

ONLINE ARCHIVING WITH THE BRITISH LIBRARY – THE EMERALD EXPERIENCE



Catherine Spivey

In October 2001, the international on-line management publisher Emerald (formerly MCB University Press) signed an agreement with the British Library to supply its electronic journals on voluntary deposit. This paper outlines Emerald's initial considerations for this initiative. Emerald is working closely with the British Library to address the joint logistical elements associated with its implementation, including copyright and system-related issues.

Catherine Spivey, Business Development Co-ordinator, Emerald Publishing, 60-62 Toller Lane, Bradford BD8 9BY, UK

Tel: +44 (0) 1274 777700

Fax +44 (0) 1274 785200

E-mail: cspivey@emeraldinsight.com

Background

The British Library has taken an initiative to promote archiving of electronic journals and Emerald is co-operating as a member of the Joint Committee on Voluntary Deposit (JCVD). The JCVD is a committee of representatives from publishers and the legal deposit libraries (LDL), set up to monitor and develop the voluntary deposit of non-print publications as set out in the guidelines of the *Code of Practice for the Voluntary Deposit of Non-print Publications* [September 1999].

The Code of Practice typically concerns media such as CD-ROMs and disks, but, since Emerald no longer publishes all of its content in these forms, we have agreed to deposit our online electronic journals in a format conducive to the outcomes of this ongoing project (Table 1).

What constitutes an e-journal?

In his IFLA address, *Depository, copyright and the notion of a document*, on 20th August 2002, Jon Bing (University of Oslo, Norway) noted that, in the hard-copy arena, there is not a legal requirement to store everything – does this translate into the electronic arena too?¹ Should all material on the Web be looked at as an electronic version of a paper document? What criteria do we apply for depository? Once Emerald began to investigate how we might deposit our e-journal content, the complexities became clear.

Our initial in-house discussions focussed on answering the question 'What constitutes an Emerald e-journal'? Is it simply the e-replica of the hard-copy journal, or do services such as 'key readings', 'reference linking' and 'document delivery' for full-text

Table 1

Annex 1 of the Code of Practice covers the deposit of electronic online material, viz.:

A 1.4. '... It is... very important that...activity should go on jointly between publishers and the legal deposit libraries to carry out trials of the deposit of certain types of online...publications, in order to explore issues, to test deposit and archiving procedures, to identify how the principles of deposit for online publications might need to vary from those for offline publications, and to provide guidance to DCMS to help in the framing of legislation addressing both offline and online publications.'

A 1.5 '... It is recommended that demonstrator projects should be established and carried out on the trial deposit and archiving of, and provision of limited access to, various types of online publications (primarily journals and monographs)... It will be important to carry out projects which address the problems and technical issues relating to publications of a range of types and formats, both those which are relatively straightforward and some which are more complex.'

referenced material not simply add value to our Emerald Fulltext database, but actually form components of the e-journal? Moreover, if this indeed is the case, how could we deposit these elements in an LDL? Branding, browse lists and the search engine are also fundamental elements of Emerald's journals in the e-environment.

Our initial report was deliberately simplistic (Table 2). It identified all the elements of Emerald's e-journals and then indicated which elements are 'industry standard' and which are generated by our Internet Publishing Service Provider (IPSP). Our use of the term 'industry standard' means that either the function itself, or the way we store that element of the data, is an industry standard.

What do we archive?

In May 2002, we met with Andrew Davis, Digital Library Acquisitions Co-ordinator, and his colleagues at the British Library, to explore further the implications of voluntary deposit. Table 2 formed the basis of our discussion.

The British Library is naturally keen to focus on the aims of preservation. These were identified as:

- To preserve the object over time
- To allow for continued functionality
- To keep an historical record of access over time
- To allow for future developments

In order to meet these aims, it is essential to store archive content in ways that are:

- Structural – to be able to view the object and

maintain its accessibility over time

- Descriptive – to enable users to find the e-journals
- Administrative – to control and govern use of the e-journal content

It seems appropriate, therefore, to use the metadata as the common denominator. Since metadata varies from publisher to publisher, the British Library would need to create a standard across the data sets within their repository. Consequently the British Library is trying to encourage standardisation across the publishing industry, for example using ONYX.

IPSP-provided functionality, such as search engines, will also vary from publisher to publisher, and therefore would be difficult to archive. Given the emphasis on preservation and access, the British Library will create a search engine that encompasses all data sets. This would add value to the LDL users. Similarly, using unique object identifiers as a trigger, the British Library/LDL could implement linking functionality to link to referenced material from within the articles themselves. This would ensure that they remain in synchronisation with the latest reference-linking technologies and allow for continued functionality.

Once we had agreed this, we needed to look at the format of the metadata and the requirements for depositing the e-journal content. Would we need to deposit a single copy of the articles, or a copy of the article in all its available formats, e.g. PDF, HTML, SGML and RealPage?

Jon Bing defined this as 'the notion of a

Table 2

Component	Constituent parts	Format available from Emerald	Industry standard?	IPSP generated?	Notes
Journal Information	Notes for Contributors, Editors, Editorial Advisory Board listing, subscription information etc.	HTML	Yes	No	This information is not date-stamped and is updated as required. It is not exported currently and is not exported currently and is therefore not tagged for export purposes. Includes links to other areas of our sites.
Header Information	© information, bibliographic info, DOI reference.	SGML	Yes	No	
Abstract	Abstract text	SGML	Yes	No	
Keyword(s)	Keyword text	SGML	Yes	No	
Content indicators	Research Implications, Practical Implications, Readability and Originality scores	SGML	Yes	No	Only appear for Emerald journals that appear in our Emerald Management Reviews database.
Article Type	Identifies if the article is wholly theoretical or is theoretical with practical implications.	SGML	Yes	No	
Full text	Header and abstract information, Article text, images, references	SGML, HTML PDF's and .tiffs	Yes	RealPage version only	We cannot export the RealPage version.
References	Flat file references	SGML	Flat files only	Reference Linking	Emerald can only export the flat files*
Non-Article Content	Seven umbrella categories – includes Editorial, Book Reviews, Conference Reports, Calls for Papers.	HTML	Yes	SGML - yes	
Key Readings	Related 'keyword' generated abstracts from Emerald Management Reviews	SGML	No	Yes	This is created programmatically, at the time of download, and therefore cannot be exported.
E-WIRE	Related abstracts from Emerald Management Reviews	SGML	No	Yes	This is created programmatically, at the time of download and cannot be exported.
Internet Research Registers	Access database with added data	.asp	No	Yes	This cannot be exported currently.
Search Engine			No	Yes	Emerald does not own this.
Browse Lists			Yes	Yes	Emerald does not own this.
LoC Classifications			Yes	Yes	Emerald does not own this.
Branding	Intangible	Logo only	Yes	Yes	Can only export the logo not the brand experience

*We can include the article's DOI with the SGML, but the technology which drives our value-added services (e.g. Reference Linking), is owned by our IPSP not Emerald.

document'. 'Document' is a term that is neutral to technology. In Norway, the law of depository makes a firm distinction between the medium and the information itself. So, if the information is provided through different media types, the resulting 'document' is unique, and therefore each document type must be deposited.

As a publisher we have already identified that we cannot export the RealPage version since this is created by our IPSP. If this is required, then we will need to rely on our IPSP to provide this content for deposit. The LDL will also have their own concerns such as: What happens if the format (e.g. HTML, Java or XML) is no longer supported? How will the LDL maintain access to that content if the browser has upgraded to the latest protocol? These issues will need to be addressed by the JCVD as the project moves forward.

Costs

Naturally the cost implications are a real issue. Who funds the LDL in their quest to store the e-content? Should there be a time limitation on the data stored, or should it be stored in perpetuity? The server capacity to do the latter would be huge and so, therefore, would be the cost. There will also be costs attributed to storing – and supporting the accessibility of – documents in the various formats.

For instance, Emerald's reference linking functionality – to non-Emerald content – is delivered in conjunction with our IPSP and CrossRef. In very simple terms, our IPSP submits referenced material to CrossRef, and Emerald is charged for every match that is made to full-text articles within the CrossRef system. In the depository environment the end-user risks losing this feature unless the LDL re-creates it. The LDL might therefore wish to consider whether reference linking should be activated solely across their depository database, or whether the LDL should use CrossRef. Who would bear the costs for the latter?

Once e-journal deposit becomes a legal requirement, then one could assume that the national government would be responsible for maintaining the funding for the LDL – but what about the costs to the publisher?

Many publishers do not create their e-journal services themselves. They deposit hard-copy journals with their IPSP who, in turn, creates and displays the e-journal. Such publishers would therefore rely on their IPSP to supply all their e-content to the LDL. This may result in additional expense to the publisher, since the IPSP would probably charge the costs back to them.

Even those who do create their own e-journal content face problems. A concern for Emerald is the way in which our data is currently stored in-house. As can be seen from Table 2, our journal information can be split into four constituent parts:

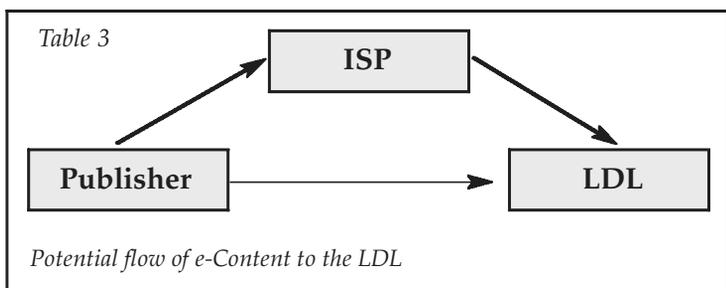
1. Information about the journal itself – Editors, Editorial Advisory Board (EAB) details, Notes for Contributors, subscription information, etc.
2. The full-text articles
3. The non-article content (NAC) – editorials, book reviews, conference reports, etc.
4. IPSP value-added services – reference linking, key readings

As previously mentioned, the elements listed in point 4 are ones that Emerald does not own, nor are they industry standard features. For instance, although we own the abstracts that are used to create key readings, we do not own the search function used to collate and display the abstracts used to fulfil this service. Consequently, it would be extremely difficult to export this data as part of the e-journal issue. Also, since it does not appear within the hard-copy journal, we have decided – for the moment anyway – not to include this in the test data that we have sent to the British Library.

In terms of the way we store elements listed in points 1-3:

- Journal information is created in-house and displayed as HTML on our website. It is not exported, nor is it time-stamped, so the date of any amendment (e.g. changes to EAB membership) is not logged in the HTML version.

Table 3



- NAC is also stored as HTML on our system, but this is exported to our IPSP, who then converts it to SGML and displays it on the website. This is time-sensitive because new NAC is added for each new issue, although the previous issue's NAC is naturally retained in the archive.
- Full-text articles are stored on our SGML database, exported to our IPSP, and then they use their own version of SGML to display it on the Emerald Fulltext website. Again, this is time-stamped according to the date that the issue is produced.

Emerald both stores and exports each of these separately, and our IPSP marries the information together on the web. When it comes to depositing our e-journal content we are therefore faced with logistical problems. One dilemma is how could we combine this data ourselves, to ensure that a true copy of the e-journal is provided to the LDL? There may be considerable costs if we are to develop our systems to improve the way in which our data are currently stored. Alternatively, we could pay our IPSP to create a full export of all elements. For test purposes, we have been able to export these individually as a one-time request. However, as the project moves forward, we will need to work closely with the British Library to work out a strategy for continual feeds.

In addition to the direct costs – developing our in-house systems for exporting all data-elements, possible IPSP charges for exporting data, resource issues – there are, of course, indirect costs associated with the exploration of the impact of voluntary e-deposit on our internal production systems. At present, there are four Emerald staff on the task force, all of whom are working with the British Library to address the needs of this project. The Electronic Publishing Services Ltd (EPS) are working with the Joint Committee for Voluntary Deposit (JCVD) to, 'assess the cost, both to publishers and to the LDL, of extending statutory deposit obligations to non-print publications, such as electronic publications and microforms' (EPS Press Release, 15 July 2002).

Copyright and legal deposit

The affects of copyright on legal deposit are another issue of concern for publishers and the author community. There is a fundamental

difference between the notion of a library and the notion of a depository. In very basic terms, the former is there for accessibility, the latter for preservation.

Content acquired for legal deposit would not be used in the same way that a library would use a purchased subscription. Consequently, it has been suggested that there needs to be a statutory licence to copy for the purpose of legal deposit only. This is not a sold copy and it cannot, therefore, be made available to multiple users on a network, but it is available via single PCs in reading rooms with no output facilities. Copyright remains with the copyright holder in all instances.

However, publishers can only deposit content for which they either own the copyright or hold an exclusive licence from the author. Not all elements of a journal are copyright assigned and therefore we would be unable to deposit this content with an LDL. This makes nonsense of the LDL's purpose of preservation, potentially resulting in incomplete journal issues where the author has retained copyright for his articles. This does not meet the needs of the LDL, the author, the publisher or the scholarly community. For instance, an editorial is an integral part of a journal's content – especially where that issue is themed. Historically, this content has been produced under 'work for hire'. It is assumed that the copyright is therefore assigned to Emerald. Yet, until 2003, this has not been explicit in our editorial contracts.

In the hard-copy environment, legal deposit is a statutory obligation for the publisher, but the printer may also have played a role in servicing that requirement. Therefore, it is worth considering the extent to which IPSPs play a role in electronic preservation. They do not own the copyright for the material, but, for some publishers, they are the means for the creation of the e-journal.

What can others learn from our experience?

Our experience so far has highlighted a number of considerations for our publishing process. These may be of relevance to other publishers, too, for instance:

- Do you own copyright in all facets of your e-journal content?
- If not, could you export only that content for which copyright is owned?

- If journal information is stored across multiple systems, could you export this in one deposit to an LDL, and would you have to develop your in-house systems to do this?
- Could you export all formats of your e-content, or would you have to request this from your IPSP?
- Would you need to change your author and editorial contracts to ensure that export for the purpose of legal deposit is allowed across all content?
- How will this affect the archive files?

On 20 August 2002, the IFLA seminar, *National Libraries with CLM*, focused on requirements for the legal deposit of e-publications. The speakers were drawn from an international spectrum, with a view to examining the various problems faced through this initiative. An overview of how South Africa is tackling the requirements for legal deposit is summarised as an addendum to this article.²

Where are we now?

In May this year, Emerald exported sample e-journal data to The British Library. Andrew Davis and his team at The British Library are currently comparing this against other publishers' data, to try and find commonality in the meta-tags. They aim to design an import procedure that will facilitate the creation of a repository for e-journal content and, ultimately, access in perpetuity to those journals.

Emerald are also looking into what internal changes will need to be made to the way we host our data on our internal systems. We will continue to work closely with the British Library to ensure that the *Guidelines of the Code of Practice for Voluntary Deposit of Non-print Publications* are met.

Summary

This project has proved to be a shared learning experience between Emerald and the British Library, and especially within the Emerald team itself. It has highlighted a number of discrepancies within our own publishing process,

particularly in relation to the technical creation of our content and its dissemination. Wider implications include:

- The impact that copyright law may have on the degree to which we, as publishers, can honour the requirements for e-deposit;
- The extent to which the publishing industry relies on third parties (e.g. IPSPs, CrossRef) for the provision of additional functionality in the online environment;
- The costs associated with our involvement in this project.

It has been particularly useful to be involved in the project from the outset, so that these considerations are explored and, hopefully, resolved, before the requirement for "voluntary deposit" is legalised. The work of the JCVD and EPS will also prove to be invaluable to the future success of e-journal deposit.

Resources

Joint Committee on Voluntary Deposit:

<http://www.bl.uk/about/policies/codeprac.html>

Electronic Publishing Services Ltd:

<http://www.epsltd.com>

The Digital Library Federation:

<http://www.diglib.org>

The Mellon Foundation: <http://www.mellon.org/>

Further information about the Mellon Foundation's e-journal archiving program is available at:

<http://www.diglib.org/preserve/ejp.htm>.

Version 1.2 (2000) of their guidelines for deposit are available at:

<http://www.diglib.org/preserve/criteria.htm>

Digital Preservation Coalition (DPC):

<http://www.dcponline.org>

References

1. Bing, Jon. Depository, copyright and the notion of a "document". IFLA address, Glasgow, 2002.
2. Letshela, Phegello Z. and Peter J. Lor. Implementing legal deposit of electronic publications in Africa: progress report from South Africa and Namibia. IFLA address, Glasgow, 2002.

Addendum

A Comparison: The African Experience

Phegello Z. Letshela and Peter J Lor, National Library of South Africa, in their IFLA presentation, *Implementing legal deposit of electronic publications in Africa: Progress report from South Africa and Namibia* (20 August 2002), highlighted the fact that countries in Africa are still struggling with the challenges of e-publishing. Most content that is published electronically is lost and there is a need for more comprehensive Legal Deposit Acts in this region.

In Namibia, for example, there was no legal deposit act prior to 1996. The 1996 Act borrowed heavily on the Norwegian Act in its creation. South Africa's 1842 Legal Deposit Act was also recently revised in 1994 to include electronic content. This act caters for content that was either created in South Africa or created elsewhere for specific distribution in South Africa.

In Nigeria there are 25 LDL and their law covers all published material. However, they are not as far ahead with the deposit of e-content.