B. Cesnik

Monash Medical Centre
The Centre of Medical Informatics
Clayton, Victoria
Australia

Synopsis

Digital Libraries

Introduction - The library of the past

Libraries have always been the organisations that serve the need of a society to store its ever-growing knowledge of the world in which we live. The purposes to which such knowledge is put, the manner in which it is represented and the responsibilities of those who are its custodians has changed dramatically over time, never more so than in the last century. As we look forward to this, the third millennium, libraries are becoming 'digital' and their structure, function, role and activities are broader in scope and face greater challenges than ever before.

Libraries, digital or otherwise, have always represented the knowledge 'asset' of a society. Historically, past cultures are often remembered for the sophistication and importance of the libraries that they created. Explore the history of the library of Alexandria as an example of the relationship between societal advancement and the role of libraries in reflecting achievement.

For most of history, libraries have had to deal with the limitations and possibilities available in the physical world. The ability to use 'words that last' to store and communicate knowledge has always been a powerful tool, this was well understood by some of the original librarians, those of the church (regardless of denomination). The ongoing and meticulous work of theologians in documenting knowledge and using it to shape culture is well known.

The limitations of this pre-Gutenberg time were the sheer task of writing, the availability of the medium on which to write (and its impermanence) and the physical requirement for safely storing and archiving the results of such efforts. We still marvel today at the detail and care with which knowledge was recorded by hand in centuries past, at the same time we note the need to care for the frail medium of the time (ancient inks and ancient paper).

The liberation of print by the invention of the printing press is a story well known to us all. In a very short period of time, those previously responsible for the laborious task of hand writing library works found they had to completely reinvent their role as carers of knowledge. Dissemination of knowledge escalated dramatically as printing presses could produce print at orders of magnitude faster than any collection of scribes. I am sure that the accompanying reduction in what was then seen as 'quality' of print distressed those who felt that books were a combination of art and words, not print alone (an early example of change management ?).

From this point on, the world of libraries changed as they adopted new roles in the dissemination of knowledge. The concomitant growth in available knowledge (it was now much easier and therefore cheaper to produce new books) resulted in libraries having to invent new and more robust ways in which to order this knowledge to facilitate retrieval, in this we can see some of the the origins of knowledge management.

This situation of a print-based world, operating within the domains of the physical, is the story of libraries in the past centuries. The last century saw the world, once again, undergo another massive change in possibilities, the digital world had arrived.

The birth of the Digital Library

The 20th century brought together changes in technology, society and the exponential growth in 'that which can be known' to challenge the libraries of old and the paradigm of print and geography.

Electronic storage capabilities appeared and still continue to offer ever increasing abilities to store not merely print but also the variety of electronic media we now increasingly rely upon such as sound, images and even spatial information such as the visible human project. With this capability, the relationship between knowledge storage requirements (books) and physical need (buildings) has been usurped. What previously required ever increasing physical space for storage now requires less and less space as digital storage becomes ever cheaper and smaller. The school

children of today frequently access encyclopaedias on CDROM that only three decades ago occupied shelves of volumes of books.

Processing power continues to advance with Moore's hypothesis of transistor density doubling every 18 months still holding true. While the physical limitations of silicon and other rare compounds is expected to reach its limit by 2010, we are already seeing biological storage and processing methods which operate at the molecular level opening new horizons. Storage and speed will not be the limiting factors to our digital world, again a major difference to the very real limitations faced in the pre-computing library world.

The most dramatic technology advance coupled with the above is that of computer networking. In the physical world, the transmission of knowledge embodied in print remained a physical task with time, cost and physical requirements presenting major barriers to communication. Interim technologies such as photocopying and facsimile transmission assisted in overcoming some of these barriers but have always represented mimicry of the physical, one way delivery not interaction.

Early computer networking allowed libraries to enhance access and interaction with knowledge. Initially within the walls of a library but rapidly to the local geography in which they existed (such as Universities). The telephone system allowed this 'reach' to extend further but until the arrival of the Internet, libraries were not faced with a truly global network for which they themselves were not the architects. This transition from a repository to a site of access has been a major feature of library change in the 20th century.

So, at the start of this, the 3rd millennium, we live in a world where libraries now have tools more powerful than ever before, massive storage capabilities, rapid processing power coupled with a global network. This network continues to extend its reach

not merely between institutions but directly to the homes of us all. I watch, sometimes in wonder, as my 10 year old daughter sometimes loses the distinction between information on a home computer (of massive computing power compared to its mainframe ancestor) and that which she finds via the Internet, to her the network is the medium. The digital library of today is fortunate in that its real task is to manage knowledge and facilitate its dissemination, adoption and application. In reality, this has always been the case, today, the barriers are falling but the nature of the task has become infinitely more complex.

The Digital Library as Knowledge Manager

Access to ordered knowledge kept in libraries has, for most of history, only been available by journeying to a location and physically interacting with the repository itself. Access to such repositories has steadily increased so that in the last century we have seen local libraries develop in communities, University libraries shift more of their space over to people rather than storage and, as mentioned, the application of early networking to extend reach.

This attention to accessibility has, as its primary reason, the desire by societies to incorporate knowledge into their lives. The physician of today faces the challenge of lifelong learning in the face of ever increasing (and almost overwhelming) knowledge growth. Additionally, the desire for the rapid conversion of such knowledge into action is also increasing. Witness the development of the Cochrane collaboration and its global efforts to examine knowledge to provide evidence for care. The translation of this evidence into practice is a challenge we have not yet addressed, this is not a criticism it merely highlights the role that the digital library faces in responding

to this change in how society sees the purpose of knowledge. Not a dry, archived collection of material but a dynamic, ever changing and growing resource to better care for ourselves and our world.

Libraries have always been the seat of knowledge management. Whether the task has been (historically) how to order physical material and cross reference it for flexible retrieval, or to provide new and innovative means whereby users can interrogate knowledge repositories in real time, libraries have worked to acquire 'knowledge about knowledge'.

The outcomes of these endeavours now extend well beyond the mere ordering of collections. Having spent all of history studying the nature of knowledge, the digital libraries of today offer us tools and approaches relevant to the generation of knowledge itself, not merely its management. The Human Genome project represents one of humanities most ambitious of endeavours and presents a knowledge management task of massive proportions. It is easy to observe how the approaches taken draw from the skills and experiences whose origins are from the world of libraries.

The Digital Library – the current challenges

The papers presented here each explore different aspects of the challenges and possible solutions we are all exploring to deal with the explosion of not only knowledge but also connectivity and totally new paradigms of knowledge navigation and analysis.

The paper by Pandolfini, Impicciatore and Bonati takes a specific knowledge domain likely to be explored by parents seeking to manage cough in their own children. The evaluation tool they describe is developed by consideration of evidence based resources

regarding cough including respected sources such as MedLine, Cochrane and The American Academy of Paediatricians. The resultant checklists created aim to ascertain the technical qualities, completeness and quality of health information present on the Internet. Their results show the likelihood of incorrect advice being found to be high which reinforces the need for such information to be seen as an additional, not primary source of information. Their advice that further work is required to develop new strategies in the production, validation and distribution of on-line information is echoed in the other papers presented.

While it is attractive to think of the digital world of instantaneous access to well managed knowledge repositories, available to all and relevant in time, place and context this is not yet the reality we have. One of our best known and respected resources is that of MedLine, the now nearly globally available means of accessing information and knowledge that has been examined, categorised and then wrapped in a set of tools to facilitate access by those seeking to use its wealth to inform their activities, yet its origins are in a world of paper.

In the paper by Murray and Anthony, a taxonomy is offered to described the approaches taken in making traditional publications available via the World Wide Web. This taxonomy recognises the range of options from the straightforward transfer of paper content to a WWW format to the previously impossible structure of end user, on-going interaction around publications adding discussion forums, messaging systems and 'evolving' publications resulting from on-line, community interaction. They note that replication of the paper format to an on-line version is the more common approach we have today.

This highlights the reality that paper has been with us for a long time. In publishing we use it to encapsulate our

reporting of knowledge for the perusal of others. Regardless of whether the 'document' we create is electronic, transferred via email and then made available via the web, the paradigm of paper still remains as predominant model of dissemination.

With this also comes the vexed issue of copyright. Those who generate knowledge through work, the application of resources (which are always constrained) and the need to continue to garner support for their work must have the recognition they deserve. Those referencing the knowledge of others in advancing their own pursuits accept the obligation for attribution recognising that knowledge grows in an interactive fashion where the work of today forms the foundation for that of tomorrow.

In a totally global, digital world, how do we (and thereby libraries) protect this right of recognition without limiting the accessibility we desire? Further, as hypertextual representation of knowledge increasingly becomes the norm, what must we do to revisit the traditional approaches to copyright protection. When a 'document' is richly interlinked to distant electronic publications so that the borders of intellectual ownership become blurred, what defines the scope of the document?

The paper by Pagesy, Soula and Fieschi offers insights into the development of this hypertextual reality. They describe an approach to improve knowledge navigation in hyperlinked resources that applies adaptive presentation and navigational techniques to enhance user representation, identify motivation in knowledge searching and match knowledge resources to the individual. The paper does not explore the impact of such enhancements to copyright but it is easy to see that the issue will merely grow, not disappear.

The legacies of paper will not disappear so soon. As a simple example,

those planning scientific conferences today usually have one similar discussion that all share. Should the proceedings of the conference be published in paper and digital (usually CDROM) form or can we consider digital only, few to date have made the leap into 'paperless' proceedings. Many libraries still require the paper version thereby removing some of the tension in the discussion but we are rapidly approaching a world of digital dominance, libraries face this problem at orders of magnitude greater than those faced by a conference committee. Add to this the reality of historical publications, for which no sound economic argument can be raised to justify conversion to a digital medium, and the scale of the challenge for the digital library becomes even more apparent.

Digital Libraries, the new challenge

The computing technologies of the 20th century introduced ways to do old work better but they also presented entirely new ways of working. As libraries advance in their role as knowledge managers perhaps the greatest 'new way of working' will be to move knowledge into action as a part of action itself. Consider that a physician working with patient information increasingly comes to expect assistance in decision making to be embedded in the systems they use, prescription generation is an obvious yet powerful example.

The knowledge in the decision support system is that distilled from the known research and study of relevant areas, knowledge regarding drug-drug interactions which has been developed from laboratory research as well as epidemiological research through randomised controlled trials. Our present processes for moving knowledge to care remain cumbersome. A Medline search regarding a patients

condition may be of assistance but still requires the processing of this knowledge into action.

Even when available, how reliable are these knowledge stores? In their paper 'Evaluation of Consumer Drug Information Databases' Choi et al evaluate prescription drug information contained in commonly available drug information databases. While these are distributed on CDROM, there is no real distinction between this medium and the networks we use daily. Their conclusions regarding the most reliable of the considered resources is interesting however, the evaluation approach they offer again shows how the rigour we apply to evaluation of evidence is no less necessary in a digital world (in fact it becomes greater). They, as do others, predict the increasing access of such resources by consumers themselves and the need to evaluate extended capabilities of user interaction to prove their benefits in enhancing utility.

Given the enormous explosion in knowledge, it is no longer enough to merely provide access to knowledge. Tools for its representation and manipulation and, critically, allow for its incorporation into other information regarding care are the modern challenge for digital libraries.

In the paper by Goldman et al, a technique for data mining of text databases is explored. The desired ability to extract new knowledge or confirm known hypotheses is evident in this work with the approach taken being applied to a previously noted observation (left-versus-right lung ratio in thoracic lung cancer). Knowledge discovery systems such as this are becoming more sophisticated and prevalent as we attempt to make greater use of the information stores we are creating and connecting.

The Unified Medical Language System has some of this functionality apparent in its purpose. Beyond the mapping of nomenclatures, the enhancement of knowledge represen-

tation and advancing our understanding of how to describe healthcare, this project also contributes to the knowledge models we require to advance support tools for care, at the point of care.

Kahn, in his paper exploring the use of the Standard Generalized Markup Language (SGML), provides insight into a growing approach to the information explosion. Rather than a distinction between free form textual material and structured data, can we make available information that is a composite of both, in other words, documents that carry their own structured information with them? While SGML has been with us for many years, its children such as the Hypertext Markup Language (HTML) and, more recently XML are changing the way in which libraries are able to publish.

If we are to realise the Decision Support grail of 'the right information at the right time to the right person in the right context' there is no doubt that the role of our digital libraries will loom large.

Digital Libraries – Today's reality

It is reasonable to observe that we are in a period of profound transition as we move from the physical to the digital. The digital libraries we hope for and see emerging face great and exciting challenges in balancing the reality of a knowledge explosion that started occurring prior to the tools we now see becoming part of day to day knowledge management.

New forms of knowledge will further contribute to this explosion both in volume and character (what is the role of the genome in the medical record?). As Medical Informatics continues its pursuit of transforming knowledge into care, it will ever increasingly rely upon and contribute to the development of the digital library.

Libraries have been 'home' to knowledge since time immemorial, the fact that the physical walls of the past are rapidly becoming invisible does not change the reality of that history.

The connected and rapid digital libraries of today and tomorrow will further serve the integration of knowledge globally while defending the recognition of those who create such knowledge, the challenges and opportunities are as great if not greater than they have ever been. So it is reasonable to assume that, whatever their appearance (physically or virtually), our libraries will continue to be an intimate part of the fabric of care, our scientific endeavours and our understanding of knowledge itself.

The term 'Library' has traditionally described a physical place set apart to contain books for reading or borrowing, the term needs to be expanded in its meaning to encompass the digital age. The term 'Publish' has two meanings, the first is to issue copies made by printing (or other processes) a book, periodical, map, piece of music, engraving or the like. The second is shorter and successfully includes our new future – 'to make publicly or generally known', the quality of that which we make known, the means whereby we interact with such resources and the nature of the information itself are exciting fields of endeavour.

Address of the author:
Branko Cesnik
The Centre of Medical Informatics
Monash Medical Centre
Level 1 E Block
Clayton Rd
Clayton VIC 3168, Australia
E-mail: branko.cesnik@med.monash.edu.au