a computer resource person, someone who can peruse Gleason Sackman, Yanoff, Noonan, December, Current—Cites, etc. and keep up to date, who can install programs and support them, who can show the occasional faculty member how to use a scanner and OCR, in short, a guru. It will be necessary to train such people, for I do not know of the existence of one at present.

I am painfully aware that I have only scratched the surface; there is much to be done, and it would be well if we got at it. Se non è ben trovato, è vero.

NOTES

1 Much of this talk is personal narrative and concerns the library I use the most, that of the University of Illinois at Urbana-Champaign. I feel that much of what I say is generalizable and applies, mutatis mutandis, to all of us. What I have to say is too important to worry about the conventions of polite conversation and eschew the first person.

2 I cannot resist quoting here a snippet I used to give to my students in “Bibliography and Methods” (Marchand 1955, vii):

ON OWNING BOOKS. “The old-time scholar accumulated his house full of books on a smaller income than that of today’s young man. Books were important to him; they were the fabric of his life; he did without other things; he wore his coat a year longer and carried home 20 new volumes. Today’s scholar will never be able to afford books, no matter what his income or his wife’s income, until he feels that books are at least as important to him as table cloths, toothbrushes, cat food, rugs, whiskey, hats, newspapers, cameras, and all the odds and ends that now come higher on the list. . . . Let us have no nonsense about the library taking the place of books in the home or about the great number of volumes published every year. . . . There is a deep personal and psychological difference between owning a library and using someone else’s. It is a little like the difference between owning and renting a house, between belonging somewhere and merely passing through. It might even be the difference between a scientist and a technician or between a scholar and a scholarly technician. . . . We might remember what George Savile, Marquess of Halifax, wrote in 1690 or thereabouts: ‘The struggle for knowledge hath a pleasure in it like that of wrestling with a fine woman.’ If the scholar or the reader finds pleasure, and not merely duty, in the struggle with learning, then he will want to live with it in his own house and not merely to sample it in the library.” August Frugé said it, in the Saturday Review for 16 July 1955.
For an example of an early use of the print chain, see the cover of *Literary Data Processing Conference Proceedings*, (Bessinger, Parrish, and Arader 1964) and the explanation on the title page.

FAQs, lists of Frequently Asked Questions, are archived by rtfm.mit.edu, in the pub/usenet/news.answers subdirectory. One of the best ways to find out about a subject is to download and read the FAQ on it.

For example: *Network-News*, no. 12 (November, 1993), p. 5: "Wishlist for the user: 1. A break key. How many times have you waited for the system to tell you it couldn't locate anything when you already saw the typo?" Amen.

There are, of course, bridge programs. One can use MacinDos to read and write to some Mac disks, and the Macintosh people have tried to make the Mac operate like a PC on occasions. GhostScript is an excellent add-on for reading and printing PostScript files, but these are add-ons, and they just add to the confusion.

This note, featuring an attack on electronic cardfiles and a *plaidoyer* for a return to the card catalog, caused a flurry of remarks on HUMANIST and BI-L, the list devoted to bibliographic instruction.

With most notebooks and Artec's WalkScan/256S, for example, one can have under six pounds of equipment, and one does not need an extension cord or an outlet.

See *InterNIC net-happenings*, April 6, 1994. Carl UnCover intends to provide the table-of-contents service free.

For information on TrueType fonts, specs, etc.: FTP ftp.microsoft.com in the subdirectory /developer/drg.

Findtext is a grep-type program created by Jeff Proise. A copy may be obtained from the disk included with his book, *PC-Magazine DOS 6 Techniques & Utilities* (Proise 1993), or from *PC-Magazine*’s bbs. List is by Vernon Buerg. It can be obtained from most bulletin boards.

For a somewhat harsher view of this dysfunction, see Smith 1990, tempered somewhat by the rebuttal of Phyllis Franklin (1993). The preservation of primary materials and the threat to their existence by imaging technology is another story (see Tanselle 1993).

Also posted for FTP in the following groups: HUMANIST, IOUDAIOS, and RELIGION.

For example, the texts by John of Trevisa deposited at FTP: blackbox.hacc.washington.edu, subdirectory /pub/text/trevisa, including the Gospel of Nicodemus, the Defensio Curatorum, De Regimine Principum, and Polychronicon. Neither I nor any of the 500 members of MEDTEXTL were aware of their existence, and some of us were actively engaged in work on Trevisa. Occasionally, it happens that one finds a trove and forgets the address. There are a number of works in Slavic languages deposited on a server somewhere in California.

The *Catalog of Available Online Books* scarcely scratches the surface of what is available, though I suppose we should not complain about such a pioneering effort.

Gleason Sackman is the editor of InterNIC net-happenings, an excellent newsletter for keeping up with what is going on the network, including E-D-U-P-A-G-E, ALAWON, EDUCOM, The Internet Hunt, and many others: net-happenings@is.internic.net; subscribe at listserv@is.internic.net.

Scott Yanoff, *Special Internet Connections*. Updated every week or so, this short list (usually five or six pages) provides access information and brief notes on about 100 popular, new, or interesting network resources. To subscribe, contact Scott Yanoff at yanoff@csd4.csd.uwm.edu. You can also retrieve past issues by FTP: csd4.csd.uwm.edu, /pub/inet-services.

Dana Noonan’s Nnews, available from ftp.nodak.edu, subdirectory /nnews. There are a number of files, with various dates. I would get them all.

Another excellent keeping-up source is from John December, known as the December lists: ftp.rpi.edu, /pub/communications/internet-cmc.

Current—Cites, available from ftp.lib.berkeley.edu, in the subdirectory /pub/Current Cites; there are a number of files, well-labeled. This is a very good source, leaning towards library, with summaries of articles from a number of journals. Also available in *InterNIC-Happenings*. 
REFERENCES


*Catalog of Available Online Books* [Online]. 1993. Available: gopher@wiretap.spies.com


*Straight from Heaven*. no date. [Computer program]. Lakewood, Ohio: Most Significant Bits.


