European Journal of Surgical Oncology 1998; **24**: 134–143

EDUCATIONAL SECTION

Scientific information sourcing in the electronic age

David A. Rew and Humphrey Dunn*

University Surgical Unit, University of Leicester, Leicester, UK and
**Medical Librarian, Glenfield Hospital, Leicester, UK*

Introduction

Information is time and time is a valuable resource. Clinicians face a perennial challenge in managing data overload. We need scientific information and data to pursue and to enhance our professional practice. This information may be general, in the form of press articles, television, video and radio programmes, or specific, in the form of scientific and review articles in the specialist press. Continuing Medical Education (CME) encourages us to learn about and to implement significant and beneficial changes in our specialist practices. Our patients expect us to be current with topical media comment in our subspecialities.

There has been an uncontrolled proliferation in scientific journals, books, newsletters and magazines; in conference reports, programmes and abstracts; in distributed electronic media such as compact discs (CDs) and tapes; and in the professional resources of the Internet. This provides us with the problem of sorting through a mass of information to extract the valuable, the relevant and the consequential from the noise. We also need to recognize that which is beneficial and of real utility from that which is frivolous among the raft of new technologies at our disposal.

Media reports would have us believe that in the brave new world of computers and digital data, everything we need to know will be at our keyboard fingertips. This is an oversimplification. We are in a period of transition, of uncertainty and of rapidly increasing complexity for the knowledge-based media. It is the purpose of this article to consider the current problems, developments and opportunities in forms of information management and retrieval, and in particular the significance of electronic data management in relation to published scientific material.

Correspondence to: D. A. Rew, Senior Lecturer in Surgery, University Surgical Unit, Glenfield Hospital, Groby Road, Leicester LE3 9QP, UK.

0748–7983/98/020134 + 10 $12.00/0

The process and psychology of data assimilation

The way in which the human eye, ear and brain assimilate information is the limiting factor in the design and utility of data retrieval systems. Vision is the most important conduit for information, whether from the printed page, the microfile reader, or the computer screen. The brain can assimilate information very rapidly by visual scanning. There are two general forms of data assimilation during reading. Through browsing, scanning or speed-reading, large volumes of data, such as the contents of a journal, magazine or book can be sifted. During focused reading, the reader pays greater attention to the detail of the document. Auditory cues can be sufficient for blind individuals to use the Internet.[1]

The greater the mass of information material in the professional environment, the more effective must be the process of scanning and selection. In the early part of the 20th century, when information was scarce, slow and expensive, a single newspaper and a general review journal might have satisfied our individual information needs. In an era of exponential growth in all media, even the choice of medium to consult becomes a time-consuming exercise. An example from television illustrates this. If we are to receive 50 or even 500 channels of television in new digital broadcasting format, then the time taken to select a programme will become a limiting factor. Similarly, in a world where there are 100 or so surgical journals alone, how does the specialist select the best sources of information? This problem is enormously compounded on the Internet, where the quantity of unrefereed material exceeds by many times the quantity of peer-reviewed material.

Paper-based media

Printed material will have continuing utility. The printed page remains the best way to present large quantities of information to the eye. It is a tried and tested medium for longitudinal data storage. Paper-based records have survived as the medium of archiving for hundreds of years.

Nevertheless, paper has a number of disadvantages. The information in books and journals is only current at the time of going to print, with the risk of a rapid depreciation in its utility. Books and journals more than a few years old are rarely consulted, and then only very selectively. Much information on store on paper in a medical library is thus a testimony to its historic collectability rather than to its current utility. This is unfortunate, because much medical knowledge in older books and journals has continuing utility, be it in the case report, the clinical trial, or the once common but now defunct diseases such as smallpox. Most documents disappear into store as shelves fill up with time, removing their knowledge content from effective circulation. Thus, medical knowledge becomes less an accumulation than a rolling programme, wherein old lessons can be forgotten and wheels reinvented. A retrieval medium which gave a 50-year-old article instant presence and prominence in a search strategy would be a considerable advance.

Another disadvantage of paper is that its information content must be copied and transcribed mechanically or manually to be of use. Optical character recognition (OCR) software programmes address this problem, by converting printed text into digitized characters suitable for use in a word processor. However, such systems are not yet sufficiently reliable or widespread as to make electronic assimilation of the contents of the printed page a simple or cost-effective task.

Paper also imposes significant costs. It occupies physical space for display, storage and archiving. The cost to libraries of providing a wide range of journals is becoming prohibitive. The proliferation of journals and associated costs makes it increasingly unlikely that new and low circulation titles will be acquired by any other than principal reference libraries. They are thus unlikely to reach a broad audience by conventional means. Libraries are thus among the powerful advocates of electronic data archiving and retrieval, and library investment is increasingly towards electronic infrastructure. This risks a cycle of decline for many specialist printed publications.

There are forms of image storage other than paper. These include microfiche, which reduces the space required for storage and speeds up retrieval. Computers also allow image storage, whereby the image of a page but not its content can be stored and manipulated. However, at a time of rapid change in technology and industry standards in electronics, we cannot be sure that the technologies will survive to allow the reading of current storage media. While paper has a finite life except in the most carefully controlled conditions, we cannot be certain that modern media, such as CDs, tape and disk drives will not decay or suffer irreversible damage and decay with time. Such, for example, was the fate of many early films on celluloid.

The utility of electronic medical data

Electronic, digital document data offer immense theoretical advantages to users and readers. The data are transportable to many media, including the computer screen, the printed page, and the archive database, and can be incorporated into other documents and files. They are interactive, allowing assimilation, conversion and modification in many different ways. They are flexible and transmissible, such that data can now be moved in large quantities, virtually instantaneously, to any point on the globe at very low cost. The reader does not need to be collocated with the original document, as in a library, or to occupy time and space travelling to and from the reference point. The speed of access and search of large databases is rapid, and large volumes of data are archivable in limited space, such as CD-ROM (Compact Disc Read Only Memory), or tape and disk drives. A conventional CD can hold around 600 million bits of binary data per disk, where an A4 page of text comprises around 2000 bits of data.

The electronic format has allowed new types of data. The CD-ROM index of published work, including abstracts of the document, has superseded the *Index Medicus* and other paper-based archives as the source of choice for information about recent scientific papers (1966 onwards). Examples of electronic indices include Medline, supplied commercially by organizations such as Ovid and Silver Platter. These systems are based upon public archives provided by the US National Library of Medicine. Medline is also available free "online" across the Internet using World Wide Web browsers. Information on obtaining free access to Medline is provided at a variety of sites.[W5, W6] Public access Medline providers include Medscape,[W7] Internet Grateful Med,[W8] PubMed,[W9] Healthgate Medline[W10] and Docnet.[W11] In the UK, such database resources include BIDS (Bath information and data services).[W12, W13]

Another form of electronic textual material is that contained within Internet Web sites. The World Wide Web format, whereby browser software and communications protocols such as HTML links are standardized, must be distinguished from Web site content, which is under the local control of the data provider, be it an individual, a commercial or a public organization. Unlike the formal journal publishing process, Web site publishing does not demand peer review or formal vetting. It is thus a useful medium for the publication of original or unconventional material which does not fit readily into existing journals. This includes previously published work, and the text of materials such as books which may be of particular scientific interest as, for example, historical scientific works or the full text of a doctoral thesis.

The World Wide Web provides a much larger and more eclectic source of medical data and text than is formally catalogued in such professionally structured database systems. A number of on line general purpose "search engines" such as Yahoo, Alta Vista or Excite, can be used to find this unstructured material. These retrieve lists of documents and web sites which contain the words and phrases specified by the user. Current awareness services[2, W1–W4] are another example of an electronic search service. Using one of a number of organizations, a user may specify information to be obtained from the Web, such that personalized newspapers, professionally relevant news services, or recent research findings will be notified automatically.[W14–W18] Certain search engines are tailored to medical use.[W19–W22]

Advanced electronic data sets

Electronic textual data can be interlinked with other media, including sound (audio) and vision (video or still pictures), creating interactive multimedia presentations for CD or Internet distribution. Multimedia systems can provide a powerful vehicle for the distribution of educational sound, picture and video material, as for example with human anatomy.[W23, W24]

Yet another form of electronic data of unique value may be described as scientific interactive data. Many branches of science, such as pharmacology, biochemistry, genetics,[W25] particle physics, astronomy and geology, produce vast data sets for distribution, comparison and online interrogation, including three-dimensional and virtual reality analysis. Such sets include gene and protein sequences, and three-dimensional chemical structures.

Data provider organizations

The Internet and the World Wide Web have provided a particular stimulus to public and private organizations, and primarily museums and libraries, to convert their knowledge base into digital form.[3] Many major medical and scientific libraries, museums and reference collections are now being digitized. Examples include the Library of Congress[W26] and the British Library.[W27] The skills of librarians and archivists themselves find a natural synergy with the Internet in electronic data management. Networking of institutional holdings creates a greater searchable resource. For example, the many manual catalogues of the British Library cover more than two centuries of documents. These are now being integrated into one Online Public Access Catalogue (OPAC97) services for universal access from remote terminals.[W28]

Libraries, museums, galleries and collections benefit from electronic conversion in a number of ways. It allows the publication of the electronic image and allows access to rare documents and manuscripts without the risk of damage in handling. It makes items available to a much wider audience. Electronic distribution reduces the risk of damage to original documents through human contact. It allows the electronic distribution of material. This reduces the need to visit the institution in person, which in turn reduces the demands and costs for provision of public access and reading space. It also allows the remote and low cost storage of original documents. It avoids the need to loan out or to transfer material between institutions, thus keeping collections intact, and reducing loss and damage.

Collections of images, including art, photographs, drawings and film can similarly be made available without the need to handle the originals. In the case of museums and galleries, catalogues and images of artefacts can be made available to a wide audience. Thus, for example, the British National Science Museum and the Wellcome Institute[W30] have made available important collections of medical artefacts. The electronic version can be incorporated into an educational, multimedia format which enhances the value of the original. There are now many such examples on the Web.[W31–W33] Moreover, many artefacts are excluded from display by virtue of bulk, constraints of space, inadequate local storage or fragility. Electronic formatting brings these hidden objects to the front of house of the electronic museum.

Tools for electronic data access

How can the non-expert and occasional user take advantage of these data resources? For those without easy access to the Internet, many hospital libraries now provide access to CD-ROM archives (such as Ovid Medline) on a stand-alone personal computer. These software systems allow the downloading of references as a printout or for direct transfer into a reference management software package such as Endnote or Reference Manager.

Users with an internet connection can undertake such searches online with a standard Web browser such as Netscape Navigator or Microsoft Internet Explorer. Some services, such as BIDS, require prearranged codeword authorization, while others are currently free at the point of use.

Selected data and text can be downloaded from the Internet by following simple instructions, either to be printed on the local printer, or to be stored locally in an electronic file for further use. Most usefully, many documents can be printed out using standard software to a quality equivalent to that of professionally printed text, thus bringing to reality the concept of the journal held in electronic form centrally, but distributed at minimal cost to the point of use in publication format. Various enhancements, or 'Plugins' to the browser software, allow Audio, 3D, Virtual Reality and other interactive features to be accessed.

Disadvantages of the electronic medium

With these apparent advantages of electronic data and the Internet, we might reasonably ask why and whether paper media will survive at all. At the present time, the advantages of digitization are offset by considerable inconveniences with Internet use. Firstly, the would-be electronic reader needs to establish a connection with the Web through an account and registration with an Internet provider, which may be a private supplier or a public body such as a University. The exponential growth in users and web site providers continues to outstrip the rate of expansion of the Web infrastructure and the capacity of the cable connections in many parts of the Net. Thus, the apparent gains from electronic transmission are offset by considerable inertia.

Secondly, personal computers, even where highly specified for speed .and memory, with clever software browsers, remain sluggish to use in practice, with a number of keyboard and mouse iterations to be performed in loading and unloading relevant software. Computer monitors lack the resolution, image capacity per page, the page-turning capabilities, the tactile features and the portability of books and journals. Portable computers, high resolution screens and larger video monitors can help reduce the utility deficit, but not completely.

Thirdly, public search engines are insufficiently precise in

their search strategies for many requests, and are defeated by their own massive archiving capabilities.[3] The resources on the Web are growing faster than the ability of current systems to organize and catalogue them automatically. Thus, a seemingly simple search on a set of terms such as 'Scientific, Hypothesis, Medicine' would recently have yielded some 70,000 hits on an engine such as Alta Vista, creating an impractical perusal task for the inquirer. Allied to this is the problem that some search terms fit a number of classifications. The inquirer may then have to follow a number of leads down a number of avenues, at which point the World Wide Web can sometimes resemble a series of blind-ending tunnels rather than a Net.

New forms of search structure are needed to represent data and to aid selection of material relevant to the search within a reasonable time.[4] These include new forms of cluster analysis, graphical representation of the word content of documents, and three-dimensional models such as rooms and shelves on which to organize retrieval strategies.[5]

New rules may also be necessary to organize the Web. At present, anybody with a suitable computer and a Web connection can publish material, leaving classification to others. There is a clear need for all web documents to contain intrinsic classification codes, as with key words on paper documents, allowing archiving systems to register and to classify the content and the author's intentions much more clearly.

The library of the future

We may expect to see considerable changes in the way in which we access scientific information.[6] The electronic search and the Internet search will become commonplace. The medical library will become a computer-orientated institution. Even small libraries will command instant access to all of the major medical databases and information sources. Their utility will not be constrained by physical holdings of books and journals and librarianship will be redefined by computer and Internet literacy. Libraries will become repositories of expertise on electronic media search, archiving, access and storage systems. Small libraries in district hospitals will be as powerful as those in the largest university hospitals. The economics and cost base of electronic libraries will be transformed. More journals will become available in electronic form, with the facility to print out copyrighted material at the point of use. Contracts with publishers will obtain access to electronic image and text versions of journals, such as are currently being introduced. For individual users, the personal computer connected to the Internet has already become the library on the desktop.

Conclusions

We are now in a period of transition. The infrastructure of the Internet cannot yet deliver sufficient speed and convenience to many users. Many readers may still be unfamiliar with the technology and opportunities of the Web, and may not have acquired the skills and confidence to use computers fluently.

The software and search engines of the World Wide Web are not yet entirely adequate to the task of structured and super-selective searches. Paper-based media remain important for archiving purposes and readability, but the balance is likely to change. Computers currently find their greatest utility in cataloguing rather than in archiving the entire texts of documents, for which paper retains superiority of convenience, of presentation, and of data capacity for complex images.

We recognize the important and potentially beneficial implications of developments on the Internet for the *European Journal of Surgical Oncology*.[7] Electronic dissemination has the ability to make the content of limited circulation specialist journals such as this one much more widely available, once copyright and income issues are resolved. This is likely to be in the form of group licensing arrangements. The electronic formats will run in tandem with the printed version, which will remain the document of record. Publishers are beginning to experiment with this format.[W35, W36]

In the present century, the telephone has become a universal tool of communication. The Internet now does for the written word and for the image what the telephone has done for the voice. We may expect many technological developments to improve the speed, the capacity and the effectiveness of the networked electronic medium. Nevertheless, the information revolution is already upon us, and it has profound implications for us all.

References

1. Raman TV. Websurfing without a monitor. *Scient Am* 1997; **276:** 65.
2. Kiley R. Current awareness services on the Internet. *J Roy Soc Med* 1997; **90:** 540–2.
3. Lynch C. Searching the Internet. *Scient Am* 1997; **276:** 44–8.
4. Anon. The Internet: bringing order from chaos. *Scient Am* 1997; **276:** 41–73.
5. Hearst MA. Interfaces for searching the Web. *Scient Am* 1997; **276:** 60–4.
6. Lesk M. Digital libraries. *Scient Am* 1997; **276:** 49–52.
7. Rew DA. http://www.hbuk.co.uk—the *European Journal of Surgical Oncology* and the Internet. *Eur J Surg Oncol* 1996; **22:** 563–8.

Relevant Web site references (current, November 1997)

(Prefixed by http://www. except where indicated by http://...)
(Bracketed text indicates title or role of site.)

Advice on using Internet search engines

W1. hamline.edu/library/bush/handouts/slahandout.htm
W2. http://scout.cs.wisc.edu/scout/toolkit (Scout Toolkit)
W3. searchenginewatch.com (Search Engine Watch)
W4. monash.com/spidap.html (The Spiders Apprentice guide to Search Engines)

Advice on Medline searches

W5. http:/surf.to/glenfield (Our own library guide to Web Resources)
W6. grhlib.demon.co.uk/medtut.html (Dr Felix's MEDLINE Tutorials page)

Free access to Medline on the Web

W7. Medscape.com (online access to Medline, and a medical dictionary)
W8. http://igm-01.nlm.nih.gov/index.html (Internet Grateful Med)
W9. http://www3.ncbi.nlm.nih.gov/PubMed/ (PubMed, including access to Genbank DNA and protein sequences, genomes and 3D biomolecular structures)
W10. healthgate.com/HealthGate/MEDLINE/b.search.shtml (Healthgate Medline)
W11. docnet.org.uk/drfelix (Dr Felix's MEDLINE page)

Online database search facilities

W12. bids.ac.uk/websearch.html
W13. journalsonline.bids.ac.uk/JournalsOnline (BIDS)

Current awareness services in medicine

W14. medmatrix.org (Medical Matrix)
W15. ohsu.edu/cliniweb (the Oregon Health Sciences Library)
W16. reutershealth.com (Reuters Health)
W17. cdpc.com (US Centre for Disease Prevention and Control)
W18. webmedlit.com (WebMedLit)

Medical search engines

W19. glenlib.demon.co.uk/medsearch.htm (Our own library search tool)
W20. http://205.232.26.2/mwsearch/index.html (Medical World Search)
W21. ohsu.edu/cliniweb/ (Cliniweb)
W22. http://omni.ac.uk/ (Omni)

Medical illustration

W23. vesalius.com (interactive anatomy images for surgical teaching)
W24. macromedia.com (the programme "shockwave" for interactive imagery)
W25. sanger.ac.uk (gateway to data on the Human Genome Project)

Libraries, scientific and medical museums

W26. loc.gov (US Library of Congress)
W27. portico.bl.ac.uk (The British Library gateway to Web resources)
W28. http://opac97.bl.uk/ (British Library OPAC project)
W29. http://copac.ac.uk/copac (Cooperative library resource project)
W30. nmsi.ac.uk/collections (UK National Museum of Science and Technology, and Wellcome Institute Medical collection)
W31. comlab.ox.ac.uk (Virtual Library and museums listings)
W32. asap.unimelb.edu.au/hstm (WWW Virtual Library of Science and Medicine)
W33. w3.org/pub/DataSources/bysubject/Virtual_Libraries (A list of virtual libraries on the Web)
W34. http://sunsite.berkeley.edu/InternetIndex (The Librarians Index to the Internet)

Commercial online publishing

W35. europe.idealibrary.com (The online journal library of Academic Press)
W36. journals.wiley.com/cancer (The journal *Cancer* from John Wiley and Sons)

This list is not exhaustive. Its sites are predominantly in the US and UK. Readers who identify other sites of general or educational interest to surgical oncologists are invited to submit a short review of the site to 'Highwayman', care of the Editor, for occasional publication as 'page fillers' in the journal.

Self-assessment exercise

Based on the Educational Section in the February 1998 issue.
'Endothelin-1 and tumour development' by E.H. Asham, M. Loizidou and I. Taylor
and 'Appraisals of transplantation for malignant tumours of the liver with special reference to early stage hepatocellular carcinoma' by R. Pichlmayr, A. Weimann, K. J. Oldhafer, H. J. Schlitt, G. Tusch and R. Raab.

(Answers given on p. 143)

1. Endothelin (ET) is a small vasoactive peptide, initially isolated from endothelial cells.
2. ET-1 secretion is limited to normal vascular endothelial cells.
3. Endothelin function is limited to its vasodilatory effect on vascular endothelium.