Scholarly communication in an electronic environment; problems and challenges

Introduction

It is considered unfashionable in many quarters to announce oneself as a librarian. Yet librarians have produced one of the most remarkable feats of international co-operation in the last fifty years, which has gone unheralded, unannounced, but whose characteristics should be informing the development of standards for the Internet and the World Wide Web. Indeed so great is that success that we take it absolutely for granted. Broadly speaking it is possible to identify any recent published work from anywhere in the world and whether or not one has identified a location for the item to borrow it or a facsimile of it either free or for a very small sum. This article will reflect on that success and on its relevance to the development of electronic information where a dangerous complacency assumes that our skills are so minor and/or our influence so limited that we must cede the field to larger players. It will argue that our skills are in fact entirely appropriate to developing the scholarly infrastructure which is so glaringly missing from the Internet and the Web. As a convenience, the building blocks will be split into two groups, reflecting IFLA's original core programmes of Universal Bibliographic Control (UBC) and Universal Availability of Publications (UAP). This model is used because it is through international bodies such as IFLA that this success has been sponsored and achieved.[1]

Universal Bibliographic Control

Throughout the first half of the twentieth century the dream of universal bibliographic control was largely embodied in those monumental multi-volume works of erudition, the Library of Congress Catalogue and the British Museum (later British Library) Catalogue. Although they filled great banks of shelves in their dress of green and blue, no research library was (or is) complete without them. But they were never easy to use and followed sets of independent – one hesitates to say arbitrary – rules devised by great libraries conscious of their eminence and independence. In the last few decades a quite different spirit has pervaded the profession and it has been quick to recognise the opportunities presented by automation and the globalisation of research. The international nature of scholarly literature has been reflected in the growth of international (or at least interoperable) standards which at least complement if not replace these great works of scholarship – and national libraries have been at the forefront of this standardisation. A large number of committed professionals, operating under the banners of national libraries and professional organisations has produced and maintained a string of standards which collectively have made the identification of almost any published item a straightforward task. Almost more importantly, these standards have been enthusiastically embraced by the profession and by related groups such as publishers and booksellers. The Anglo-American Cataloguing Rules, the MARC standard, ISBNs and ISSNs and the growth of national bibliographies all represent a huge and successful professional effort over the last four decades to globalise and standardise bibliographic control. Subject access as well as author/title access has also been fully explored. There were, of course, diversions as library thinkers made detours into the exotic world of faceted and other classifications, but the dominance through constant revision of both the Dewey and LC Classification schemes has also forced order and structure on to an essentially chaotic publishing output. Curiously, the financial pressures of the last two decades have also helped to gain control of what is published. Research libraries have introduced coherent collection development policies in an effort to focus their collections and as a response to those financial pressures. At the same time consortia have appeared which share responsibility for acquisition and good catalogue access is a pre-requisite for such sharing. As library housekeeping systems became universal in the 1970's and 1980's in the developed world at least, co-operative catalogues grew and retrospective conversion on a quite massive if unco-ordinated scale took place. Although not yet complete it has massively enhanced access to collections throughout the world from throughout the world.

Curiously, what may be seen as the one major gap in this roll-call of success has been filled very effectively by commercial publishers. The abstracting and indexing of journals has a long and largely commercial history; but it too has provided a near comprehensive system of access which has kept pace comfortably with the great post-war expansion of scientific literature. It might however be argued that the system is weakest here in its coverage of non-scientific literature and in its coverage of non-English language and minority interest titles. But even in this long-standing activity, Eugene Garfield's innovative concept of citation analysis brought a new thrust to the exploitation of journal literature, although the citation indices cover only a fraction of published titles.

Political and economic instability in many parts of the globe may have prevented the total establishment of UBC, but as a profession we have developed and actively maintained a vigorous and robust bibliographic infrastructure. All of this has required a constant but largely unremarked stream of diplomacy and activity of the very highest order. There is no self-evident reason why a spirit of international co-operation should have triumphed over national ambition.

Universal Availability of Publications

If UBC can be considered a professional success, the development of UAP is even more astonishing in its final form. Universal Availability of Publications has been a particularly British led activity – indeed IFLA's UAP office is based at the British Library in Boston Spa. If bibliographic control is a necessary pre-requisite of access to collections, inter-library lending is its acme. Of course libraries have lent books to other libraries or scholars for centuries, but the development of standardised national and international systems with first Urquhart then Line the successive Directors-General of the British Library's Lending Division at Boston Spa acting as leading thinkers, advocates, drivers and promoters of internationalism has created the hard-won rather than inevitable system we see today. Inter-library lending is now a core activity in support of scholarship rather than a peripheral and exceptional one. And yet it is not self-evident that subject-based document delivery systems such as that of the National Library of Medicine or co-operative systems such as that run by OCLC for its member libraries should be interoperable with systems run by a major national library such as the British Library. The development of a common currency; the very act of trusting fellow professionals – and even more implausibly their readers - in small libraries in countries at the other end of the globe to act in a uniformly responsible way; the meshing of different copyright and fair use traditions and legislation; all represent quite remarkable acts of international co-operation and trust. Not all is perfect, of course, since not everything is deliverable and sometimes the scholar must still move to the book. Nevertheless it is broadly true that scholars anywhere can identify and acquire or gain access to the most obscure works of scholarship, wherever held.

In sum, the combined efforts of whole teams of committed librarians sitting in committee rooms throughout the world has created universally adopted systems of bibliographic description and mechanisms for accessing the literature wherever it may be found over recent decades. It might be argued that these systems are a natural consequence and expression of our professional skills and interests, but even that view would have to acknowledge the sheer scale and complexity of what has been achieved. The system runs so smoothly that it is ignored and its lessons are not being used to inform discussion on the emergence of exactly the same issues in the electronic world we face. And the task is much harder this time. In the world of print the field was largely ours to make of it what we would, and where co-operation was required it tended to be with empathetic groups such as publishers who understood the need for standards and valued the contribution we could make as a profession. In the new environment we are jostled and compete with computer scientists, publishers, authors, lawyers, learned societies and agents, all of whom feel they have a role to play and skills to offer. This is perhaps most simply illustrated by the creation of the ubiquitous URL as a standard. It requires a very particular skill to create a descriptor of an ugliness and lack of structure which makes one yearn for the relative simplicity of faceted classification.

The Future

Using the model of UBC and UAP it is possible to review the position and the issues in the electronic environment in the same way. The lack of thought and imagination here is in stark contrast to that of the printed word and there is a real danger that scholarly communication will be seriously compromised at least in some disciplines, through a naïve blindness to the glaring problems of the Internet as a medium for scholarly communication. A whole host of problems surrounds the electronic equivalent of bibliographic control. Some are a function of the medium while others are variations of old issues where the new environment poses new challenges. A thread which runs through all of the problems is the failure of the academy to recognise that the problems exist. There have been spasmodic attempts by individual scientists or groups to open up the debate but these have made little real headway[2]. The Internet is seen as a great and liberating development, but it is not a neutral development and requires very substantial international effort if it is to be made usable for sustained scholarly communication rather than short-term gratification.

Electronic UBC

There is a whole range of issues to be addressed[3]. Some of these are being worked on by groups and some are not. Probably all of them are soluble. But there is no real debate of the underlying issue of the future of scholarly communication. There is if anything a vague assumption that such communication will adapt to what is available, rather than there being explicit statements of what is required. This constant plugging of holes in the dam avoids the need for a debate on how water is best managed.

The very act of naming and identifying electronic objects consistently is fraught with difficulty. A book is a static object which does not change over time. Some electronic information has the same characteristics, but most does not. In an electronic environment there is a need to reference objects as they move and change over time and place. The temporary nature of URLs is notorious, with figures ranging from 75 days to six months being given as the average life of a URL[4]. This author was involved in teaching a course recently which involved citing some 64 URL's. These have changed or disappeared at the rate of four a month over the course of one semester – and that in the field of information management! Even national libraries, library associations and departments of information studies and librarianship cannot manage this process effectively. PURLs or permanent URLs are being worked on to resolve this issue.

Even where the URL remains constant issues of version control and quality assurance remain unresolved. Issues to do with network costs will lead to an increase in mirror sites. Such mirrors will inevitably be "out of synch" for periods of time. Worse, sites contain unofficial copies of unknown provenance and accuracy. At least some of the projects seeking to put large volumes of text into electronic "libraries" deliberately and ostentatiously ignore issues of version and quality putting up any out of copyright text which is available rather than only texts which are of scholarly worth. The seriousness of this problem cannot be overemphasised for the continuity of citation is central to scholarship and without it scholarship cannot flourish. The ability to refer back and forward to agreed texts and articles so that others can replicate the work is as critical to the Arts as the Sciences. Some attempts are being made to deal with this problem, the current favourite being Digital Object Identifiers[5]. The International DOI Foundation is a non-profit organisation although it stems from the commercial world and DOIs may be seen as a sort of electronic equivalent of the ISBN. However a significant if unquantified proportion of the material held in any library and any organisation (such as a university) and in any medium is either non-commercial or out-of copyright and any new system must be able to embrace everything from incunables to examination papers, yet must remain a permissive rather than a mandatory system.

The issue of the authority to name objects is also difficult and shows no sign of being resolved. At present anyone can publish anything on the Internet and can name any object

without reference to any standard or organised registration body and with no obligation to maintain the name over time. There is no minimum requirement such as, for example, giving the date of the last update or version of the object. This is compounded by the fact that many of the reference points we take for granted in the print world disappear. A book published by Cambridge University Press implies a set of values, standards and scholarly rigour that is understood. But an address incorporating the phrase "cam.ac.uk" could be anything from a university press to a student p.c. in a rented room. The persistence of object names is a long way from having a settled structure – and there is little evidence that the official bodies in scholarship understand the threat this poses. It is ironic that the breathtaking growth of the World Wide Web in particular – measured in months rather than years has not given time for any rational or ordered examination of the implications for scholarship.

Metadata and the standards for the description of objects is rather better developed. The Dublin Core standard first produced by Stu Weibel at OCLC has very rapidly developed enthusiastic international acceptance and now has an international development structure with participation in standards work from Europe, USA and the Pacific Rim. But even in this area where librarians have intervened successfully much work remains to be done. Cataloguing has historically described static and largely immutable objects. The Internet offers new genres of multimedia and even services which will require appropriate description. This work remains to be developed. Large scale efforts such as OCLC's Project CORC are also reviewing how metadata can link to existing MARC tagging structures. But there is a worry that electronic objects differ fundamentally from printed objects. A whole class of new information is required to describe electronic objects adequately and many of these will be inherent in the terms and conditions of sale rather than the object itself.[6] Many items will have multiple copyright permissions with images, tables, data and text all having different owners. Some data will be leased rather than purchased and will have different restrictions on different categories of users at different periods of time. Not all data is available to everyone on the Internet

Searching the web is a much more difficult exercise technically than the optimistic designers of web-crawlers would have us believe. Issues of relevance matching known to the information community for many years have only just begun to prove significant for the Internet. Web indexing systems are breaking down as their architecture collapses under the weight of data. It is increasingly common to undertake a search on Lycos, or Excite or Infoseek and recover hundreds of thousands of hits in apparently random order. Much work is going on here but designers despair at the inability or unwillingness of the public to master Boolean searching and most systems still have a long way to go to beat a half way competent reference librarian. The search engines have themselves recently come under proper scrutiny. Rather to everyone's surprise it has become apparent that they address only a fraction of the estimated 320 million web pages. Coverage varied from a best of 34% for Hotbot to a worst of 3% for Lycos.[7] Within that, up to 5% of links were "broken" although "pages that timed out were not included in these statistics"[8]

Web searching has undoubtedly transformed the ability of researchers to acquire a whole range of current reference information, but is dramatically poor at discovering scholarship and research. Relatively few electronic journals are accessible on-line; contractual issues make them difficult to use; other scholarly data is difficult to obtain. One cannot with confidence find appropriate resources which are appropriately available. Perhaps the largest scholarly directory of Internet resources is BUBL[9] and it lists a mere 12,000 addresses – a figure less than a small departmental library.

Electronic UAP

In the print world, once one fulfils the basic qualifications for membership of a library, all of its contents are available for use (with isolated exceptions), and the interlending service provides access to resources not held locally. The position is quite different in the electronic world where we will require validation of the rights of the user in relation to each object.

Typical developing contracts allow access to specified groups for specified periods (eg all matriculated second year law students for the third semester only). Another variant is that access is only from specified IP addresses or locations. User authentication is regarded as an essential element of electronic commerce, but it too lacks basic elements for the furtherance of scholarly activity. At present there are no good ways of proving membership of the "data club" when away from the parent institution. Scholars visiting another institution, students on vacation or researchers on field trips are difficult to validate. There is then a very knotty problem surrounding usage data. On the one hand commercial publishers wish to collect usage information as a marketing tool. They are, however, unwilling to release this information to libraries so that they can judge whether usage justifies subscription. Conversely many users do not wish anyone to know what they are reading or researching. Traditionally, libraries have preserved the anonymity of user data except where criminal acts are suspected. Is this a right or simply a custom?

The preservation and archiving of electronic information has only just begun to surface as a very complex issue. The Data Archive at the University of Essex has existed for some twenty-five years and has perhaps as a clear a picture as anywhere of the so far intractable problems of storing, refreshing and kite-marking information. The problems are staggeringly complex technically and staggeringly expensive to resolve. Although some progress is being made on the legal deposit of commercial material, little appears to be done on the non-commercial and primary materials of scholarship. A preliminary study, the CATRIONA II project, was funded as part of JISC's e-Lib programme to examine the nature and range of electronic materials produced by institutions. It found that there was a great deal and that institutions had no mechanisms for dealing with such material whether in terms of archiving or rights management. There are no standards or control or approval mechanisms for institutions or data repositories. This position may be compared with that in the United Kingdom where archives are expected to meet the BS5454 standard and the Historical Manuscripts Commission takes an active interest in the state of repositories and where archivists have specialist professional training. A new class of electronic material, what Clifford Lynch of CNI has called "endangered content"[10] is emerging, where the formal and informal records of disciplines are effectively at risk through neglect. Archives collect papers, but institutions do not sample or preserve the electronic mail or word-processed files of their scholars. Lab books are routinely preserved by scientists but it is doubtful if any institution has a policy for the preservation of digitally captured images or data from research equipment. No publisher has yet given a guarantee to preserve material for more than three years and yet authors continue to sign away copyright in all media for all time.

The structure of networks - network topology - is barely discussed as an issue due to a naïve assumption that there will be an infinitely expanding amount of bandwidth which will somehow be made available to scholarship. But the fact that bandwidth is available – even if true – does not make it affordable. And as yet there is no evidence to support this view of availability. The major American research universities have abandoned what they perceive as the failing Internet provided by telecommunications companies to create Internet II as a private network attuned to their needs. In Europe the relatively modest ambition of the European Union to link existing research networks through the TEN-34 Project has been "shaped by a series of non-technical influences such as non-availability of required public services"[11], while "standard PNO (public network operator) services in Europe could not fulfil the requirements of the R&D community in Europe"[12]. Equally the assumption that we should accept a simple commercial approach to network planning should not be allowed to go unquestioned. At present in the UK, bandwidth for Higher Education is acquired in the light of use rather than as a result of scholarly or educational policy decisions. Thus bandwidth expands at a great rate to the East Coast of North America in order to meet traffic growth. There is almost no debate on whether policy not use should drive such acquisition and that bandwidth should be routed say to Southern Africa then India, Singapore, Australia and then the West Coast of the United States, opening up markets and scholarship by policy rather than apathy. Parts of the developing world remain free of

Internet access and communication at least at scholarly level and yet Malawian history and medicine or Papuan folklore remain subjects of significant scholarly interest even if they do not have the commercial clout of big science. There is a creeping form of cybercolonialiasm in the assumption that only the United States has digital material of value to the world. There is a certain irony in the fact that for large parts of the day in Europe the United States is almost uncontactable due to the press of traffic on the networks. This issue has been recognised in Australia where a decision has been made to encourage local web scholarship and information proiders. The Australian Vice-Chancellors have agree to use network charges to discriminate against overseas websites and in favour of Australian ones[13].

One of the most important strands of scholarship is the output of the small learned societies. We may suppose that few of them will be able to run 7x24 servers giving access to their materials, far less ran mirror sites on different continents to provide good and robust access. Yet no discussion appears to take place of how the products and output of small learned societies are to be mirrored around the world and what standards and quality controls will apply to mirror sites. Again the scholarly community is silent while the commercial giants of the STM world dictate the shape of electronic scholarly communication – despite the fact that large scientific publishers are aberrant rather than the norm.

And of course the network is not nearly as robust as the telecommunications companies would have us believe. A new sub-jargon of messages such as Error 404 and time-outs has become part of the norm of daily life on the network. A recent Dilbert cartoon pointedly and uncomfortably accurately suggested that all of the time saved through automation in the information age had been lost by people sitting at web browsers waiting for images to load. Video and film remain pathetically inadequate while networks do not yet give the reliable quality of service required for multicasting. It should be self-evident that for research institutions working at the leading edge of scholarship and indeed telecommunications, the standard services provided by Internet Service Providers will always be inadequate. Instead, like Dr Johnson the academy prefers to admire the tricks of the dog and wonder that they re done at all rather then questioning whether they are done well or badly.

There is however one area which does require positive comment. And this is the encouraging broadening of the definition of what constitutes scholarly content. Services such as the Arts and Humanities Data Service[14] based at King's College London or the excellent SCRAN project[15] funded by the museums of Scotland are much involved in the digitisation of museum and archive collections and in recording everything from the performing arts to archaeological sites. This growth of new forms of content is emerging rapidly and brings with it some novel and clear academic thinking on such issues as new licensing models and standards relevant to the academic world rather than the commercial world.. Such services also highlight the important role of collection managers in the digital environment in terms of presentation as well as preservation. But again there appears to be little concerted effort by the official organs of scholarship to build formal cross-domain linkages. In the UK it has been left to government to bring Museums, Libraries and Archives together under a new single government agency – regrettably for reasons of economy rather than scholarship.

Conclusion

It has been the purpose of this paper to argue that the very significant skills we have brought as a profession to making the printed word uniformly and universally available have been overlooked. An electronic environment is being created which is inimical to scholarship and which is largely being designed by commercial and entertainment forces which are irrelevant to the scholarly process. Even if that environment is modified and the issues described are resolved, it will remain an essentially hostile commercial environment. The academy remains largely unaware of the dangers – particularly in the area of preservation of both primary and secondary research resources. Our electronic house is built on shifting sands and a much more active approach is required from the profession to demonstrate that we can, like Sisyphus, reclimb the hill of bibliographic control and access and use that most basic skill of

library school courses – the Organisation of Knowledge – to define scholarly requirements for the emerging information society

References

1. McCallum, Sally H. (ed). IFLA Medium Term Programme 1998-2001 The Hague,1998 describes the current standards work undertaken by these programmes
2. Shaw, D. & Moore, H. Electronic publishing in science UNESCO/ICSU, Paris, 1996 is one such example, where the International Council of Scientific Unions is trying to open up a debate
3. Most of these issues were first discussed in an unpublished paper given at the European Union Telematics Conference in Barcelona, February 1998 by Clifford Lynch of the Coalition for Networked Information. This paper expands and explores the list given by Lynch
4. Parker, Sandra Paper given at 1999 Career Development Group Conference in Leeds
5. Paskin, Norman DOI: Current status and outlook D-Lib Magazine Vol 5 no 5 1999, http://www.dlib.org/dlib/may99/05paskin.html
6. Law, D.G. How microwaved is your POODLE? *Catalogue & Index* 114 p1-6, 1995
7. Lawrence, Steve and Giles, C. Lee Searching the World Wide Web *Science* 280 pp98-100, 1998
8. *ibid.*
9. This excellent gateway may be found at http://www.bubl.ac.uk/
10. European Union Telematics Conference, *supra*
11. Behringer, Michael The Implementation of TEN-34. Paper presented JENC8, the 8th annual Joint European Networking Conference, May 1997 and later published in DANTE IN PRINT, No 28 at http://www.dante.net/pubs/dip/28/28html
12. *ibid.*
13. News report in the THES No 1322, 6th March 1998
14. http://www.ahds.ac.uk/
15. http://www.scran.ac.uk/