Digital Libraries: The Infrastructure and the Training Needed

Alan Hopkinson

Abstract

In many developing countries there has always been a problem in acquiring and storing printed materials because delivery was expensive and storage difficult in a humid climate. In the digital era these problems are solvable and avoidable. However do not imagine that you can throw away all the bibliographical standards that a physical library needs. Digital libraries need a good index and need quality catalogue records and good classification to supplement full text indexing which on its own can make retrieval of relevant materials difficult. Standards like MARC and classification schemes are indispensible as are standards necessary to create the information retrieval structure. Additionally librarians must work together in teams and provide good information literacy training.

Keywords: Digital Libraries, Electronic journals

1. Introduction

Much of the literature about electronic libraries is devoted to the implementation of software to host digital materials or deals with the technology of digitising one's own collection. Much of the published literature deals with projects some of which have become the foundation of technology for commercial applications. Dempsey discusses these kinds of aspects in his paper "The (Digital) Library Environment: Ten Years After" where he considers how the digital library environment has changed in the ten years since Ariadne was first published[1]. However the majority of accesses to digital libraries worldwide are to the commercial digital libraries of journal articles and conference papers, with the recent addition of monographs which their publishers have allowed to be loaned for a set periods or sold outright over the internet. This material is devised by the owners of the material to be reasonably accessible with the proviso that they must protect the copyright of the authors and their own income from the sale or licensing for use of the material. Moreover since much material that used to be in print is now becoming digital only, this has management implications for libraries. Libraries therefore have to set up new infrastructures to supplement the old printed collections and move towards a digital library which is unlikely to be totally digital for the foreseeable future but will be a hybrid library, a mixture of digital and print.

2. Academic Material in Libraries Outside the West

In countries which have not published their own materials, there has always been a problem for libraries acquiring books and journals. In the days of print journals, they were too expensive to purchase at the price they were available in the industrialised world, and additionally there were postage costs to be accounted for. In the case of journals there was a feeling that a run should be

complete and so when as often happened subscriptions could not be kept up the library felt it was failing in its responsibilities.

2.1 Electronic journals

Digital journal (e-journals) have been in existence now for some years. Initially they were made available on CD-ROM for the benefit of developing countries, but nowadays a large number of academic institutions outside Europe and the USA have bandwidth good enough to download articles from e-journals, if not as good as in the west. CD-ROM is now an obsolete medium except for organisations which are usually not-for-profit and provide materials for use in developing countries. For example CABI (Commonwealth Agricultural Bureau International) publish in partnership with KIT (the Royal Tropical Institute, Amsterdam) TROPAG and RURAL, a bibliographic, abstracting and indexing database that brings together literature on tropical agriculture from the developing rural areas of Africa, Asia, the Pacific and the Americas. This database is available on CD-ROM as well as on the Internet (updated bi-annually). E-journals for the most part, especially those aimed at the industrialised world, are available today only on the internet and are not bought like traditional journals. A new publishing model has been devised and they are licensed and not purchased outright. Publishers have set up bundles of journals for sale to try and generate as much revenue as possible. Libraries often gain access to journals they may not know about and which may not be useful to their readers. Publishers have to take into account how easy it is for users to copy material and they have had it declared illegal under copyright law to make these copies beyond a certain number regarded as being 'for fair use'. The photocopier made a difference to publishing as libraries had to set up mechanisms to ensure that the readers did not make copies of articles and sell them, so users can only photocopy items for their own use. Digital material is even easier to copy and so the publishers have strictly defined through licensing what a user can do. Organisations like eifl [2] who claim to be "Working in collaboration with libraries in more than 45 developing and transition in Africa, Asia and Europe, [to] enable access to knowledge for education, learning, research and sustainable community development" have been instrumental in ensuring that organisations in developing countries are able to access this material which is available using a new publishing model. However many journals have back files which may or may not be available only while the current subscription is kept up. Even the wealthiest libraries in the world have been concerned that if a publisher goes out of business or gets taken over by another organisation, libraries which have paid for a license for journals may not in the end get access. To alleviate this, libraries have formed a consortium based at Stanford University in the USA known as LOCKSS, Lots of Copies Keeps Stuff Safe, which obtains permission from the owners of digital data to host the material on multiple servers belonging to the consortium members. Access to this can be activated if a company goes out of business or even if a server goes down for a limited period of time. It could for example be activated even if the internet connection fails between the host and the customer.[3]

2.2 Electronic books

E-books have been around for a number of years but they are only just beginning to take off. They have existed as CD ROMs for some time since publishers do not seem too concerned about their

being copied even though it is not difficult with computers today. At the University of York for example they have available on CD ROM for loan Cambridge Grammar of English, Digital Domesday, Hengwrt Chaucer, History of Parliament, Margaret Thatcher Public Statements, and Pevsners Buildings of England[4]

As compared with e-journals, it has been even more difficult for publishers to come up with a purchasing model which would suit libraries and or individuals. A book is much more valuable as a commodity for a publisher than an article in a journal so a publisher is likely to make much greater effort to ensure it is not freely copied to the extent that it may not be easily available at all. Additionally the e-resource hosts have had to come up with the technology to support the solutions devised. Library book suppliers have become involved since their customers have been pushing them to make e-books available to their users. These suppliers have set up hardware platforms for the storage of e-books. Library users can download these or read them in a web-browser. Either way the user is identified by logging in to the platform. Different library suppliers have a different selection of books on their platforms. Basically the users can be told to access the URL of the suppliers, then log in making themselves known to the platform.

The digital material is usually if not invariably accessed through Adobe. PDF files which are readable by Adobe have incorporated certain features to prevent the files being copied or if they can be stored on a computer they will expire after a certain amount of time, the number of days requested by the user in this case. Incidentally this often requires the library to have the latest version of Adobe mounted on the servers in the library. If you are reading a document through a browser, it is often loaded page by page to the computer (to stop mass copying of a file). When Adobe is set up it can be set up in such a way to disable some features essential for reading digital material of this kind.

In one model, the book is 'virtually' issued to the reader so that no other user can read it. It can be issued for a number of requested days and the platform will usually impose a maximum. A library can purchase more than one copy of a book so more than one reader can read it at once.

Another model again requiring the user's identification allows the library to buy a certain number, say 500, accesses for a year. The library will have to buy 500 again the second and subsequent years. If 500 are exceeded within a year the library can pay for more.

A third model permits records multiple simultaneous 'virtual borrowings' and averages them out. If the access is frequently more than one at any one time the library will be asked to contribute to extra licenses. This requires 'supervision' by the system with a report which currently would be made by manual intervention.

Any of these models may have added to them a facility to enable the user to purchase a copy for themselves. Problems may occur as a user in a library will not want the file to be restricted to the computer at which it was purchased. The user may have to create an account on the server from which the file is downloaded. In practice it appears users can save files to a USB storage device. It also requires the library reader to have a debit or credit card account.

3. Moves Towards open access

Books are usually published by a single author. The publisher is protecting the author's rights by setting up mechanisms to prevent mass copying. The author will receive royalties. However traditionally authors are not paid for journal articles. Publishers have set up infrastructures to enable journal articles to be read. In the past they printed journals and posted them to libraries or individual subscribers. Now they provide internet access and other value added services which are not there in traditional journals such as access to full text which can lead to researchers discovering additional articles of interest to them. The indexes may be available universally across the internet. The actual content will only be available either to subscribers or on an individual charge basis. The publishers do not usually worry too much about individual articles being copied as they make their money from subscriptions. Libraries in many countries have made agreements with publishers to allow library users to copy proportions of a journal for their own use. This is not easy for a library to police.

Newspapers are still treated in this way as people want to buy a physical newspaper to read. Many newspapers made available a digital version alongside the printed making it freely available for publicity purposes. Sometimes it is a reduced version of the published version and may have advertisements which generate extra revenue for the publisher. This is changing as quality reading devices are introduced which enable the user to download the content of newspapers and read them. Some specialist newspapers which are mainly purchased by subscription have provided a digital copy alongside, to subscribers only. The New York Times similarly made a subset of its content freely available with the possibility to subscribe for more. The Times (of London) has just started charging for its digital version as indeed it has always charged for its back files (to defray the cost of digitisation for earlier issues), but current issues are automatically available in digital form.

To return to e-journals, a situation can arise where an author writes an article but cannot allow his students to see a copy because neither the author himself (usually) nor the library at the institution where he teaches (often) subscribes. Authors are usually nowadays offered an electronic equivalent to an off-print for their own use to appease the many authors who would not otherwise see their own articles in print. This copy will usually last for ever but the publishers do their best to ensure it is not copied.

A number of organisations have felt there is an injustice here. Universities and research funders pay for research to be done. Publishers make this available without paying for the privilege and in doing so are alleged to make a handsome profit. E-journal publishers such as Elsevier have made large profits. The British House of Commons Select Committee on Science and Technology summoned Elsevier to answer questions as a result of which they have allowed content to be placed on servers, but not in the published format, sometimes without illustrations. Universities and research institutions have been encouraged to set up servers with free access from the internet for their own researchers' output and place there, with permission, the text of the research.

Various open access initiatives have been formed. At Southampton University Stefan Harnad a professor in electronics and computer science, was perturbed by the situation where academic

work is no longer affordable to the researcher and the publishers are in effect censoring the work by price. He has therefore developed software called eprints and established an organisation eprints.org with its website to distribute the software. This is freely available to anyone to use to set up an electronic library of journal articles. This software was used by 269 libraries worldwide as of December 2010[5].

The UK Joint Information Systems Committee of the UK Higher Education Funding Committee have set up a project called SHERPA[6] which catalogues the permissions that can be given for authors to place the articles they have written on their institution's repositories. Most universities have set up repositories for their academics' articles to be stored in digital form.

4. Library Catalogues and Cataloguing for the Digital Era

Libraries traditionally have purchased books and catalogued them. Journals usually receive one entry in the library catalogue giving details of the holdings. What happens with electronic materials? There can be few libraries in the industrialised world that do not have an online public access catalogue. These catalogues store records which can be imported through the MARC record structure. This is a standard developed in 1966 for the Library of Congress to enable it to circulate electronic records rather than catalogue cards, the original intention being that libraries would produce their own catalogue cards. Catalogues around the world, ranging from OCLC's Worldcat to the smallest special library, can download records from each other's databases. Records are available for digital materials as well. MARC records consist of fields which can be separately manipulated for display in a catalogue. Serials Solutions is a company that can provide records for journals to supplement those catalogued by the cataloguers. Serials Solutions work with electronic journal suppliers such as EBSCO to provide a record of what the library is entitled to[7]. Many journals are, as mentioned above, now sold as bundles and libraries would not have time to catalogue all the journals to which they are entitled without the services of organisations like Serials Solutions. A data element has been added to their MARC catalogue records to hold the URL where the electronic journal or e-book is found. The situation is made more complex by the fact that there may be different routes to electronic material not all of which have been licensed by an individual library. Free journals are included as well. Additionally libraries can add their own e-books which are freely available on the internet. For example, in the UK, government reports used to be published and printed by HMSO (Her Majesty's Stationery Office). When an important report was released people would queue outside their bookshop to purchase the report. Today these reports are available online so the cataloguer in the library needs to create a record for them and link to its location (URL) on the internet. Subject librarians need to work closely with cataloguers to get the material included in the catalogue on time.

These records are available only to those catalogue software packages which have implemented the MARC format. Systems need to be developed in line with standards. Most packages have implemented the MARC standard including the CDS/ISIS family of standards and ABCD which is currently being developed to supplement CDS/ISIS. The MARC standard is itself dependent on other standards. The ISBN comes to mind. ISBN is a 10 or 13 digit number which uniquely identifies any

edition of a book. Records can be selected from external databases using the ISBN to get the exact record. A cataloguer with a book in front of him or her can request the records from external databases using the ISBN. In the case of ISBN used with e-books the booktrade has decided to allocate a separate ISBN to each copy from each different host as the booktrade is using ISBN for rights management, to determine the royalties that should be paid via each supplier. ISSN is often used as a unique in systems for digital journals. Underlying traditional catalogues are cataloguing rules. To the same extent as before, they are necessary in the digital era to provide consistency of access points. Digital materials are too numerous to catalogue in the traditional way and because they are often accessible in full text they do not need so much cataloguing and have been identified by the Dublin Core system as the Dublin Core Metadata element set which again has been adopted first as an American NISO standard and subsequently as an ISO standard.[8], [9] But MARC is the main standard for the storage and exchange of bibliographic records and it in turn derives its data element definitions from the Anglo-American Cataloguing Rules, soon to be superseded by RDA, Resource Description and Access which was published in 2010[10]. MARC records also come with classification numbers included to help index the collection and many records include summaries which can also be indexed.

Potentially every computer could store its data in a different way but in practice systems use the same character sets which have been developed from standards. Libraries needed diacritics long before other IT systems and so there is a suite of character sets for different languages and scripts. Theses have now to a large extent been superseded by UNICODE (ISO 10646).

If your integrated library system does not use these standards it will be difficult to participate in the international information community which all libraries have to be part of today.

5. Setting up the Infrastructure

It is not self-evident how a library must set up its infrastructure for e-materials. If a library has many runs of journals some of which finish abruptly it may be that they finish abruptly because they are now digital and no longer published in print format. A note should be put at the shelves; or students can be persuaded instead to find anything they want by looking in the catalogue first before going to the shelves. Journals and electronic books, some of which may be free-of-charge, may be added by the cataloguer to the catalogue as mentioned in the previous section. Journals and books can be seen physically in a particular location. How can library users know about digital materials which are not physical but virtual? It is necessary to promote their existence in a more active way than it was with printed materials. Training may be required for librarians in this. In some countries there is a tendency for librarians to feel they do not need to be involved with electronic materials. They are found on the internet and only accessible through computers so they are the province of the IT department, or so goes the thinking. IT professionals on the other hand are not specialists in searching for or through the content of electronic materials. They do not wish to be bothered to give out passwords. Librarians are the best people to look after electronic libraries and training is needed for them to learn to set up an infrastructure.

Establishing the infrastructure is not that difficult. The catalogue can and probably should be the basis for it. In the MARC field 856 there should be placed the URL which the licensor of the material being made available will let the library have in order to enable access. On the other hand this can usually be found otherwise by doing a search in Google for the title of the journal or other resource. Some electronic libraries particularly for single monographs require the library user to go into their own website and register themselves there so it will not be possible to go direct from the library catalogue to the resource itself. One such example is the British Library's pilot database of UK theses eThos where the user must be taken to the front page of the database and then make their own search within though downloading an electronic copy is completely free-of-charge. This is so that the British Library knows who has been searching for and downloading theses.

Many electronic resources are freely accessible. These include government publications from around the world, journals which are open access (found through the Directory of Open Access Journals[11], for example). Links to the URL can be made from the catalogue to those which are likely to be of use to the users of the catalogue. Those journals which have commercial publishers will require some kind of authentication. The most common way for this to be done is by IP address authentication. The library registers its IP address range with the publisher and the publisher's computer is aware that any access from the computers with those IP addresses is legitimate. This does not allow users to access from other locations. A further method is by username and password which will allow access from anywhere. Of course there are opportunities of fraudulent use by users offering their passwords to other persons. Additionally there is a system called Shibboleth which in the UK is supported by Eduserv through a system known as Athens. Publishers allow access through this system which allows usernames and passwords identifying a user belonging to an institution to gain access from inside or outside of the IP range. More information was made available about this at the CALIBER Conference in 2009 in Pondicherry in the workshop led by John Paschoud. [12]

Additionally there is a product called EZproxy (pronounced easy proxy) which enables a systems librarian to set up access for all resources requiring IP authentication from outside the registered IP range by means of a proxy server. The URL included in the catalogue includes the address of the proxy server and can be an alternative to the normal URL via IP address.

We mentioned above the need to make access available via the catalogue. This does not preclude additional lists of resources which are available, such as lists of journals or databases in alphabetical order or classified by subject. Of course this will be a necessary way of dealing with access in those universities which do not yet have any integrated library system with a public catalogue. Incidentally these lists of electronic journals and databases can also be provided by companies such as EBSCO and Serials Solutions.

6. Training

Librarians need training to be able to serve their users in the digital era. Setting up the appropriate infrastructure, indeed understanding the digital information topography are best learned by experience. At Middlesex University we have been evolving along with and alongside the information world over

a period of 15 years. We have provided the physical infrastructure and the training for users that is necessary. As a result of having developed our own system, we have been able to invite people from developing countries under the Commonwealth Professional Fellowship scheme to see how we do this and take the knowledge back and share it with colleagues in their own countries[13]. We participated in a project to improve the physical infrastructure of the Yerevan State University library by providing under a European Union funded project servers for a virtual learning environment and training in the use of new technologies. We realised that the most important activity was to train the librarians in using the new technology, after persuading them it was their responsibility and not that of IT staff. This resulted in a further EU-funded project which expanded from Armenia to Georgia and Uzbekistan[14]. Library staff and university lecturers went to Robert Gordon University in Aberdeen to participate in a master's course in library and information science and one of the results of that was that on their return some of the participants from the Ilia State University Library in Georgia produced in their own language a very attractive poster and leaflet promoting the electronic resources that are available to their members.

Libraries today need to provide training in information literacy. There needs to be an induction session for all new students in every university library as to how to access the electronic resources that are available. This can be quite difficult for new students to follow as they have so many other things to learn on arrival, so there must be follow up sessions and tools including documentation on the library website to help students and other members of the university. Library staff of course also need to be trained to be one step ahead of the students. There will be today in most universities in developing countries staff who have studied recently in universities in the industrialised world and they will expect to see the same kind of facilities available. In an earlier TEMPUS project that the author undertook with Robert Gordon University in Syria[15] we set up a website to access the electronic resources which had been negotiated by eifl on behalf of the Ministry of Higher Education in Syria[16]. We held a seminar using the website as a way of introducing information literacy to the librarians and this was attended also by a law lecturer who wanted, for himself, to know what was available as he had found e-resources invaluable during a doctorate in western Europe.

7. Conclusion

Becoming part of the universal digital library but providing an infrastructure to take into account the hybrid library requires a library to become part of a universal digital library. The internet requires a much more sophisticated interface to enable the merging into a library's collection of external digital resources. It also requires the library management system to adhere to standards to enable this interface and requires also a certain level of information literacy in its library staff to develop the same for the users of the library.

Acknowledgements

The projects mentioned in section 5 were funded by the European Commission under the TEMPUS Programme.

1. Dempsey, Lorcan. The (Digital) library environment: ten years after, Ariadne, 46, (February 2006) http://www.ariadne.ac.uk/issue46/
2. eifl. (electronic information for libraries)Enabling access to knowledge in developing and transition countries. http://www.eifl.org/
3. LOCKSS Home, 2008. http://lockss.stanford.edu/lockss/Home
4. University of York Library.CDROM databases on university networked PCs. http://www.york.ac.uk/library/electroniclibrary/cd-romdatabases/
5. Open Access and Institutional Repositories with EPrints. http://www.eprints.org
6. SHERPA Opening access to research. http://www.sherpa.ac.uk/
7. Serials Solutions. 360 Services. Seattle, Serials Solutions, 2010.http://www.serialssolutions.com/360-all/
8. ANSI/NISO Standard Z39.85-2007, Washington, DC, NISO, May 2007 [NISO Z39-85]
9. ISO Standard 15836:2009 Geneva, ISO, February 2009 [ISO15836]
10. Joint Steering Committee for Development of RDA. RDA: Resource, Description and Access Instructions, Washington, DC, ALA; London, Facet Publishing; Ottawa, CLA, 2010,ISBN: 978-185604-749-4
11. Directory of Open Access Journals http://www.doaj.org/
12. Paschoud, John. Access management tutorial. 2009 https://gabriel.lse.ac.uk/twiki/bin/view/Projects/Caliber2009AmTutorial
13. Commonwealth Scholarships Commission. Professional Fellowships: case studies, London, ACU, 2010 http://www.cscuk.org.uk/apply/pfcasestudies.asp
14. Zargaryan, Tigran New Masters Programme in Library and Information Science [website]. Yerevan, National Academy of Sciences Fundamental Scientific Library, 2009- http://www.flib.sci.am/eng/Tempus/index.html
15. Information available at http://www.tempus-nto.org.sy/pdf/4-2-0/2005%20Round1/M030A06.pdf
16. SYReLIB e-library London, Middlesex University, 2007. http://www.lr.mdx.ac.uk/tempus/syria/eresources.htm

About Author

Mr. Alan Hopkinson, Head, Library Services and Bibliographic Division, Middlesex University, London.
E-mail: A.hopkinson@indx.ac.uk