Medical Information in the Networked World: Easier to find the citation, more difficult to understand the industry

Tony McSeán,
Director of Library Relations, Elsevier & Past President of EAHIL

When the organisers very kindly invited me to introduce today's event, they asked me to provide a general introduction to the way the distribution of medical information has developed. To make such a request to a person of my age is a potentially dangerous business, of course. Many of us need little encouragement to float away on clouds of nostalgia, describing the days when real librarians roamed the tangled jungles of printed indexes and everyone had to copy everything out by hand because photocopying was so expensive. Please bear in mind that I am talking exclusively about scholarly scientific and medical information, and that much of what I have to say is not at all applicable beyond these limits.

However, I personally have no such feeling of nostalgia for the stone age, and although I have been a librarian long enough to remember primitive man's (by whom I mean of course Eugene Garfield) first efforts at computer-generated indexing I think such tales, like the sporting triumphs of one's youth, are best left untold. So, my intention over the next 15 minutes is to recount in no detail at all how we arrived were we are, and to set the scene for Maria Martens and my colleague Dominic Vaughan to provide you with more detailed insights into what might lie in store for us.

Let us pass quickly over the first the first computer system I remember for anything other than an inability to produce readable print. The US National Library of Medicine's original Medlars system was a batch system. With immense care, experience and trepidation you would formulate a Boolean search question. You would then, in effect, telex it to Washington where it would be run against the database overnight. You would then come into work the next morning to find somewhere between zero and five million citations printed out (I exaggerate of course, it only looked like five million) and you tried again the next night. It was like playing chess by post without knowing the size of the board. Astonishingly, people made Medlars work, although in retrospect it might have been easier to memorise the entire database rather than just its detailed structure.

Medlars very soon became Medline. At first, dialling in to Medline was both fearsomely expensive and involved acts of skilled criminal vandalism on telephones and switchboards, but the essentials of online bibliographic searching were all there. Portfolios of databases large and small were constructed by Dialog and others and a new industry was created. Searching was limited to bibliographic citation and (sometimes) abstract and keywords. Despite these limitations the periodical literature opened up as never before and interlibrary loan services entered their golden age. Users and librarians alike were enchanted with their new information horizons. It was also essentially a mediated, librarian-only activity because of the professional skills required, the expense, technical fragility and terrible interfaces - and while some readers were irritated by this few saw passed the novelty and questioned the text-only nature of the system and the need for two-stage access using old-fashioned print and photocopy.

For a short while, the CD-ROM looked as though it might develop into a local alternative to online searching, particularly in countries with very bad telephone networks. Optical discs were even used for experimental full-text distribution. But in both cases the technical and storage limitations caused problems from the very beginning. I remember watching the Adonis system (based on what seemed like thousands of 30cm discs) in use at the British Library and feeling like I had stepped into one of those darkly comic passages from a Bunuel film.

The demise of the CD-ROM was begun by cheap 32bit desktop computers with big hard discs, and then finished off by the internet. Suddenly there was a usable interface, low-cost access to the whole world, integration with desktop computing, standards that let you join in, and above all dial-up that didn't need five technicians to accomplish successfully. Speaking personally, I remember reacting to my first sight of a browser with a mixture of extreme eagerness and excitement and a chilling worry about how long before this developed into something that would make librarians as obsolete as transatlantic passenger ships. To a considerable extent this is still my reaction, and I shall return to this later.

I spoke just now of the "demise" of CD-ROM, but of course one of the characteristics of librarianship is that no technology or service ever actually dies altogether - even something as unpleasant and useless as microfiche. They all just linger on adding to the cost and complexity of our operations - and the switch from what we might call the Dialog generation

of online searching to where we are now has taken over a decade and is still incomplete. However, the limitations previously taken for granted were soon only too apparent. End-user searching swept the academic world , using both general-purpose search engines and friendlier provisional interfaces of which Ovid is the best-known example. Text-only suddenly came to seem terribly old-fashioned. Everyone involved in the publishing cycle realised very quickly indeed that the web seemed to have unlimited potential for the development of electronic alternatives to the printed journal. What we have today - and what Dominic, Marie and subsequent speakers will be talking about - are various attempts to harness that potential: to deliver all the functions of the scholarly periodical in a way that fulfils all its traditional functions, and in a way that ensures that the system is sustainable

We are moving (and in the case of many scientific and medical institutions have already moved) to an information environment when an overwhelming proportion of the information researchers need is instantly available to them at their desk or laboratory bench. It is probably also available at home, and when they are away from home it is probably no further away than the nearest Starbucks clone. Almost all of the information not instantly available is quickly identifiable online and available either through purchase or via the worldwide interlibrary loan network which is probably librarianship's greatest achievement. Increasingly, libraries such as the British Medical Association are providing these loans as PDFs delivered to the desktop within a day of asking. The end of the printed journal is in sight, and the end of the library as a place scientists need to visit is already here in some disciplines and coming soon for the whole of science, technology and medicine.

This is already having tangible effects on people's behaviour and on the written record itself. It has been calculated that simply making journals available via the desktop PC saves each researcher on average 20 hours per year compared with having to go to the library to read them. This may not sound a lot but for some of the academics I have met while watching football it represents a 100% gain in productivity. At the same time, the improvements in search tools and the easier availability of full-text are encouraging researchers to read more — a steady increase in the numbers of journal articles read each year has been shown in many studies, right across the range of science, technology and medicine. As the internet has spread, we have seen an enormous growth in the number of papers co-authored from multiple institutions and from multiple countries. (I was at this point intending to list my own

international co-authorships- which extend to Brazil, Norway, Japan and Iceland - but since this is an oral not a written paper there is no self-citation to be gained so I will spare you.)

Amidst all this complex, rapid and highly visible change it is easy to forget that the underlying fundamentals remain unchanged almost from the very beginnings of the modern periodical. The basic functions of the scientific periodical remain intact, and when the printed record is replaced, its online equivalent will need to be at least equally good at each of the four basic functions of scholarly record. What is being talked about today needs to be judged by the following criteria:

Registration - identifying a particular discovery or theory as the work of named individuals from named institutions

Certification - an assurance by peer review that what is reported is valid, original and significant

Awareness - getting the information to the scholarly community and to others with a desire to see it

Archiving - ensuring that a secure and accurate copy of the published document is retained over the very long term

Sustainability - ensuring that any new model of operation is acceptable to the authors and editors, as well as self-sustaining financially and operationally. It is a misreading of reality to think that researchers and academics are keenly interested in the technicalities of publishing and information retrieval. As authors they want to be dealt with efficiently and to be valued. And as readers they want the tiresome information-seeking component of their work to be completed as quickly, and with as little distraction as possible from what they regard as their real work.

It is important to keep these five basic concept in mind when evaluating new models for scholarly publishing. The web offers many exciting new possibilities for dissemination of knowledge, many of which you will hear about as today progresses, many of which tend to be pulled together under the general heading of Open Access. Within the Open Access genus

you will find two families of ideas which are fundamentally difference - what Stevan Harnad has termed Gold OA and Green OA. Gold OA is like old-fashioned publishing except that there is a fee for publication and no fee for distribution. Publishers of all types are currently experimenting with variations on what is sometimes known misleadingly as this "author pays" model. There are specialist OA publishers such as BioMed Central, publishing collectives of varying degrees of formality, and now several commercial publishers offering author pays (around $3,000 usually) alongside the subscription model. Green OA is the self-archiving option, where the author posts a copy of the published paper on a personal web site or institutional archive. Green OA is wholeheartedly supported by a substantial majority of the major medical publishers.

I would like to close by spending a minute or two talking about the profound long-term implications of these developments for libraries and the library and information profession. The importance of *the library as place* is declining rapidly, to the point where the physical library may disappear from scientists' lives altogether (as it already has in many pharmaceutical and chemical companies). Without the social and service reinforcement of a major public space, libraries will need to work much harder to brand and to market our services and to make sure policy makers stay aware of the value we provide on their investment. Already there are many researchers who think that services like Elsevier's Science Direct are free-to-all and not a product of their library service's care and investment.

Will librarians disappear? Not immediately. There is no doubt that at the moment quality information services depend on skilled professional mediation and intervention and that good research depends on good information support. We add crucial value research that the current range of self-service tools cannot match. Even Google Scholar. We will certainly need to work more in the field and less in our office - taking our skills out to where they are needed, now that we no longer need miles of shelves and tons of dead wood in order to apply them effectively.

What of the longer term future? I am less optimistic. Institutions will still need information specialists to ensure that information provision matches the quantity and quality required by staff and students. Whether we will continue to be able to beat the web tools of the foreseeable future is another matter. For years received wisdom was that chess grandmasters would always beat their electronic rivals because of their pattern recognition skills; massively

parallel systems and sheer processing horsepower has shown that to be false. Current research in semantic linkages and pattern recognition are, I feel, likely to chew away at our human advantage as it has the chess player's.

For me one of the fundamental lessons of the take-up of web services is that they do not have to be better or even as good as the non-web services they have replaced. They only need to be good enough. And on that worrying note I will thank you for your attention and hand you over to what I hope will be more cheerful people.