Archiving E-Journals Consultancy - Final Report

Report Commissioned by the Joint
Information Systems Committee (JISC)

Maggie Jones

October 2003

CONTENTS

Executive Summary

The National Electronic Site Licensing Initiative (NESLI) encouraged widespread take-up of electronic access. UK universities are increasingly moving to e-only access to e-journals through licensing content. Publishers are also increasing their e-journal content. However the implications of licensing access, as opposed to owning content, means that there needs to be procedures in place to ensure that libraries are assured of continued access to content they have paid for, even following the termination of a licence.

The Model Licence has provided some reassurance, through the introduction of archiving clauses. It is now timely to explore options for implementing those clauses. Although libraries can see many benefits in moving to electronic access, and regard parallel print and electronic as unsustainable, concerns about VAT and archiving are two frequently cited obstacles in making the move to e-only. Publishers have never been expected to undertake a preservation role in the print environment, and the diversity of publishers and publishing means that there are inevitably many different interpretations of the archiving clause in the Model Licence.

This problem is of course by no means confined to the UK and efforts to address the issues associated with ensuring continued access to licensed commercial content have been taking place. In the U.S, seven projects focussing solely on scholarly e-journals were funded by the Mellon Foundation. These concluded that new organisations will be necessary to mediate between libraries and publishers. Two very different approaches (LOCKSS and JSTOR) have now been funded for 18 months to further develop their work.

Services, such as OCLC Digital Archive have also been developed and offer some prospects for trusted third party services for the U.K. Developments within the UK, such as anticipated legal deposit legislation and its potential implications for licensed UK e-journals also needs to be taken into account.

It is premature to speculate to what extent these initiatives and developments can assist UK universities but it is clear that there is strong potential for them to play a role. Related developments, such as e-print repositories, are not expected to offer a replacement to licensed e-journals but their parallel development needs to be monitored closely.

Although there is no imminent danger of loss of content to licensed e-journals, there is an urgent need to provide a co-ordinating archiving service for the UK, which can develop in stages. Beginning with a dedicated resource to act as a central liaison between publishers and libraries, and to explore viable options for ensuring continued access to licensed content. This will provide the necessary reassurance to UK libraries that archiving of e-journals content will be followed up on their behalf, without committing potentially vast sums on establishing a UK repository.

Acknowledgements

This study was made possible through funding provided by the Joint Information Systems Committee (JISC) through its Digital Preservation and Records Management Programme. The JISC Journals Working Group (JJWG) provided valuable contextual information and guidance throughout the project. The co-operation of libraries and publishers in the surveys and follow-up interviews added greatly to the report. Attendees of the Workshop gave their time generously and were able to develop a way forward in what is still a very complex area in which much activity is going on worldwide but there is little which is sufficiently developed to offer a clear option.

1 Background to the Consultancy

1.1 Concerns about future access

The long-term archiving of scholarly publications is a key concern of many academic authors and customers of Scientific, Technical, and Medical (STM) publishers. Subscribers to e-journals, particularly academic libraries whose core business depends on being able to cite and ensure future access to published works, are also deeply concerned with ensuring provision for continuing access to licensed publications.

1.2 JISC involvement

JISC has been involved with national site licensing of journals since 1995, beginning with the Pilot Site Licensing Initiative (PSLI). This was followed by the National Electronic Site Licensing Initiative (NESLI) which, as the name suggests, focussed on encouraging widespread access to electronic journal content negotiated with publishers via a managing agent on behalf of Higher and Further Education Institutions. During the third phase of JISC journal negotiations JISC staff undertook negotiations with publishers but an ITT sent out during 2003 will appoint a negotiating agent for 2004 journals deals and beyond. As a key funding body for both the provision and development of digital content for Higher and Further Education in the UK, JISC has a critical role to play in its long-term preservation and access, in collaboration with a number of partners.

With the conclusion of the Cedars project, JISC funded a consultancy to evaluate previous licences; explore with publishers and other stakeholders archiving and access provisions; and evaluate future options for archiving of licensed e-journals and access arrangements. In the first instance it was suggested that CURL would be invited to utilise existing expertise built up in the Cedars project and extend the contract for the project manager for 12 months to complete the consultancy. Maggie Jones was subsequently appointed to undertake the study between May 2002-April 2003.

1.3 The Cedars project

JISC has funded CURL through the Cedars project to address fundamental research issues in this area and Cedars has been particularly active and made a substantial contribution to developing work on preservation in digital libraries.

The Cedars Project[1] was undertaken between 1998-2002 and it identified the need to work closely with publishers at a very early stage of the project. In July 1999, Cedars hosted a Publishers seminar, to provide an opportunity to build a good relationship between the digital preservation aims of the Cedars project and publishers. Some publishers contributed to the project by providing material to be used to test the Cedars Digital Archiving Prototype.

The final year of the Cedars project focussed on the production of five Guides covering various aspects of the work of Cedars. *The Cedars Guide to Intellectual Property Rights* referred to "the need for clear communication between libraries and publishers in clarifying rights and responsibilities connected to ongoing maintenance of the digital information."[2]

1.4 The Model Licence

There are two interwoven issues associated with licensing of e-journals:

- The need for librarians to have assurance of continued access to material they have paid for under licence.
- The need to have assurance of long-term preservation for all scholarly digital resources, including e-journals.

The current Model Licence[3] attempts to accommodate both of these requirements by the incorporation of three clauses. Clause 2.2.2 refers to the need for libraries to have assurance of continued access to material they have paid for, following termination of their licence. Options suggested are the Publisher's server or from a mutually acceptable third party; an archival copy delivered to the Licensee; or a central archiving facility operated on behalf of UK HE. Clause 5.4.1 refers to the need for an archive for licensed material and 5.4.2 reiterates the need for continued access following termination, as specified in 2.2.2.

The assumption behind these clauses is that legal, commercial, and technical issues will dictate that the initial responsibility for ensuring retrospective access to licensed material must lie with the publisher, regardless of where a long-term digital repository will reside. It is also important for libraries to be clear on what their payment covers. The assumption in the Model Licence is that they are entitled to expect permanent access, as was the case when they purchased subscriptions to printed journals. It is therefore envisaged that a collaborative arrangement between publishers and libraries will need to exist to ensure that licensed content will be available not only during the period of subscription, but also beyond.

1.5 The role of the Publisher

It is acknowledged that long-term digital preservation cannot reasonably be expected to fall to the publishers and it is therefore a high priority to develop strategies and recommendations for implementing the recommendations referred to in the Model licence.

Recently, the DTI report on *Publishing in the Knowledge Economy* recommended on behalf of publishers and customers that:

> '…in order for the UK to protect access to important research material and to ensure that small and not-for-profit publishers are not unfairly disadvantaged, the archiving of digital research should be organised at a national level by Government. A content repository together with

access rights should be created by Government and industry for use by academic and research institutions.'[4]

1.6 Library/Publisher Collaborations

Seven e-journal archiving pilot studies[5] were funded by the Mellon Foundation in US universities for one year during 2001-2002 and a number of potential business models were explored with publishers during this planning phase. Two very different models (LOCKSS and JSTOR) have subsequently been funded by Mellon to build on the work of the earlier studies.

The Digital Preservation Coalition was launched last year and copyright libraries and publishers together with JISC are involved in its activities. Publishers are becoming more aware of the issues and are increasingly involved in collaborations to explore potential solutions for long-term archiving. A joint ALPSP/DPC Seminar entitled Archiving – Whose Problem Is It?[6] was held in November 2002 and brought together representatives from JISC, libraries, and publishers to discuss archiving from a range of perspectives.

1.7 Legal Deposit

The legal deposit libraries are exploring voluntary legal deposit (primarily of offline material) with publishers and legislation for electronic legal deposit is anticipated. However legal deposit of electronic materials will not provide a solution to all archiving issues. Many publications from international publishers will not be covered by legislation and continuing access by institutions in line with licensing agreements would need to be resolved. However there are clearly potential synergies that could be explored.

2 Methodology

The consultancy was undertaken in three major stages. Initial desk-top research was then complemented by interviews and surveys. Finally an invitation-only Workshop was held to bring together the main results of the research with key stakeholders.

2.1 Desk-top Research

Literature searches were conducted to investigate related developments worldwide and revealed many useful sources of information. They have been compiled as an annotated bibliography at Appendix A.

Archiving implications and arrangements of the PSLI and NESLI schemes were researched through appropriate JISC files, discussions with JISC Collections Team, and analysis of the 2002 NESLI licences.

2.2 Interviews and Surveys

Interviews were arranged with others undertaking relevant work in the area of preserving digital information.

At the British Library, a meeting was held with Andrew Braid, John Byford, Neil Smith, and Deborah Woodyard, to discuss the British Library's potential role in archiving e-journals, in the context of anticipated legal deposit legislation.

A presentation on the e-journals consultancy was given at the JISC/RLG Symposium on Collaboration and Selection for Digital Preservation , held in Washington D.C, which also offered the opportunity of a meeting with JSTOR's Kevin Guthrie and Eileen Fenton to talk through their plans to develop JSTOR's electronic archiving service.

A telephone meeting initiated by Vicky Reich, from LOCKSS also provided the opportunity to discuss the potential role this initiative might make to preserving licensed e-journals in the U.K.

Meg Bellinger of OCLC provided detailed answers to questions about OCLC's Digital Archiving Service.

Contact was also made with Christine Mahler, who had helped prepare a report on archiving of licensed e-journals for the Council of Australian University Libraries (CAUL) and revealed similar concerns within Australian HE as the U.K. There was also some email correspondence with Leacy 0'Callaghan regarding the Canadian National Site Licensing Project.

Further opportunities for networking and information gathering were gained from attendance at the Ingenta Institute's *The Consortium Site Licence – Is It A Sustainable Model?* And the ALPSP/DCP Seminar *Archiving – Whose Problem Is It?*

A related project, funded by the Arts and Humanities Board, was announced during the course of the Study. The eighteen-month Copyright and Licensing for Digital Preservation project commenced in September 2002 and, while there has been only a brief overlap in terms of timing, it has been useful to maintain contact with the project and Maggie Jones was invited to be a member of its Advisory Board. The objectives of the latter project are somewhat broader, covering copyright legislation, as well as licensing and covering a range of digital materials, as well as e-journals.

The common concerns and transferable lessons from the literature searches and interviews are discussed in more detail in Chapter 3.

Further detail was sought through surveys designed to clarify both libraries and publishers perceptions and experiences. The results of these surveys are provided as Appendix B and C, respectively. Follow up interviews with the following provided added insights and observations:

Julia Burke (Bangor); Ian Bannerman and Nancy Gerry (Blackwells); Louise Cole (Leeds); Stephen Drake (Ulster); Karen Jeger (UCL); Jill Lambert (Aston); Mark Mandelbaum (ACM); Marion Miller (Glasgow); Julian Richards (*Internet Archaeology*); Jill Taylor-Roe (Newcastle); Bridget Towler (Hull)

2.3 Workshop

An invitation-only Workshop held on 17th February was the final major milestone in the project and had the objective of exploring legal and organisational issues and business models and reaching consensus on the best way forward. A report of the Workshop and the list of attendees is at Appendix D.

3 Findings and Recommendations

3.1 The Problem

> "Without trusted electronic archives, it is unlikely that e-journals can substitute for print and serve as the copy of record, and so we have a duplicative and even more costly system – a system we all hope is transitional."[7]

This deceptively simple statement reveals a fundamental dilemma facing both libraries and publishers in deciding to move to e-only access. If it is accepted that maintaining a parallel print copy (even assuming the electronic and print versions are identical) is not sustainable, then the question remains, how can the electronic version be relied upon? Given the substantial technological, legal and organisational challenges involved in guaranteeing continued access over time, how can publishers be expected to provide such a guarantee, when they have never been expected to undertake such a role in the print environment? How can libraries be assured that the journals they are licensing access to will be available into the future?

Costs are not yet well understood, particularly in terms of large-scale digital preservation programmes, but are assumed to be substantial. Key questions posed by the Mellon projects included issues of incentive. Where is the incentive for publishers to invest in trusted digital repositories for their content, which will ensure its access beyond its commercial benefit? On the other hand, if one assumes libraries should continue to take on a preservation role, where is the incentive for them to invest in a digital repository on behalf of publishers, if the latter reaps all the commercial benefits while the former accepts all the risks?

Roles and responsibilities are not nearly as well defined as they are in the print environment. This is particularly true of licensed e-journals, where libraries licence access to the content which is under the control of publishers. For the most part, publishers are assuming an archiving role by default, at least for the short to medium term. Improved arrangements are needed for both libraries and publishers so that all can benefit from the assurance of continued access to valuable digital content and the market can continue to develop.

3.2 Journal Negotiation

Journal negotiations co-ordinated by JISC and undertaken on behalf of UK universities have encouraged use of the Model Licence, including the archiving clauses. The NESLI deals were undertaken by a Managing Agent. JISC staff then undertook negotiations on behalf of the community but problems were encountered in reaching acceptable deals with publishers in a timely manner. Furthermore, JISC journal negotiation operates as a loose consortia, and it was time consuming and complex to negotiate deals needed by the community and also encourage adherence to the Model Licence.

The third phase of the JISC journals negotiation will involve negotiating agent for 2004 subscriptions and beyond. The agent will negotiate ten deals based

primarily on data from the 2002 National Serial Requirement Survey[8], following an in principle commitment from the community. It is hoped that this strategy will enable a more rapid negotiation process and deliver content to the community which has already been identified as high priority.

Another critical aspect of journal negotiation is in ensuring regular feedback to the community, to ensure they are kept informed of the progress of deals. It is clear from comments on lis-nesli email lists, and also from the library survey, that there is concern to be kept informed of the progress of deals and also that they adhere to the Model Licence (and if not, why not).

The allocation of a dedicated resource to undertake complex negotiations and achieve the best value for money in a timely and efficient manner will bring benefits to both publishers and UK universities by providing a clear line of communication between the two. However, experience has already shown that it sometimes proves impossible to reach agreement on all aspects of the Model Licence, including the archiving clauses and sometimes it has been necessary to agree to deals which are less than optimum.

It is also clear that a ripple of concern is felt by UK universities when commitments to assured ongoing access are unclear because the Model Licence has not been used or has been significantly modified. There is also a need for a dedicated resource to work in partnership with the Journals negotiator, focussing on the archiving aspects of the Model Licence and clarifying arrangements in place for ongoing access. This is not a task the Journal Negotiator will be able to undertake as their primary responsibility will be the timely delivery of content required for the current year.

Recommendation 1:
That JISC allocate resources for a consultancy to focus on the longer term access and archiving provisions of journal deals.

3.3 Past National Licensing

The Pilot Site Licensing Initiative began in 1996. The original primary objective was to reduce the spiralling costs of print journals. E-journals were initially seen as a by-product of this but rapidly assumed a much more central issue. When the PSLI ended in 1998, the National Electronic Site Licensing Initiative (NESLI) replaced it and, as the name suggests, focussed on encouraging the widespread adoption of e-journals.

An analysis of the 2002 NESLI deals revealed a great deal of variation in interpretation and use of the archiving clauses of the Model Licence. Any amendments appear to be relatively minor. However there appears to be considerable variation between publishers on which of the three options they are likely to prefer following termination of a licence, whether or not they have amended the archiving clauses, so even an apparently standardised licence is still subject to varied interpretations and practices.

Given the overall complexity of many of the deals, and the overriding of previous licences by the current licence, it would be incredibly difficult for individual libraries to be absolutely clear what their rights were, much less follow them up.

A small number of publishers dominate the NESLI licences. The total Number of titles licensed in JISC/NESLI deals for 2002 was 5025 from 17 publishers. 4038 of these (or 80%) were from 6 publishers (Elsevier; Blackwells; Springer; Kluwer; Taylor & Francis; Wiley).

Not surprisingly perhaps, there is some overlap here with data from both the Morgan Stanley Report. *Scientific Publishing: Knowledge is Power* [9]and The National Serial Requirement Survey Report[10] conducted for JISC. The former indicated 2001 major global market players as Elsevier, ACS, Thomson, Wiley, IEEE, Kluwer, McGraw Hill, Taylor and Francis, and Springer. The NSRS report showed a similar result, with the Top 10 publishers both independently cited and cited from journal articles being Elsevier (including Academic Press); Blackwells; Wiley; Taylor & Francis ; OUP; Springer; CUP; Kluwer; Sage; ACS/Nature.

3.4 The Model Licence

The Model Licence was developed as a collaboration between libraries and publishers to assist in clarity and understanding of the rights and obligations of both parties. Since 1999, three clauses have been incorporated which relate specifically to arrangements for future access in the event of the licence being terminated. The major clause is as follows:

2.2.2 After termination of this Licence, the Publisher will provide the Licensee and its Authorised and Walk-in users with access to the full text of the Licensed material which was published and paid for within the subscription period, either by continuing online access to the same material on the Publisher's server or by supplying an archival copy in an electronic medium mutually agreed between the parties which will be delivered to the Licensee, or to a central archiving facility operated on behalf of the UK HE community without charge. Continuing archival access is subject to the terms and conditions of the use of this Licence.

A difficulty here is that it is often uneconomic, or technically difficult for individual institutions to undertake responsibility for maintaining a physical copy of any e-journal content they have paid for, if supplied on CD-ROM. As the central facility does not yet exist, this leaves only continued online access from the publishers site. However most publishers are not able to guarantee continuing access indefinitely from their own servers given the risk inherent in any commercial enterprise.

After several years in operation, it is perhaps timely to re-visit the archiving clauses of the Model Licence, and explore with publishers what mutually acceptable arrangements and conditions for implementing these clauses are possible.

Recommendation 2:
That JISC facilitate ongoing discussions with publishers about mutually acceptable arrangements for implementation of the archiving clauses of the Model Licence.

3.5 Views of UK universities

The views of UK universities were sought from a survey sent to 26 UK HE/FE institutions involved in NESLI deals during 2002. Nineteen responses were received and nine of these agreed to follow-up interviews. Despite an increasing trend towards e-only access, benefits of e-access, and a view that parallel print and electronic journals being unsustainable, there was some unease about moving to e-only, expressed by both libraries and some of their users.

Concerns about archiving and VAT were the two issues most frequently cited by respondents as inhibiting the move to e-only access. In terms of archiving provisions, survey respondents had concerns about either leaving this responsibility entirely to the publisher, or undertaking it themselves, through receipt of an "archival copy". The following quote from the JISC Continuing Access and Digital Preservation Strategy provides a good summary of the views of the survey respondents.

> Digital preservation represents a complex set of challenges, which are exceptionally difficult for institutions to address individually. National action in this field is therefore appropriate to the community and UK wide remit and mission of the JISC[11].

Guaranteeing continued access to licensed e-journals was felt to be an issue more logically and efficiently dealt with at the national level, rather than at the individual institution level, and also by a trusted third party. This is not to say that distributed models, such as LOCKSS (which is discussed in more detail below), are not worth pursuing. However the random, distributed effort involved in individual libraries investing stretched resources into costly, duplicative procedures to preserve licensed e-journal content, (as opposed to the controlled, consortial replication offered by LOCKSS), is not the way forward.

A further source of concern was that the "archival copy" provided by publishers to libraries was almost invariably CD-ROM, and, as noted earlier, many libraries are making a strategic move away from these. The generally resource intensive nature of providing access from them causes difficulties for libraries. Indeed some survey respondents indicated that they would prefer to rely on print and/or interlibrary loan in preference to a CD copy of their licensed content.

Concern was also expressed at the option for the Publisher to undertake archiving responsibility. Some respondents felt it was dangerous to put so much control in the hands of a single party, others felt that only the larger publishers were likely to have the necessary infrastructure. In any case, all publishers are subject to the vagaries of market forces and the implications of mergers and bankruptcy were also noted.

The Model Licence was viewed very positively but some noted that there are no procedures in place to implement the options in the archiving clauses.

There was a clear preference for the central archiving facility foreshadowed in the Model Licence. A few noted that the central service referred to in the

Model Licence had not yet appeared; though there was acknowledgment of the complexities involved in guaranteeing continued access.

The need for a reliable and trusted archive was at least implicit in much of the discussions with libraries and the report from RLG and OCLC on Trusted Digital Repositories[12] is relevant in this context. A Task Force[13] has been established to make progress on the report and this work. JISC is represented on the Task Force.

A summary report of the libraries survey is at Appendix C.

3.6 Views of Publishers

It proved to be extremely difficult to elicit a good response rate from publishers, though the encouragement of the Publisher's Association provided an additional two responses and two further publishers were invited to participate in either a face-to-face, or 'phone meeting, which yielded valuable additional data. In all, data was gathered from seven publishers (out of an initial seventeen contacted). There could be several reasons for the difficulties in gaining a good response rate. Competing demands on time make surveys unattractive for everyone and one commentator suggested publishers have a particular dislike of them. It was also difficult to be certain of the best contact person, particularly within large publishing firms, when the NESLI representative may not have been the most appropriate person to discuss archiving arrangements with.

Despite these difficulties, the resulting collation of results at Appendix D was valuable and strongly reinforced the need to encourage ongoing constructive dialogue with publishers on these issues. This is particularly so as both the analysis of NESLI licences, and the survey results and meetings revealed a lot of disparities between publishers, not just in terms of their size and business models but also in their interest and understanding of archiving issues.

Additional useful publisher contacts and views were gained through attendance of the Ingenta Institute's Seminar on national site licensing and the ALPSP/DPC Seminar on archiving. Publisher representatives were also invited to the JISC Workshop held as part of this consultancy and this provided a fairly evenly balanced distribution of library/publisher representation.

Recommendation 3:
Ensure ongoing and constructive dialogue between libraries and publishers by developing relationships with selected publishers included in JISC deals.

Recommendation 4
Ensure ongoing collaboration and partnership between libraries and publishers through a range of mechanisms, including regular meetings, publicity, and further joint forums to address archiving issues.

3.7 Volumes, Formats, and Risks

The data gathered in the publishers survey and analyses of market trends all point to a rapidly escalating increase in the quantity of electronic journals, already evidenced over the past five years and likely to continue at the same steep rise over the next few years.

In addition, the number of formats is also increasing, as evidenced in the small survey of publishers shown in Appendix D. Very few of the journals produced by publishers are e-only at this stage but this percentage is likely to rise rapidly. The "safety net" of a print copy, regarded as unsustainable in any case, will be removed entirely over time. There are trends towards increasing numbers of e-only titles by publishers but a parallel trend is that the print and e-versions, where they exist, are unlikely to be completely identical as the capabilities of the technology to provide greatly enhanced capabilities are exploited. The small survey of publishers reported at Appendix D indicates a multiplicity of formats being used.

This is not to say that there is imminent danger that content will be lost. It is clear that publishers are beginning to invest significantly in making more and more of their content available online, in response to market demand. The commercial advantage these strategies provide makes it highly unlikely that valuable digital assets will be in danger at least for the short to medium term.

The urgency is not so much in imminent danger of loss of content but in providing assurance to libraries on continuing access. JISC should continue to take a leadership role in terms of providing a dedicated resource to further research, planning, and dialogue with relevant stakeholders. Neither individual publishers nor individual libraries can be expected to undertake this role.

Recommendation 5

Maintain an active watching brief on initiatives and services being developed to provide archiving of licensed e-journals by trusted third parties.

3.8 Organisational Models

In terms of architecture, the JISC Digital Preservation Strategy proposes a modification of the important OAIS Reference Model[14] for use in the JISC Information Environment. The OAIS Reference Model focuses solely on archival repositories. The proposed JISC adaptation of the model can accommodate a wider range of repositories, including hosting services, publisher services, and preservation archives and their inter-relationships. This allows for scenarios, such as escrow arrangements with preservation

archives which might be particularly suitable for commercial materials, like licensed e-journals.

The proposed architecture is seen as being particularly applicable to third party services being developed to support preservation planning or remote storage. It may be one mechanism for implementing continued access for commercial materials.

Two broad possibilities are discussed for implementing the JISC adaptation of the OAIS model in the Strategy:

- The OAIS archive undertaking full preservation responsibilities. Storage would be the Archival Storage function in OAIS undertaken according to preservation standards. The remote storage in this case would be for the off-site back-up copy(s): this is considered a best practice employed in JISC archiving services. Preservation planning would encompass all functions required to ensure the accessibility of the content over a long period of time.
- A publisher/hosting service. Ingest, storage, and access represent the normal business functions rather than archival functions in the OAIS model. Remote storage/escrow would represent a third party guaranteeing preservation and continuing access on behalf of the publisher and customer.

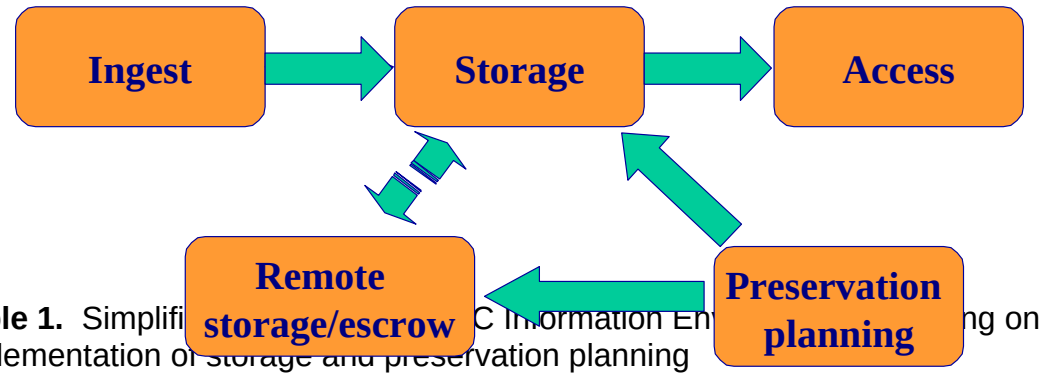

Table 1. Simplifi C Information Env ng on implementation of storage and preservation planning

[A Continuing Access and Digital Preservation Strategy for the JISC 2002-2005]

One current example of the model above is the KB/Elsevier agreement described below. However, it can be implemented in a number of ways, with trusted third party service providers guaranteeing longer term access. Current access can be from the publisher, and this may well be the preferred delivery mechanism, as noted at the JISC Workshop, but the key thing is to ensure there are appropriate arrangements in place for the longer term.

Recommendation 6
Explore the range of options for implementing the JISC model for individual journal deals.

3.9 Related Initiatives and Projects

The British Library

The British Library is clearly going to be a major player in any national preservation model for the U.K and some survey respondents noted their potential role in helping to preserve licensed e-journals. Planning and preparation has been undertaken by the BL for the national role they will play in preserving digital materials and is now focussed around four categories. These are voluntary/legal deposit; UK websites; purchased digital materials, and digitisation. For the purposes of this report, deposited content is the primary focus.

The Voluntary Deposit Scheme has been overseen by the Joint Committee for Voluntary Deposit (JCVD), which has representatives from publishers and voluntary deposit libraries. This has been seen as a precursor to anticipated legal deposit legislation and a means of testing the practical implications of legally deposited digital publications. The Voluntary Deposit scheme has involved predominantly physical, or "hand held" formats and an estimated 90,000 items have been deposited to date. A number of issues have been addressed during this planning period. Two key and overlapping issues are:

- The need to assure publishers that their business models will not be undermined by legal deposit. A secure network has been built to ensure access can be restricted to deposit libraries.
- The need to understand more about cost implications for both publishers and deposit libraries. A report commissioned by JCVD published its findings in October 2002[15] and contains a wealth of valuable data. Significantly, it found that publishers were relatively unconcerned by delivery costs for legal deposit, which were seen as minor, but were much more worried about access/usage, metadata and rights issues which they believe have the potential to add significant costs for them.

A meeting, as part of the Study, was held at the British Library with BL staff in August 2002 revealed that, for the moment, while e-serials are largely duplicated by print, the print copy is regarded as the preservation copy and only the print copy is deposited and preserved (although in some cases the BL may .licence access to electronic journals as part of their other services). This will not of course be viable as the trend towards electronic publications which do not precisely mirror their print counterparts and also the trend towards e-only continues.

Clearly, any archiving service established for e-journals licensed by UK HE/FE will need to ensure that there are strong communication links with the British Library and other UK Copyright libraries. There are also issues of access rights to licensed material taken in under deposit, and any additional roles would be subject to negotiation between the deposit libraries and publishers (e.g. the KB/Elsevier agreement). Finally, it needs to be understood to what extent there is likely to be overlap between titles licensed by UKHE/FE (which of course include material from all over the world) and titles deposited at the BL, which will be UK-based.

KB/Elsevier Agreement

In August 2002, the Royal Dutch National Library (KB) and Elsevier announced an agreement [16]that the KB would become the first of Elsevier's official archive sites. This is an interesting development in that it evokes the traditional role of the library, particularly of national libraries, in undertaking

preservation responsibility, while also asserting the commercial role of the publisher. It potentially provides the assurance and trust required by libraries that the content they are licensing will be retained over time, while respecting the commercial rights of the publisher. By implication, this allows each partner to focus on their particular strengths. In the case of libraries, to focus on preservation, for publishers, to further develop the capabilities of their product in line with customer demand.

Funding has been provided by the Dutch Ministry of Culture, through their financing of the KB. Access will be limited to onsite users, similar to arrangements traditionally in place for deposited materials, with the KB permitted to undertake an interim service role in the event of catastrophic disaster rendering Science Direct inoperable for a long period of time. Elsevier is the largest STM publisher in the world and is investing heavily in digitisation of backfiles so this is a major archive. It was estimated that more than 7 terabytes of data would be delivered to the KB from Elsevier, once backfile digitisation is complete. It is premature to speculate to what extent this model could be applied to other publisher/library partners but it will be useful to monitor the experience of the KB/Elsevier agreement closely.

JISC/Academic Press

On a more modest scale is the arrangement between JISC and Ingenta for hosting Academic Press e-journals licensed under PSLI. The Public Site Licensing Initiative (PSLI) ran from 1996-1998, with three publisher participants (Blackwells, IOPP, Academic Press). The original objective was to reverse the spiral of declining funds for journals and to reduce the overall cost to the HE system of *paper* journals. At this point, e-journals were regarded as a by-product but their importance became increasingly apparent as the scheme progressed and problems of electronic delivery were encountered. Academic Press had separate licenses for electronic and print and licensed directly with HEFCE and the other three funding bodies, who then sub-licensed individual sites. Blackwells and IOPP contracted with the funding bodies to license sites directly.

At the end of the scheme, JISC put out a tender to host the e-journals in the Academic Press archive, consisting of 174 titles, to ensure continuing access for subscribers in the scheme. This has been undertaken by BIDS, (subsequently Ingenta) since then. A meeting with Terry Morrow and John Morrison on 23rd August confirmed that the titles took up c, 200 gigabytes of storage and served c.100 institutions. Most of the AP archive is SGML (mainly headers but some full-text) or PDF. Ingenta estimate that it will be c. 5 years before problems could be expected to be encountered with PDF because of changing technologies.

A more pressing issue was the implication of the Elsevier take-over of Academic Press and implications for delivery following the closure of the IDEAL platform from January 2003. Storage space is relatively cheap so marginal increases in storage would be expected to attract marginal cost increases, up to a certain point. If and when a bigger machine is required this

would entail a major expense. Ingenta also emphasised that their major business is hosting and aggregating content so they would be interested in taking on more but stressed that digital preservation management is a different matter.

Hosting the AP titles at Ingenta has clearly been very successful and has retained the archive of the 174 titles since 1998 and made them accessible to authorised users at modest cost to JISC. As Ingenta themselves point out however, hosting is not the same as preserving information, the former involving a relatively passive role while the latter requires an active involvement in preservation planning to overcome inevitable technological change. Hosting can provide an excellent breathing space while a more robust technical infrastructure is developed. Despite a wide variety of formats noted in the Publishers survey, relatively simple formats still predominate for the time being and this provides a simple, low cost solution for the short term.

Open Access Models

Open access journals offer an alternative model to the traditional toll access, in which institutions purchase access to content from publishers. The open access model levies a charge on the individual researcher, or their institution, for submitting a research article. Exponents of this model argue that it is a more logical and fairer distribution of research funding, benefiting both academics, whose works are disseminated freely, and institutions who fund that research. In terms of assured ongoing access, the assumption is that making material freely accessible and providing hosting by trusted partners will provide the necessary guarantees.

One example os the arrangement between BioMedCentral[17] and PubMedCentral[18] in the U.S and INIST[19] in France. PubMedCentral is a digital archive of life sciences journal literature managed by the National Center for Biotechnology Information and the US National Library of Medicine. Access to PubMedCentral is free and unrestricted. Peer reviewed research articles from participating journals are made available, with time delays depending on the publisher. Other content, such as reviews, essays and letters, is also made available at the discretion of the publisher so it is not an exact replication of the journal. [Note: The latter may be a problem for some users, some library survey respondents noted that some users were concerned that certain material was excluded from the electronic version of some journals.]

The Institut de l'Information Scientifique (INIST) of the National Center for Scientific Research (CNRS) has shown a strong commitment to open access, reinforced with its agreement with BioMedCentral. This agreement allows CNRS researchers to submit their research papers to BMC without charge, in return for INIST hosting BMC journals.

BioMedCentral is also a participant of LOCKSS and is searching for more archiving partners.

JSTOR

JSTOR[20] is familiar to UK HE institutions, many of whom subscribe to the JSTOR service. The JSTOR fee structure is based on an archive capital fee, which funds an endowment that provides for future data and software migrations, and an annual access fee. The JSTOR model was originally based on making digital copies of print journals and making these available to members who can access material on the basis of a "moving wall". This determines the age of the most recent issue which will be available in JSTOR and it has been very successful in combining library's needs for assurances on preservation without threatening the publishers' business models.

Email exchanges, supplemented with a face-to- face meeting with Eileen Fenton and Kevin Guthrie provided further information on the new phase of JSTOR, which is demanding a very different approach to the traditional JSTOR model. The latter has been built around print journals, and is therefore optimized for page-based materials. "Born digital" materials require very different models, systems, and processes, and the Mellon grant will enable JSTOR to develop these. The Mellon funding of $1.3m over eighteen months is supporting a new functional unit within JSTOR. This planning stage will conclude on 31st March 2004.

Work has begun on drafting technical requirements and initial system design for the archive. Possible business models are also being explored, and a study aimed at understanding all costs incurred by libraries in processing and incorporating e-journals into their collections is under way (and hence possible incentives to use a third party). A similar exploration of the economics and incentives of e-journals from the publisher perspective will also be undertaken. It would be very useful to analyse this data as it becomes available and also to conduct a similar study for the UK. JSTOR has built up a good relationship with UK universities over the years, as evidenced in the library survey.

Recommendation 7

Undertake a study into the economics of journal archiving in UK libraries, to complement the findings of the JSTOR study.

LOCKSS

The LOCKSS[21] model, based at Stanford, creates low cost, persistent caches of journals content which is housed at the institutions authorised to licence content from them. It effectively permits the institutions licensing content to "own" the content they are paying for, as they do with print. Beta testing between 1999-2002 has already demonstrated that its model and protocol are technically viable. A number of publishers of interest to NESLI/JISC deals have participated in LOCKSS testing, for example Blackwells, Project Muse, BMJ, OUP, ACS.

Four UK research libraries have been involved in LOCKSS trials, they are the British Library, Imperial College, University of Leeds, and the University of Cambridge. The additional Mellon funding is intended to support the next stage of LOCKSS, to manage content as bibliographic entities rather than as web-addressed files. LOCKSS is scheduled to launch in 2004, pending successful beta testing. Like JSTOR, it has developed a relationship of trust with publishers.

A telephone interview with Vicky Reich on 27th January 2003 included the following issues:

- The need to raise broad awareness of LOCKSS and the potential role it can play.
- The need to encourage further participation from UK research libraries.
- The desire to encourage more active participation from publishers.

It would be very worthwhile exploring the LOCKSS model further, and its potential role for UK licensed content.

Recommendation 8 Undertake a technical analysis of LOCKSS and its potential role for preserving UK content.

OCLC Digital Archive

OCLC[22] have been actively involved in undertaking research into digital preservation for a number of years. They have collaborated with RLG on two working groups, one developed a framework for preservation metadata (and will now move into developing recommendations and best practices for implementing preservation metadata) and the other was in defining attributed and responsibilities for Trusted Digital Repositories. It seems logical that OCLC should extend its range of services to include archiving. They have worked hard at building up a good relationship with publishers and offer a range of funding options for participants. There are currently 4,500 primarily research journals under contract to OCLC, covering a range of subject areas.

Recommendation 9 Undertake a more detailed analysis of the potential role of OCLC and JSTOR as trusted third party providers, for UK licensed content.

U.S National Digital Information Infrastructure and Preservation Program (NDIIPP)

This aims to undertake building appropriate infrastructure on a very large scale within the U.S. However there will certainly be transferable lessons for elsewhere, it has become axiomatic to say that many of the issues of digital preservation are global so the results of the extensive planning and research for NDIIPP will certainly be of value to the U.K. (and vice versa).

Appendix 9 of the NDIIPP plan[23] describes a high level overview of the proposed architecture and preservation network. An assumption made by the technical team is that "the NDIIPP will never be finished in any static sense but will instead need to be able to evolve continually to be able to integrate new forms of hardware and software, and to preserve digital material of new format and types."

To accommodate this need to evolve, the architecture includes four layers, each of which encompasses a different set of functions and rules for use. A further assumption was that the infrastructure will need to be built modularly rather than monolithically. This broad conclusion was also reached at the JISC Workshop on Licensing and Archiving, where it was recognised that a single monolithic structure to support the archiving of licensed e-journals is likely to be neither feasible nor desirable.

Recommendation 10
Adopt a staged approach to the development of a centrally co-ordinated archiving service for the U.K, enabling a flexible, distributed approach and utilising emerging trusted third party services, as appropriate.

E-Print Repositories

The rapid escalation of e-print repositories has been regarded by some of its champions as a potential replacement for more traditional scholarly communication provided by licensed e-journals. The emphasis to date has been on encouraging scholars to deposit content into the archives, rather than on preservation requirements. Indeed some proponents of e-print repositories argue against considering preservation requirements at this stage as being largely irrelevant for the time being and something which can be considered later (if at all).

As these repositories will be expected to contain valuable scholarly resources, it is to be hoped that their preservation will be taken into consideration and a related JISC funded study is currently investigating this[24]. In the meantime, the immediate concern of libraries is whether they can rely on continued access to material they have paid for through licences. While there will be overlapping issues to be addressed in preserving licensed e-journals and e-prints, at this stage they need to be considered separately.

The development of e-print repositories should be watched with interest and offer the potential for more rapid access to scholarly research. However, this study recommends they should not be regarded as a complete substitute for licensed e-journals but as a parallel development.

Recommendation 11
Maintain a watching brief on the development of e-print repositories (taking into account the findings of the recent JISC funded study), but do regard them as a parallel development to archiving licensed e-journal content.

3.10 Legal Issues

The presentation on legal issues given by Emanuella Giavarra at the Workshop on Licensing and Archiving on 17th February posed three key questions:

- How can libraries be assured of continued access to the content they have paid for?
- How effective is the Model Licence in providing protection for assured ongoing access?
- What is the preferred and most achievable archiving model?

Copyright law and contract law are the two legal options for ensuing continued access to content paid for, the latter being invoked through the Model licence clauses 2.2.2, 5.4.1 and 5.4.2. Three potential options are provided for in the Model Licence, archiving by the publisher, archiving by the individual institution, or archiving by a central facility operated on behalf of UK HE/FE.

Currently, contracts are between individual institutions and publishers. The choice of options is left to individual institutions, and the choice and decision to enforce is left to them. In practice, there are currently only two of the three options specified in the licence available, either an archival copy delivered to the individual institution, or archiving by the publisher. A copy of a CD-ROM containing content previously paid for is frequently offered to institutions cancelling their subscription but, as noted in previous chapters, this is the least preferred option for libraries and cannot be considered to be a viable strategy.

A central facility would need to negotiate new contracts with each and every publisher. The JISC Workshop concluded that there is no immediate enforcement problem regarding licence clause 2.2.2 since the NESLI and JISC journal deals have been renewed. However, it was agreed that greater protection is needed, in particular for cases where deals are not renewed and only CD ROM's can be provided and/or where renewal contracts override the access obligations in previous contracts.

The ability to reliably track access rights across time will be crucial for a central archiving service. E-Journals are particularly complex as licence agreements and subscribers may change annually. This implies archives of content, will need to be accompanied by archives of licences and subscriber information, and/or "moving walls" for opening up access will need to be agreed between parties. Developing the capacity to handle complex rights was seen as an essential development by both libraries and publishers at the JISC Workshop.

A further issue raised by the Legal issues breakout group at the JISC Workshop was the national acquisition of backfiles from several publishers. It was felt that this model would particularly benefit smaller publishers as most of them find it difficult to comply with perpetual access and other requirements, such as Athens.

Recommendation 12

JISC should undertake a requirements study for Digital Rights Management in the JISC Information Environment.

3.11 Costs and Funding Models

UK Legal Deposit.

For the British Library, two cost scenarios were constructed. One assumed that the BL would receive 60% of unique e-serials on deposit by year 3 (compared to an estimated 45-50% now) and a steady rate (estimated 75%) of e-monographs. The second scenario assumes the BL will receive 80% of e-serials by Year 3. The total costs over three years were estimated at £2.6m and £2.7m respectively[25]. These figures were also distributed by Anne Foster at the Workshop on Archiving and Licensing in February 2003.

There is no indication at this stage of whether this additional funding will be allocated to the British Library, but the annual estimate equates quite well with additional funding provided to the KB shown below.

National Library of the Netherlands

The Ministry of Education, Culture and Science in the Netherlands funds the National Library of the Netherlands. Between 1998-2002, the KB received 1.45m euros (c. £1m) over four years, plus research funding to prepare for their new deposit system for digital objects. Structural funding of 1.14m. euros(c.£800k) p.a. for ongoing support of this activity has been allocated for 2003 and beyond[26].

NDIIPP

$100m (c. £62m) will be released in stages for the development of the National Digital Information and Infrastructure and Preservation Program[23]. $5m has already been allocated for the initial planning year and preparation of the NDIIPP Plan[27]. This has now been approved by Congress, releasing a further $20m for the next stage, which will include building a prototype infrastructure capable of supporting digital preservation of a wide variety of digital formats with a wide variety of rights management and access requirements.

OCLC Digital Archive

Development costs over an 18-month period were $2.4m (c.£ 1.5m) for staff and admin and $675k (c. £422k) for hardware/software[28].

JISC Academic Press Archive

Ingenta have provided a hosting service for the 176 e-journals licensed by Academic Press under the PSLI scheme, since 1999. These titles are estimated to require c. 200 gigabytes of storage and access is provided to c.

100 institutions who have licensed access to the content of the archive. The cost of this has been somewhere between £9k and $11k p.a.

JISC Workshop Discussion on Business Models

A range of possible business models were explored by the JISC Workshop during the breakout session on Business models. Capital development funding could be taken from public funding (a submission to the Comprehensive Spending Review, or the proposed Research Libraries Network were suggested); a one-time investment by foundation members (i.e. UK university libraries); private or in kind donations (i.e. publishers). Ongoing support and maintenance could be funded by public funding, subscriptions from foundation members, pay-per-view from other institutions and researchers, publishers, institutions and archives collaborating on assigning metadata, content storage and managing and administering user access. Questions raised in the workshop related to incentives, what were the incentives for publishers and libraries? It was felt that there needs to be further work on exploring this. Demonstrating that the model is cost-effective was also seen to be critical. A related conclusion was that the UK is probably not a large enough a market to act on its own and the international scope of the content licensed could require wider collaborative arrangements .

Summary

It is clear from these figures that planning and development costs of preservation archives are substantial. The Mellon funded Yale E-Journals project provided some interesting broad categories of cost life-cycles of an e-journal archive. These are:

> 1) Development and start-up (most difficult); 2) ongoing maintenance and problem resolution (easier parts and likely to become marginal over time); 3) collaboration and standards (tricky) many people have been working on digital preservation, so there will be similar, but not identical solution's coming to light; 5) migration (difficult and probably v. expensive)[29].

The hosting services described in this report are essentially operating under 2) above, i.e. ongoing maintenance and problem resolution. These can provide a relatively low cost option for the short-medium term and also effectively utilise existing infrastructure. Longer term guarantees will be considerably more costly.

OCLC Incentives Paper

The issue of incentives is a recurring theme. It was noted in the Mellon e-journal studies, it is underpinning much of JSTOR's business modelling, it was raised at the JISC Workshop, and it is the focus of a recent OCLC Report. The report is based on the assumption that economics are fundamentally about incentives. It suggests that necessary first steps in building sustainable digital archives must start with an examination of conditions under which there

are insufficient incentives to preserve, and determining how this can be remedied.

> …as digital preservation moves beyond the realm of small-scale, experimental projects to become a routine component of a digital asset's life-cycle management, the question of how it can be shaped into an economically sustainable process begins to overshadow other concerns[30].

There is a great deal of work to be done before absolute assurances to ongoing access to e-journals licensed by UK universities can be assured. There are potentially substantial costs in building a sufficiently robust infrastructure and the question of incentives to undertake this will need to be addressed in the context of relevant third party services being developed elsewhere.

Recommendation 13 Rigorous business modelling and examination of incentives be undertaken for a small sub-set of JISC deals (in the first instance) in order to determine the sustainability of proposed solutions. This should include an assessment of potential savings in storage space if a robust, reliable e-journal archiving service was operational.

4 Conclusions

4.1 Collaboration between Libraries and Publishers

This is a constantly recurring theme throughout the literature searches and from feedback from the surveys and the Workshop. In many ways it seems self evident but there is a need to nurture this relationship and to ensure mechanisms are in place to enhance mutual understanding.

Recommendations relating to this include providing a dedicated resource to establish ongoing dialogue with publishers involved in JISC journal deals, establishment of a PALS Steering Group, joint library/publisher forums. The difficulty of maintaining unambiguous communication with publishers is increased because of the extremely diverse nature of publishers and similarly diverse knowledge and interest of digital preservation. It is therefore important to engage with individual publishers to understand their specific concerns, as well as to undertake more wide ranging opportunities for exchange of views.

The Digital Preservation Coalition has a cross sectoral approach, and includes publishers in its members. However there is also a need for a bi-lateral approach so that issues relevant to the specific problems of ensuring continued access to licensed content can be focused on and satisfactorily resolved.

The archiving clauses in the Model Licence need to be explored in greater detail with publishers, with a view to finding mutually acceptable means of implementing them.

4.2 The Needs of UK universities

From undertaking this study it is clear that UK university libraries require more assurance and support in implementing the Model Licence clauses. They require assurances of ongoing access to all licensed content they subscribe to, wherever it originates. They are looking to JISC to continue to provide leadership in licensed journal content, and also to the British Library because of the latter's crucial role in any national model for digital preservation. It is important to balance the urgent need for action with a need to recognise the constraints of establishing a repository on behalf of UK universities. The latter is probably not a large enough market to warrant the significant costs which would certainly be required to establish a repository for all e-journals licensed by UK universities.

Moreover, it is important to monitor and assess the potential of services being developed internationally, for example, LOCKSS, JSTOR, and OCLC, to meet at least some of the needs of the U.K.

4.3 **Legal Deposit**

It is important to note the distinctions between services to provide continuing access on behalf of paying subscribers in UK universities, and the national published heritage role of legal deposit libraries. However, there may be synergies which can be explored between these functions and publishers and libraries. The British Library and UK legal deposit libraries will be important participants in the development of any national model for preservation of the national published corpus in digital form . However legal deposit legislation has not yet been enacted and in any case will be limited to UK content. UK universities licence access to a wide range of content and the country of origin is of much less concern than the quality of the content. There are also issues of service delivery, and funding, which are not yet clear so it is premature to speculate to what extent the Deposit libraries would be willing or able to play a role in the ongoing provision of licensed e-journals. However, despite these caveats, these issues should be considered further as part of ongoing discussions between publishers and libraries.

4.4 A Staged Approach

An important breakthrough in the JISC E-Journals Workshop was the realisation that it was possible to establish a service which co-ordinated archiving provisions but permitted distributed delivery. This will provide immediate action, which is urgently needed, but at a relatively modest cost, allowing the service to develop incrementally, as further knowledge and experience is gained. It also provides the necessary flexibility to accommodate a wide range of publishers and needs.

Providing a dedicated resource to build on the work of this consultancy will retain the necessary momentum. It will be essential to keep monitoring relevant developments overseas, and to establish contact with those services and initiatives, as well as to explore in greater depth the specific requirements of the UK research community. It will also be critical to maintain close communication with both the newly established JISC Journals Negotiation Team and publishers providing content to UK universities.

Further funding will need to be sought in subsequent years, for third party services and/or establishment of a UK-based repository.

Some of the proposed tasks for Year 1 include:

- Establish close links with JISC Journals negotiation team
- Follow up archiving provisions for all ten deals brokered by the JISC Journals Negotiation team
- Focus in depth on archiving provisions for one deal, as identified by the JJWG and/or PALS Steering Group as being particularly important, and make recommendations on this

- Provide the central contact point for publishers and institutions involved in JISC journal deals and other licensed content

- Establish close links with potential service providers, in particular BL, JSTOR, OCLC, LOCKSS

- Provide a focal point for UK participation in LOCKSS

- Analyse options and requirements for rights management issues

- Ensure other relevant work being undertaken by JISC is taken into account. For example JISC Information Environment, TIES, (the JISC-funded project on authentication which is due to report in September 2003), PALS Model Licence Review Group, findings of the JISC Study on E-Prints, etc.

- Prepare reports and recommendations on archiving requirements for JISC journal deals

5 Summary of Recommendations

Recommendation 1

That JISC allocate resources for a consultancy to focus on the longer term access and archiving provisions of journal deals.

Recommendation 2

That JISC facilitate ongoing discussions with publishers about mutually acceptable arrangements for implementation of the archiving clauses of the Model Licence.

Recommendation 3

Ensure ongoing and constructive dialogue between libraries and publishers by developing relationships with selected publishers included in JISC deals.

Recommendation 4

Ensure ongoing collaboration and partnership between libraries and publishers through a range of mechanisms, including regular meetings, and further joint forums to address archiving issues.

Recommendation 5

Maintain an active watching brief on initiatives and services being developed to provide archiving of licensed e-journals by trusted third parties.

Recommendation 6

Explore the range of options for implementing the JISC model for individual journal deals.

Recommendation 7

Undertake a study into the economics of journal archiving for UK libraries, to complement the findings of the JSTOR study.

Recommendation 8

Undertake a technical analysis of LOCKSS and its potential role for preserving UK content.

Recommendation 9

Undertake a more detailed analysis of the potential role of the OCLC Digital Archive and JSTOR as trusted third party providers for UK licensed content.

Recommendation 10

Adopt a staged approach to the development of a centrally co-ordinated archiving service for the U.K, enabling a flexible, distributed approach and utilising emerging trusted third party services, as appropriate.

Recommendation 11

Maintain a watching brief on the development of e-print repositories (taking into account the findings of the recent JISC funded study), but regard them as a parallel development to archiving licensed e-journal content.

Recommendation 12

JISC should undertake a requirements study for Digital Rights Management in the JISC Information Environment.

Recommendation 13

Rigorous business modelling and examination of incentives be undertaken for a small sub-set of JISC deals (in the first instance) in order to determine the sustainability of proposed solutions. This should include an assessment of potential savings in storage space if a robust, reliable e-journal archiving service was operational.

References

1. CURL Exemplars in Digital Archives (Cedars).
<http://www.leeds.ac.uk/cedars/>

2.Cedars Guide to Intellectual Property Rights (2002).
<http://www.leeds.ac.uk/cedars/guideto/ipr/>

3. JISC Model Licence
<http://www.nesli.ac.uk/modellicence.html>

4. Publishing in the Knowledge Economy. (2002). Action Point 22.
<http://www.uk-publishing.info/competetive.asp>

5. Details of all seven projects, including final reports, where available, can be found at:
<http://www.diglib.org/preserve/ejp.htm>

6. Archiving – Whose Problem Is It? Joint ALPSP/DPC Seminar, 29th November 2002.
<http://www.alpsp.org/s291102.htm>

7. Waters, Don. (2002) Good Archives Make Good Scholars: Reflections on Recent Steps toward the Archiving of Digital Information. *In* The State of Digital Preservation: an International Perspective. CLIR.
<http://www.clir.org/pubs/abstract/pub107abst.html>

8. National Serials Requirement Survey (2002).
<http://www.nesli.ac.uk/NSRS_Report_Final.doc>

9. Morgan Stanley (2002). Scientific Publishing: Knowledge is Power.
<http://www.alpsp.org/MorgStan300902.pdf>

10. National Serials Requirement Survey (2002) op cit

11. JISC Continuing Access and Digital Preservation Strategy for the Joint Information Systems Committee (JISC) (2002-2005). p.A2
<http://www.jisc.ac.uk/index.cfm?name=pres_continuing>

12. RLG/OCLC (May 2002). Trusted Digital Repositories: Attributes and Responsibilities.
<http://www.rlg.org/longterm/repositories.pdf>

13. Task Force on Digital Repository Certification
<http://www.rlg.org/longterm/certification.html>

14. Reference Model for an Open Archival Information System (OAIS). January 2002.
<http://wwwclassic.ccsds.org/documents/pdf/CCSDS-650.0-B.1.pdf>

15. The impact of the extension of legal deposit to non-print publications. Study report prepared for the JCVD by Electronic Publishing Services Ltd. October 2002.
<http://www.epsltd.com/ExtensionOfLegalDeposit.htm>

16. National Library of the Netherlands and Elsevier Science make digital preservation history. Press Release 23rd August 2002
<http://www.elsevier.com/inca/publications/misc/ni21.pdf>

17. BioMedCentral
<http://biomedcentral.com/>

18. PubMedCentral
<http://www.pubmedcentral.nih.gov/>

19. INIST-CNRS/BioMedCentral Ltd Partnership Agreement for Promotion of Scientific Information. Press Release, 3rd December 2002.
<http://www.inist.fr/en/actu_en/cp_031202bmc.php>

20. JSTOR
<http://www.jstor.org/>

21. LOCKSS
<http://lockss.stanford.edu/>

22. OCLC Digital Archive
<http://www.oclc.org/digitalpreservation/services/archiving/digital/>

23. Shirky, C. Preliminary Architecture Proposal for Long-Term Digital Preservation. *In* Plan for the National Digital Information Infrastructure and Preservation Program. Appendix 9.
<http://www.digitalpreservation.gov/ndiipp/>

24. JISC Study on Archiving E-Prints
http://www.jisc.ac.uk/index.cfm?name=project_eprints_pres

25. The impact of the extension of legal deposit to non-print publications. Study report prepared for the JCVD by Electronic Publishing Services Ltd. October 2002.
<http://www.epsltd.com/ExtensionOfLegalDeposit.htm>

26. Source: Beagrie, N. National Digital Preservation Initiatives: An Overview of Developments in Australia, France, the Netherlands, and the United Kingdom. *In* Plan for the National Digital Information Infrastructure and Preservation Program. Appendix 5.
<http://www.digitalpreservation.gov/ndiipp/>

27. Plan for the National Digital Information Infrastructure and Preservation Program
<http://www.digitalpreservation.gov/ndiipp/>

28. DPC Forum on Preservation of e-Learning Materials and Cost Models for Digital Preservation, 15th October 2002.
<http://www.dpconline.org/graphics/events/index.html#elearning>

29. YEA: The Yale Electronic Archive; One Year of Progress. (2002)
<http://www.library.yale.edu/~okersen/yea>

30. Lavoie, B. (2003)The Incentives to Preserve Digital Materials: Roles, Scenarios, and Economic Decision-Making.
<http://www.oclc.org/research/projects/digipres/incentives-dp.pdf>.

Appendix A

Annotated Select Bibliography of Sources Checked during e-journals archiving consultancy

Mellon e-journal archiving projects

1. Waters, Don. Good Archives Make Good Scholars: Reflections on Recent Steps toward the Archiving of Digital Information. In The State of Digital Preservation: an International Perspective. CLIR, 2002

<http://www.clir.org/pubs/abstract/pub107abst.html>

In 2001, Mellon awarded funding for a one-year planning effort to seven universities, focussing on archiving of e-journals.
Of the seven funded projects, one (Stanford) focussed on developing technology to harvest presentation files, the web-based material used by publishers to present content to readers, and storing them in a highly distributed system. This approach is known as LOCKSS (Lots of Copies Keeps Stuff Safe). Five of the seven projects focussed on the capture of publishers' source files, including high quality images and text that is encoded in either SGML or XML. The seventh project (MIT) focussed on dynamic e-journals, for example CogNet and Columbia International Affairs Online (CIAO).

The projects were due to report in September 2002 (see YEA, Harvard, and MIT Reports below) but Don Waters identified three broad conclusions from their research:

1. Archiving now seems technically feasible using different approaches, using web-based presentation files with LOCKSS, and capturing source files
2. Participating publishers have come to view archiving as a competitive advantage
3. There is an increasingly shared understanding that an e-journal archive should aim to make it possible to regard e-journals as publications of record and to persuade publishers and librarians to consider abandoning print.

The Mellon Foundation has concluded that new organisations will be necessary to act in the broad interest of the scholarly community and mediate the interests of libraries and publishers. Much work is needed to get the right model but Mellon will not support the concept of "dark" archives in which a publisher can claim the benefit of preservation but yield no right of access.

Mellon is now funding two major approaches, preserving presentation files, using LOCKSS, and preserving source files. The LOCKSS development was funded from June 2002. A proposal from JSTOR for preserving source files was funded from September 2002. LOCKSS and JSTOR are two very different approaches and Mellon is deliberately supporting both in the belief that a single definitive approach is never going to be likely, so they should be regarded as complementary approaches, rather than competing.

Another critical impediment for digital archiving that arises again and again is the legal regime governing intellectual property. It might be premature to specify exemptions which could apply to digital information but there is a need to begin to

articulate “safe harbor” principles about intellectual property rights that could form the basis of digital archiving agreements among interested parties.

2. Yale Electronic Archive Report. 2002

<http://www.library.yale.edu/~okersen/yea/>

Yale was one of the seven Mellon funded e-journal preservation planning projects, all of which are expected to report during 2002. They will all be linked from: <http://www.diglib.org/preserve/ejp.htm>

This project was a collaboration between Yale Library and Elsevier. During the year, Yale gained a better understanding of the commercial life-cycle of electronic journals and ways in which journal production will impact on the success of the e-archive.
Initial assumptions included:

- Users will be primarily interested in content rather than functionality
- Minimally acceptable standards of access can be defined

Some conclusions reached:

- Migration is a more realistic strategy than emulation
- Preservation metadata differ from those required for production systems and add real value to the content
- Success will depend on adherence to broadly accepted standards and best practices
- A reasonable preservation process is one which clearly identifies the “trigger events” that would require consultation of the archive and plans accordingly

Intersecting needs of the three main players (i.e. authors, publishers, and libraries) were identified.
By the summer of 2000. Yale was spending > $1m annually on online digital resources.
Yale have acquired parallel print and electronic copies for many resources but it is clear that this “dual pathway” is not sustainable.

There is a need for industry-wide e-preservation standards and close co-operation between libraries and publishers because the most cost-effective production of metadata will be by publishers.
There was a general assumption that software tools being developed in other institutions would be used,as appropriate, for example, OAI as a search engine, though there was concern by publishers to use OAI for retrieval.
The mission of the archive is to provide permanent access to content BUT “This simple truth grows immensely complicated when one acknowledges that such access is also the basis of the publisher’s business and that, in the digital arena (unlike the print arena), the archival agent owns nothing that it may preserve and cannot control the terms on which access to preserved information is provided.”

Three types of archive were modelled 1) de-facto (e.g. Ohiolink) which have current licence to load all of a publisher’s journals locally; 2) self-designated archival service (e.g. Elsevier/KB agreement) and 3) publisher-archival agent partnership (the focus of Mellon investigation).

The main sticking point has been access. If there is NO access then there is very little incentive to build an archive.
Costs. Guess that the truth lies somewhere between the two polarities that “costs will be huge” or “it will be free” (open archiving literature). Detailed costs for building the YEA for e-journals have been developed and those calculations are beginning to

provide a sense of scale for such an operation. Five cost life-cycle stages of an e-journal archive were identified;

1) Development and start-up (most difficult); 2) ongoing maintenance and problem resolution (easier parts and likely to become marginal over time, particularly as standards develop, 3) collaboration and standards (tricky), many people have been working on digital preservation, so will be similar, but not identical solution's coming to light; 4) comprehensiveness (messy). There will be a fair number of journals unable or unwilling to co-operate, 5) migration (difficult and probably v. expensive).

This was acknowledged as a very simplified overview but as relatively few of the 1100+ Elsevier journals contained complex information objects there is some time left to address and solve the problems of preserving and archiving complex digital info.

A range of business models are considered, with the pros and cons of each, e.g. lifetime annuity model, where the user pays a defined quantum upfront in return for an eternity of preservation.

Licensing agreement. It was surprisingly easy to reach a general agreement between Yale and Elsevier. A 10 year term and 10 year renewal was selected. The most difficult issue to resolve was remained what constituted a "triggering event". In the end the only triggering event they could reach agreement on was when materials being archived are no longer commercially available. However, it is difficult to imagine such a scenario which begs the question of whether it makes sense to maintain a separate archive at all. It was agreed that, as a minimum, the archive would need to "exercise" the archive so need to at least make it "dim" as opposed to "dark".

Financial assumptions are that the publisher provides files without charge and the archival agency accepts perpetual archiving responsibility without financing from the publisher (as in the KB/Elsevier agreement described below.

3. MIT Libraries. DEJA: A Year in Review. 2002
http://www.diglib.org/preserve/mitfinal.html

This Mellon e-journal project intended to focus on dynamic e-journals, defined as "those that contain moving elements or make elements move". However, there were so few journals which currently meet these criteria that it was necessary to broaden the scope to include other journals from smaller publishers who were involved in innovative techniques and business models. The project found that these publishers were either non-commercial or not high-profit motivated (e.g. SPARC journal publishers). Because of these factors, they were willing to accept licensing agreements which would allow the e-journal archive to make their content accessible within a relatively short timeframe, e.g 3-5 years. They were also willing to consider subsidising the archives of their publications but are unable to do so without passing the costs back to their subscribers or members. The project concluded that there is an urgent need for further research into the technical means of preserving dynamic e-journal publications and also of identifying the technical metadata needed to support that preservation. There is also a need to build cost models for doing such preservation.

4. Harvard University Library. Report on the Planning Year Grant for the Design of an E-Journal Archive. 2002
http://www.diglib.org/preserve/harvardfinal.html

Like the Yale project, described above, the Harvard project was based on a collaboration between an archive and publisher. In this case there were three publisher partners, Blackwell Publishing, Wiley, and the University of Chicago Press.

Between them, these three publishers produced 1,137 journals at the time of the project. The decision to restrict participation to larger publishers was made to ensure effective contribution to the technical planning process. However it was hoped that as work develops, it will lead to the production of tools and infrastructure which might be shared by others.

Decisions were made on which components of e-journals should be included. It was agreed that this needed to include more than the articles contained but also other material, some of which is not currently necessarily included in the e-version of journals. This represents front matter, such as editorial boards, only the most current versions of much of the front matter of paper journals was included in the e-versions of journals included in the project. The project decided to exclude advertisements, not because they are not a potential source of future scholarly interest but because the particular challenges they pose, combined with their relatively minor content in e-journals led to the decision to exclude them from this particular project.

Access issues were discussed, and, as with the Yale project, it was not possible to arrive at an agreed definition of what should be included as a trigger event. This is an area requiring further investigation. Economic issues are also considered in the report. This is complicated by the fact that the costs of archiving are not currently known. Harvard proposes that funding for a sustainable archive should accompany deposit of content from the outset. Furthermore, this should consist of two aspects, an "ingestion fee" covering costs of receipts, quality control, and archival preparation. Secondly, an endowment should be paid to cover the long-term costs of storage and preservation activity.

5. Lots of Copies Keep Stuff Safe (LOCKSS).
<http://lockss.stanford.edu/>

The LOCKSS model, based at Stanford, creates low cost, persistent caches of journals content which is housed at the institutions authorised to licence content from them. It effectively permits the institutions licensing content to "own" the content they are paying for, as they do with print. Beta testing between 1999-2002 has already demonstrated that its model and protocol are technically viable. A number of publishers of interest to NESLI/JISC deals have participated in LOCKSS testing, for example Blackwells, Project Muse, BMJ, OUP, ACS. Four UK research libraries have been involved in LOCKSS trials, they are the British Library, Imperial College, University of Leeds, and the University of Cambridge. The additional Mellon funding is intended to support the next stage of LOCKSS, to manage content as bibliographic entities rather than as web-addressed files. Like JSTOR, it has developed a relationship of trust with publishers. A telephone interview with Vicky Reichs on 27th January 2003 included the following:

- The need to raise broad awareness of LOCKSS and the potential role it can play.
- The need to encourage further participation from UK research libraries.
- The desire to encourage more active participation from publishers.

6. JSTOR.
<http://www.jstor.org>
JSTOR is familiar to UK HE institutions, many of whom subscribe to the JSTOR service. The JSTOR fee structure is based on an archive capital fee, which funds an endowment that provides for future data and software migrations, and an annual access fee. The JSTOR model was originally based on making digital copies of print journals and making these available to members who can access material on the basis of a "moving wall". This determines the age of the most recent issue which will

be available in JSTOR and it has been very successful in combining library's needs for assurances on preservation without threatening the publishers' business models.

Email exchanges, supplemented with a face-to- face meeting with Eileen Fenton and Kevin Guthrie provided further information on the new phase of JSTOR, which is requiring a very different approach to the traditional JSTOR model. The Mellon funding of $1.3 over eighteen months is supporting a new functional unit within JSTOR. This planning stage will conclude on 31st March 2004. Work has begun on drafting technical requirements and initial system design for the archive. Possible business models are also being explored, and a study aimed at understanding all costs incurred by libraries in processing and incorporating e-journals into their collections is under way. A similar exploration of the economics of e-journals from the publisher perspective will also be undertaken.

OCLC Digital Archive

7. OCLC Digital Archive.
<http://www.oclc.org/digitalpreservation/services/archiving/digital/>
OCLC have been actively involved in undertaking research into digital preservation for a number of years. They have collaborated with RLG on two working groups, one developed a framework for preservation metadata (and will now move into developing recommendations and best practices for implementing preservation metadata) and the other was in defining attributed and responsibilities for Trusted Digital Repositories. It seems logical that OCLC should extend its range of services to include archiving. Further details of the question and answers regarding the OCLC Digital Archive appear at Appendix F.

National Digital Infrastructure and Preservation Program (NDIIPP)

8. Plan for the National Digital Information Infrastructure and Preservation Program. October 2002.
<http://digitalpreservation.gov.ndiipp/>

The report is the result of the first phase of a $175m. program being led by the Library of Congress. An initial $5m was released for the research and development of the NDIIPP plan. Now this has been approved by Congress, a further $20m. has been released to build the development of a prototype system to support digital preservation. Appendix 9 of the NDIIPP plan describes a high level overview of the proposed architecture. An assumption made by the technical team is that "the NDIIPP will never be finished in any static sense but will instead need to be able to evolve continually to be able to integrate new forms of hardware and software, and to preserve digital material of new format and types." To accommodate this need to evolve, the architecture includes four layers, each of which encompasses a different set of functions and rules for use. A further assumption was that the infrastructure will need to be built modularly rather than monolithically. This broad conclusion was also reached at the JISC Workshop on Licensing and Archiving, where it was recognised that a single monolithic structure to support the archiving of licensed e-journals is neither feasible nor desirable.

Open Access Models

9. The National Center for Scientific Research (CNRS) model is an open access collaboration between INIST/ CNRS in France and BioMedCentral. CNRS host

and distribute BioMedCentral journals and have negotiated a deal in which CNRS researchers do not have to pay to have their articles included in BMC.

Further information can be found at:

<http://www.inist.fr/en/actu_en/cp_031202bmc.php>

See also:
<http://www.inist.fr.openaccess/en/comsci_impli_projets.php>

The assumption behind open access models is that, because they are freely available, there is safety in redundancy as content can be readily hosted by third parties. BioMedCentral journals are also hosted in the U.S by PubMedCentral.

10. PubMedCentral was launched in February 2000 and provides a digital archive of life sciences journal literature managed by the National Center for Biotechnology Information and the US National Library of Medicine. It is not a journal publisher and access to PubMedCentral is free and unrestricted. It contains all peer reviewed primary research articles from every participating journal. Other content, such as reviews, essays, letters etc. are made available at the discretion of the publisher. Participating journals must supply full text of articles in SGML or XML format which conforms to the established DTD for journal articles. Supplementary material may also be submitted but PMC does not accept HTML. PMC titles are made available at the discretion of the publisher, with most releasing material within six months of publication, though some have delays of two years.

More information is available form the PMC website at:
<http://www.pubmedcentral.nih.gov>

Legal Deposit

11. Arrangements for Depositing Electronic Publications at the Deposit of Netherlands Publications in the Koninklijke Bibliotheek. 10 June 1999
<http://www.kb.nl/kb/dnp/overeenkomst-nuv-kb-en-pdf>

12. BL Code of practice for the deposit of non-print publications.
<http://www.bl.uk/about/policies/codeprac.html>

Both KB and BL assume stringent access restrictions on deposited material. BL also indicates that print will be the preservation medium where parallel print/electronic publications exist.

13. KB/Elsevier announcement 23/8/02
<http://www.elsevier.com/inca/publications/misc/ni21.pdf>

Press Release from KB and Elsevier announcing their agreement that KB will become the first official digital archive for Elsevier Science journals. The library will receive all copies of Elsevier journals made available on its web platform, Science Direct. This currently consists of c1,500 journal titles exceeding 7 terabytes of data (once all backfiles have been digitised).

The agreement provides assurance that, if Elsevier should go out of business, their files will not be lost. Journals are currently available in both PDF and a tagged,

structured text format that permits different on-screen viewing, faster network delivery and sophisticated search, retrieval and linking. Both formats will be sent to the KB.

The KB will take responsibility for migrating the content and associated software as technologies change. The new IBM e-deposit system will be ready this fall and handed over to the KB. At that time, Elsevier journals will be placed into it. Elsevier will work closely with KB to ensure permanent availability. KB will provide access onsite to visitors. Also, in the event of catastrophic disaster, KB will be part of an interim service system. If Elsevier should ever cease to make the journals commercially available, the KB can open access to all on a remote basis.

14. IFLA/IPA . Preserving the Memory of the World in Perpetuity: a joint statement on the archiving and preserving of digital information. 12 August 2002. <http://www.ifla.org/V/press/ifla-ipa02.htm>

The statement commits the two organisations (IFLA and the International Publishers Association) to working together on joint initiatives to study technical, economic, and policy issues associated with digital preservation.

National Site Licences

15. Canadian National Site Licensing Project. Principles for Licensing Electronic Resources. May 2000. <http://www.uottawa.ca/library/cnslp/docs/Licen_princip.htm>

The CNSLP project won the 2002 Innovation and Achievement Award of the Canadian Association of College and University Libraries. CNSLP is a consortium of 64 Canadian university libraries and its administrative centre is based at the University of Ottawa. 1.8 of their licensing principles states that:

> Priority will be given to those electronic resources that offer…Rights for long-term retention and/or provisions for access in perpetuity.

Email correspondence with Leacy O'Callaghan indicated that they are looking at a broad based evaluation of the project, which will be reported in 2003. Archiving will not be considered at this time but they anticipate that it will be in the next stage. A copy of the CNSLP licence was emailed as an attachment. The key archiving clause in this document is:

> 12.4 On termination of this Agreement, the Consortium, Authorized Users and Walk-in Users **shall retain the right to access and use in archived form the content of the Database for the period of time set out in Schedule 3 up to the date of termination**, except where such termination is due to a breach of the Agreement by the Consortium which the Consortium has failed to remedy as provided in clause 12.1.1 and 12.1.3, in which case such continuing access shall be provided in respect of Licensed Materials published up to the date of such breach. On termination of this Agreement, the Publisher shall at its option:
>
> a.) provide each Member, on request, with an electronic copy of the content of the Database for the period of time set out in Schedule 3 up to the date of termination, or
>
> b.) provide for continued access to the Licensed Materials on the Server for the period of time set out in Schedule 3 up to the date of termination,
>
> provided that:

12.4.1 the Consortium and each Member seeking access continues to adhere to its obligations with respect to the restrictions on use of the Database as provided in this Agreement;

12.4.2 each Member permitting access to Walk-in Users continues to limit such access to the Premises; and

12.4.3 **each Member seeking access pays to the Publisher a maintenance amount reflecting the Publisher's costs for facilitating such access as the Publisher and Member, acting reasonably, may agree.**

E-Prints

16. Jeffrey.R.Young. 'Superarchives' Could Hold All Scholarly Output. Chronicle of Higher Education. 5 July 2002.
<http://chronicle.com/free/v48/i43/43a02901.htm>

Institutional repositories are seen as creating an alternative to the unsatisfactory present scholarly journal situation which critics argue has a monopoly on scholarly output that leads to ever-soaring subscription prices.

The most ambitious 'superarchive' is D-SPACE (http://web.mit.edu/dspace). This is trying to collect research material from nearly every professor at MIT. MIT has been building a set of software tools for the past 2 years and will be testing these with 4 departments at MIT this summer.
The incentive for scholars is increased visibility. Refers to study indicating mean citations to offline articles is 2.74, compared to 7.03 of online citations. There are also different traditions for different disciplines. Scientists tend to want their research out asap, whereas scholars in humanities might have concerns about potential plagiarism.

Costs. D-Space was supported by a $1.8m grant from Hewlett-Packard. 'Officials aren't sure how much the archive will cost to maintain, though universities already have much of the equipment in place to run digital archives. Still, it is estimated that DSpace could cost up to $250,000 per year, if all the costs were added up. The hope is that free software tools will allow even small colleges to run repositories using their existing resources.'

Colleges setting up repositories need to set clear guidelines for who owns copyright.

17. SPARC Paper on Institutional Repositories.

The Case for Institutional Repositories: a SPARC Position Paper. Prepared by Raym Crow. Release 1.0. July 2002.
<http://www.arl.org/sparc/IR/IR_Final_Release_102.pdf>

Institutional repositories are defined as digital collections capturing and preserving the intellectual output of a single or multi-university community.

They have the potential to address two strategic issues facing universities, 1) reforming the system of scholarly communication by reasserting control over scholarship by the academy, increasing competition and reducing monopoly power of journals, and 2) the need for tangible indicators of a universities' quality and relevance.
Factors forcing change in structure of scholarly journal publishing:

- demand for broader access to research and for more robust digital presentation;

- significant increases in volume of research has exacerbated dissatisfaction with time-lag of print publications;
- increasing dissatisfaction with pricing models;
- increasing uncertainty over who will handle the preservation archiving of digital scholarly research.

The traditional model had librarians and academic institutions supporting archiving and this stays the same in the proposed disaggregated model though the other three functions (i.e. registration, certification, and awareness) change.
Fundamental to implementing the disaggregated model is the logical separation of the content and service components as advocated by Van deSompel and others. Thus allowing distributed open access repositories to be maintained independently of value-added services fulfilled by multiple service providers.

The main success in discipline-specific e-print servers has been from disciplines with established pre-print traditions.
Archiving. Suggests that libraries took responsibility for print, but publishers have asserted control over the digital versions of their publications. 'Institutional repositories, in the context of a disaggregated scholarly publishing model, keep responsibility for the preservation of research materials in the hands of librarians, those professionally prepared and committed to handle it.' (p. 12). **However polices and procedures must be in place to ensure e-print archives can scale up.**
Costs of institutional repositories are covered on p. 27-28. This suggest that both development and operational costs can range from almost no incremental costs to hundreds of thousands of dollars, depending on whether institutions reallocate resources or recognise incremental systems and staff resources.
Archiving costs are as yet unclear but institutions are advised to build an estimate of archiving costs into the repository budget "if only as a placeholder". In any case, the report concludes that costs are likely to be very much less than journal costs libraries now incur and over which they have little control.

Note: The JISC has funded the SHERPA project to investigate e-print repositories in the U.K. A study in potential archiving of e-prints is currently being undertaken and will report soon.

18. Kling, Rob, Spector, Lisa, and McKim, Geoff. The Guild Model. Journal of Electronic Publishing. 8 (1). August 2002
<http://www.press.umich.edu/jep/08-01/kling.html>

The Guild Publishing Model is offered as an alternative to the five models for organising scholarly communication, which are:
e-print repositories; free on-line access to peer-reviewed literature (e.g. Budapest Open Access Initiative, Public Library of Science); peer-reviewed pure electronic journals; hybrid paper-electronic journals (this is the most conservative and also the most dominant model to date) and author self-archiving. The Guild Publishing model is based on existing research manuscript model which is already active in many disciplines worldwide, e.g. computer science. Refers to the fact that publishing models which work well in one field (e.g. arXiv.org) do not necessarily work for another (e.g. biomedicine rejected arXiv model and modified it to become PubMed Central.). Open publishing models can be problematic for fields which have stringent quality control mechanisms. Shows 6 advantages and 3 limitations of GPM.
Advantages = 1. Local control. Is set up by academic departments and research institutes, based on local interest and available resources, 2. Relative ease of innovation. Not every research institute within the same institution may wish to set

up a GPM, 3. Quality. Legitimacy is based on the reputation of departments and research institutes that host the guild site (e.g. the Harvard Business School business research manuscript series), 4. Access. It can provide rapid access to scholarly literature. "Before long-term access becomes an issue, the materials must be able to be found by interested scholars", 5. **Economy. Costs are manageable for most research universities (though not negligible). Costs vary enormously depending on the amount of pre-existing infrastructure at any given institution**. 6. Compatibility with other publishing models. On the other hand, the limitations are 1. Quality indicators depend on reputations of sponsoring institute, so lesser known scholars and institutions producing interesting material might not transfer well to this model, 2. Access. Material needs proper bibliographic control. Also changed urls can cause potential problems. 3. Prior publication limitations. Some fields have strict prohibitions against any form of prior publishing for submission to their key journals.

19. AAAS Directorate for Science and Policy. (July 2002). Electronic Publishing in Science: Seizing the Moment: Scientists' Authors Rights in the Digital Age.
<http://www.aaas.org/spp/sfrl/projects/epub/epub.htm>

The report recommends that authors take control of their intellectual property by negotiating licensing arrangements with publishers designed to maximise access to and dissemination of their work. It argues that waiting for changes in copyright legislation will take too long, even assuming it is ultimately successful. The study was conducted by the American Association for the Advancement of Science (AAAS), with support from the NSF. The study draws attention to the exciting potential of electronic publishing to greatly accelerate the time taken to move from manuscript submission, peer review, and publication; the ability to reach a much larger audience; more powerful and creative capabilities for presenting and illustrating research findings. Against these advantages are the concerns of scientists that access to information is too concentrated in the hands of a few major publishers and that widespread and timely dissemination is being compromised as a result. The study urges authors to consider their guidelines for licensing agreements with publishers. These include, among other principles, the rights of authors to use their work in ways that maximise access to it by others for educational purposes, and also the concept of fair use.

Market Analyses and Commentaries

20. Office of Fair Trading (September 2002). The market for scientific and medical journals.
<http://www.oft.gov.uk>

This report suggests that competition may not be working effectively but assumes that market forces harnessing new technology may change this without the need for intervention. Examples of the development of the Internet enabling academics to 'bypass traditional commercial publishers altogether in some cases' are: SPARC's Journal of Logic Programming and Evolutionary Ecology, arXiv archives, the Electronic Society for Social Scientists, and the Berkeley Electronic Press. Reference is also made to the Public Library of Science, though there is no mention of the Budapest Open Access Initiative and it also fails to take into account the limited impact such initiatives have made thus far. However the report is dismissive of counter arguments from commercial publishers which try to justify the large price differentials. For example, Elsevier's argument that there has been substantial investment in development and delivery of electronic journals is discounted by the report which points out that the average prices of commercial journals appear to be

substantially higher than those of non-profit journals. Tables showing cost comparisons indicate an average commercial price of $487 compared to an average Society price of $229 and an average Educational publication price of $81. Studies of tests of value for money based on costs per page and costs per citation also indicate that for profit journals were five and a half times more expensive in terms of costs per page and over ten times more expensive in terms of costs per citation. The report also draws attention to the higher levels of profitability for commercial STM publishing, which is 10-15% above other forms of commercial publishing.

21. Morgan Stanley. (September 2002). Scientific Publishing: Knowledge is Power. <http://www.alpsp.org/MorgStan300902.pdf>

The main prediction of interest and concern is that benefits of scale will increasingly accrue to the large players and that libraries will trim smaller suppliers who are unable to offer bundled journals in the same way larger players can. Tables show that Reed Elsevier has a 23.3% market share of scientific publishing, ACS has 7.9% (the next biggest share). Including RE, there are 9 publishers dominating the market with 'others' representing 42.3%.

22. OCLC Library and Information Center Report. Five Year Information Trends. March 2003
<http://www.oclc.org/info/trends/>

The report 'provides a snapshot of information trends that will likely shape the information landscape of the future'. The report notes both the expansion of ePrint archives and the growth of online journals.

Appendix B

JISC Libraries Questionnaire

Summary of Responses

Background

JISC has funded a one-year consultancy to investigate archiving implications of licensed e-publications. Licensed electronic journals have been regarded as particularly vulnerable for some time and steps have been taken to address this. These include specific 'archiving' clauses inserted into the NESLI (now JISC) Model Licence, designed to provide both clarity and additional assurance to libraries as they increasingly depend on licensed access to e-journals rather than purchasing print journals.

The NESLI archiving clauses propose three options for continued access following termination of a licence:

- continuing online access from the publisher's or a third party's server;
- supplying an archival copy to the Licensee;
- supplying an archival copy to a central archiving facility operated on behalf of UK HE.

Since the third option does not yet exist, it is considered increasingly urgent to scope what is required to deliver this service to the community.

A sample survey of 26 HE and FE libraries who have been involved in 2002 NESLI deals was conducted, and also of 15 publishers currently participating in NESLI deals. Surveys were sent out during September 2002 and by the end of October, 19 out of 26 libraries had responded, providing sufficiently rich information to prepare this summary of the findings. Follow-up interviews were also carried out with 9 libraries.

Common Themes

All respondents reported an increasing trend towards licensing electronic journals, which most expected to continue into the foreseeable future. Several also referred to the practice of maintaining parallel print and electronic versions of journals as being unsustainable and several noted they were unhappy with deals which prohibit print cancellations. The most commonly cited reasons for moving to e-only were:

- convenience (24/7 access);
- space considerations (though one respondent noted that this is both long-term and difficult to quantify as a saving);
- improved access to a greater number of titles;
- increasing needs and expectations of users.

Against this, there were several factors referred to which, if not exactly inhibiting the move to e-only, are certainly causing considerable qualms about doing so. The poor quality of some aspects of certain e-journals, for example photographs, detailed diagrams etc. was mentioned. Technical issues were also commented on, for example some browsers perform differently with the same title. Pricing models were not always felt to be favourable. Some felt that it is more difficult to plan budgets as e-only licences rarely go beyond three years. Some users were also reluctant to move to e-only, and even those who are enthusiastic about electronic access do not necessarily want to abandon print. There are also differences between academic disciplines regarding the degree of take-up of e-only access. In addition, certain material of potential interest may be excluded from the e-version of journals, such as editorials, reviews, and obituaries. On the other hand, some e-journals might have additional information so there is an issue of version difference.

While not specifically given as inhibiting factors, several respondents referred to a range of additional operational snags or complications. These included not always being able to get usage statistics, confusion at the service level for users, examples included lack of clarity in some interfaces of material the user is entitled to access, titles being moved to another publisher or simply withdrawn without notice, and urls changing. A few reported user frustration with 'big deals'. This latter causes potential confusion and annoyance where the library has subscribed to these at the expense of some individual purchases which may be more valuable to users than some of the 'big deal' titles.

Far and away the two most cited concerns regarding moving to e-only were VAT, which means the electronic version is still more expensive, even where print cancellations are permitted, and archiving (expressed by both libraries and users). Some respondents expressed this as a positive quality they look for in deciding on deals, i.e if perpetual access is offered, preferably online, then that is seen as something which would strongly influence a decision to accept a deal. Almost all respondents indicated that this was an issue of great concern to them. Either the lack of an archiving policy/statement on continued access, or an ambiguous one, are factors which make the transition to e-only problematic for libraries.

NESLI Model Licence Archiving Clauses

The NESLI Model licence was regarded in a very positive light for a range of reasons, not least the reduction in administrative burden that a standard licence can provide. The inclusion of archiving clauses in the NESLI model licence was generally regarded positively, though some noted that the effectiveness of this as a means of protecting continued access into the future has yet to be rigorously tested so few respondents were completely reassured by them.

The lack of detail on how guaranteed access can actually be achieved was mentioned as a matter of concern for some. One respondent noted that the clauses are not particularly specific and pointed out that there are no

procedures in place which might provide the degree of assurance and security libraries need. Without such concrete evidence, some respondents found it difficult to place too much confidence in the clauses.

In terms of checking out this information in other licences, responses were divided. Some said they did not have the resources to check out this information. Others said they routinely raised this as an issue before signing any licence. Only one institution specifically said they will not sign a licence which does not include clauses relating to continued access to materials following termination, though many listed this as a major issue for them.

The reasons for signing a licence which does not contain archiving clauses can be characterised as a combination of pragmatism and pressure. Some feel dubious about the ability of publishers to deliver perpetual access, even where it is offered, and in any case, sometimes the wording is so vague it would be difficult, if not impossible, to pursue. Pressure from their user constituency to provide the content they are demanding can also leave librarians feeling they have little choice but to sign licences they are less than happy with.

In terms of the three options provided in the NESLI Model licence, by far the least favoured was the archival copy provided to individual institutions. Some respondents indicated that diverting access to a print copy (where available), or using ILL may be preferable to the operational difficulties of having to maintain an archival copy. CD-ROMs are particularly problematic and some institutions referred to a strategic move away from them. There are also clear inefficiencies in multiple institutions investing scarce resources in maintaining identical titles. The extent of redundancy inherent in the print environment is unnecessary in the electronic era.

This leaves continued access from the publisher's server, or from a central archiving facility (a service which does not yet exist). The former was seen to be fraught with dangers. While some felt that larger publishers might well have the necessary infrastructure to undertake archiving provisions, some worried that smaller publishers would not be in a position to do this. Even where publishers have made a commitment to archiving, the level of trust, particularly for commercial publishers, is not sufficiently robust to dispel anxiety. Some feared that publishers would exploit an increasing dependence on electronic access by escalating prices (and some felt this had already begun to happen).
The suspicion of publishers is not dispelled by extremely complex deals and even those which appear to offer promise of more flexibility needed to be very closely checked out. For example, one respondent reported that they had heard anecdotal evidence that one publisher's selective deal actually proved to be more expensive than their big deal.

The implications of mergers and publishers going out of business was also seen as a significant cause for concern in entrusting archival access solely to publishers. How can publishers commit to long-term archiving in such a fluid and rapidly changing environment? This suggests that the establishment of a

central archiving facility would be regarded as an increasingly urgent priority by many HE institutions and several respondents expressed this view.

Decision Making

Most institutions have to manage multiple licences, which adds to the administrative burden, particularly since some of the deals offered by publishers are extremely complicated. Several indicated their preference for the NESLI Model Licence. A standard licence is clearly much simpler to administer. One interesting prediction was the belief that more flexible models will need to develop. These would move away from 'big deals', and towards a "horses for courses" approach in which national deals might be made for core journals but other partners would be sought for more specialised purchases, including cross-sectoral consortia. The same respondent indicated that this could in turn lead to difficulties in having a model licence which can suit all of these needs, so that increased flexibility in deals might be gained at the cost of potential loss of clarity and standardisation in licences.

Several respondents referred to the tension between the desire to meet the needs of users and tight budgets. If users desperately needed the content, licenses could be signed which are less than optimum, either in terms of high costs and/or unsatisfactory archiving provisions.
The majority of respondents said they had not experienced difficulties following termination of a licence, though in most of those cases, there had not been experience of termination of a licence to really test this. Where there have been difficulties, e.g. "On several occasions, payments have not been recorded by publishers", they seemed to require considerable negotiation to rectify.

There were also problems of access to subscribed titles once cancellations of print subscriptions were made, requiring further negotiation. Several mentioned the need for negotiation in a range of circumstances, which is clearly time-consuming. One respondent indicated that they preferred to go direct to the supplier, rather than through a subscription agent, as they inevitably needed to become involved anyway and it often simply confused things further with an intermediary. One respondent also felt that publishers still do not understand the way libraries operate and therefore are unable to fit in with their timing and/or service requirements as well as might be hoped.

Business Models

When asked about specific experiences of and views on charging models for archival access, there was overwhelming dislike of a model which involves charging for content already paid for following termination of a licence. Objections to this model were partly philosophical (paying again for content already paid for under licence seemed inherently unjust). Partly perceived poor communication (some felt this particular model had been inadequately explained before being introduced), and partly practical (some felt it was easier to manage budgets with a one-off payment, and simpler to work out exactly what was being delivered for the money).

Some said they preferred a model used by a large commercial publisher which has one up-front fee. Even though the costs are greater than they would like, and can cause significant problems for libraries in finding such large amounts, these respondents felt they at least felt they knew exactly what they were getting for their money.

However several indicated that they were not opposed of paying towards the costs of archiving per se, some suggested a maintenance fee, as opposed to a content fee, may be acceptable.

JSTOR

15 of the 19 respondents subscribe to JSTOR and the comments are overwhelmingly positive here. A few referred to the expense of an annual, rather than one-off payment. One also mentioned the relatively small coverage and another thought the searching was a bit simplistic. But the positive comments far outweighed these concerns. Positive user response and acceptance, good value for money were mentioned a number of times. The fact that it is not-for-profit was seen as a definite asset, and it was generally regarded as very stable in terms of its archiving commitments though only one respondent actually indicated they were discarding print because of it. One mentioned that "Over the next few years it will start including access to titles we have already paid for electronically." which will be an interesting development. Mellon has funded JSTOR to develop their archiving service so it will be good to monitor this development closely. Clearly JSTOR has established excellent user confidence and satisfaction.

Central Archiving Service

Most were very much in favour of this. It would clearly allay fears about perpetual access and also an expressed concern regarding putting too much control in the hands of commercial publishers. Several respondents indicated the efficiency in having a central service and this seemed to make good sense. Specific comments included "We need the reliability such a service could offer", "There has been talk of an archiving service ever since PSLI but where is it?".

Almost everyone suggested start-up funding should come from UK Funding Councils. One felt that this might be contentious and suggested lottery funding as a possible alternative, also government funding, though this wasn't regarded as very promising! Someone else suggested RSLG as a potential source of start-up funds.

In terms of ongoing costs, most felt it was unrealistic to expect the publishers to contribute though some felt they should. One respondent noted the potential to save costs for the publisher that such an archive would provide, though it was assumed that this saving is unlikely to be passed on to the customers. Another comment was that it would take the responsibility from

the publishers but would give the library community the freedom to determine how the archive develops and is maintained.

Almost everyone indicated subscription by participating libraries would be acceptable, though most of these added the caveat, “as long as the subscription was modest”. Two respondents also said they would be opposed to a fee based on JISC pricing bands. A few mentioned the development of e-print archives and wondered if this could provide some prospects for future archival development.

Most acknowledge the difficulty of the task but it is clearly very high on the list of increasingly urgent priorities. Someone else said it shouldn’t be necessary to have separate archives for different sectors, for example business and commercial sectors also need archives, so it needn't necessarily be restricted to HE. Opening up the archive to other sectors could then lead to other potential sources of funding.

Some respondents also mentioned the possibility of a role for the British Library. In these cases, after follow-up interviews, respondents expanded on this to acknowledge that the BL can’t be expected to do everything but it would be very useful to have a clear steer from them of their archiving intentions.

Conclusion

Despite significant advantages and an increasing trend towards use of electronic journals, there remain several obstacles in the way of a wholehearted embrace of e-only access. Of these, lack of certainty in archiving is very high on the list of concerns. While most institutions are continuing with a strategic move towards e-only, they are doing so with an acute awareness of the potential dangers. The establishment of a central archiving service, as foreshadowed in the NESLI licence, would not remove all practical difficulties regarding e-journal licensing but it would certainly significantly reduce one key source of apprehension.

Appendix C

JISC Publisher Survey

A survey of publishers was conducted as part of the study. It was intended to focus on the publishers involved in 2002 NESLI deals. The objectives of this survey were:

1. Quantitative
- To calculate total volumes of currently archived electronic journals
- To establish the range of formats they are currently stored in

2. Qualitative
- To assess whether publishers have distinct archiving, storage and access arrangements for licensed material
- To assess whether publishers have rights to archive all content in materials they publish
- To canvass publisher's views and attitudes of on roles and responsibilities regarding archiving
- To canvass views and attitudes of publishers towards a potential central HE archiving service
- To identify any other related issues

The initial survey was emailed in September 2002 to fifteen publishers involved in 2002 NESLI deals. Only three completed surveys had been received by December 2002, despite reminder emails and an extended deadline. One publisher responded that they were unable to provide the information requested as provision of the online version of all of their journals was outsourced and there were insufficient staff from the customer service and marketing to undertake the research needed. Follow-up interviews were held with two publishers as a result of the first survey.

As this was such a disappointing response rate, a more generic survey excluding questions that related to specific arrangements in place for the Model Licence, was sent to members of the Publisher's Association which yielded an additional two responses. Two more publishers were contacted directly and agreed to a meeting, where they answered questions taken from the generic PA survey.

In all, data on volume, storage formats and general comments on the Model Licence were taken from seven publishers. Encouraging a response rate from publishers was very difficult, though the level of co-operation from those who did respond was very high. Reducing the number of questions to focus on the highest priority data was also undertaken in order to encourage a higher response rate.

Collation of responses

Q1 How many journals do you store in electronic form?
- **Number of titles to date**
- **Number of issues to date**

Responses:

Titles-
The total number of titles covered by the survey was 2,555 from all seven publishers but the number of titles held by individual publishers varied greatly, from 1 to 1,790.

Issues- The total number of issues covered by the survey was 254,000 from six publishers (one was unsure of the number). Once again, numbers varied substantially, from 12, to 200,300

Q2 What is the earliest date for which you hold an electronic version of a journal?

Responses:

In date order: 1823, 1874,1879,1954,1975,1996, 1998.

Q3 How many (if any) of your journal titles are stored *only* in electronic form, as opposed to parallel print and electronic?

Responses:

None, all exist in parallel print
O*nly* electronic
Very few - parallel print and electronic as a rule
None are purely electronic yet
None, all are also in print.

Summary. There are very few e-only at this stage, a total of seven titles from three publishers were reported as e-only, though one reported that this would increase "significantly" in the near future. Three publishers indicated that none were e-only at this stage, and the seventh publisher indicated "very few" but did not specify a number.

Q4 What volume of storage (in gigabytes or terabytes) is currently taken up by your e-journals?
If more than one copy is stored, please differentiate between:

a) single primary copies

b) multiple or back-up copies

Responses

a) **Single primary copies** (in ascending order of magnitude)

1 gigabyte
15 gigabytes
89 gigabytes
0.2 terabytes
2.3 terabytes
2.5 terabytes
Not sure

This indicates an estimated total of 5.1 terabytes of storage are currently being used by the six publishers who responded to this question. The six publishers have a total of 2,525 titles between them.

b) **Multiple or back-up copies**

No response or "unsure" from four respondents and N/A (print copy represents the archival copy) from a fifth respondent. The other two respondents indicated :

- 3 GB
- Back-up copies stored, three copies under direct control but numerous organisations are permitted to host content (though the latter does not provide as rich a functionality).

Additional information through a follow-up interview revealed that one respondent who did not respond on the initial questionnaire has disaster recovery plans in place - these are handled by a third party specialising in having a secure environment for digital materials. Duplicates of the publisher's journal titles are sent to this firm each week.

Q5 What volume of storage do you estimate you will have in:

a) 1 year from now

b) 5 years from now

Responses (interpolated with Q4 results to show growth pattern):

Current	**1 year on**	**5 years on**
1 gigabyte	5 gigabytes	13 gigabytes
15 gigabytes	15 gigabytes	20 gigabytes
89 gigabytes	120 gigabytes	200 gigabytes
0.2 terabytes	0.4 terabytes	1.5 terabytes
2.3 terabytes	8.6 terabytes	6.9 terabytes
2.5 terabytes	3.5 terabytes	5-6 terabytes
Not sure	Not sure	Not sure

Summary: Although one publisher indicated a reduction in volume of storage over five years (from 8.6 terabytes to 6.9 terabytes), I believe this was an anomaly but was unable to get a response when I queried the figure. However, this same publisher also estimated a massive increase in storage over a year (from 2.3 terabytes to 8.6 terabytes), which seems unusually large, even with the large-scale digitisation of backsets assumed. All other respondents indicated an increased estimate over the course of five years, in some cases dramatically so.

Note: It is interesting to compare these findings with the recent OCLC *Five-Year Information Format Trends* [1]. This report quotes Rick Anderson predicting that "five years from now, research journals will be published almost entirely online." It also shows data on the Electronic Journal Explosion provided by *The Library Journal.* This compares fewer than 75 peer-reviewed electronic journals in 1994, with 2002 figures indicating 75% of titles online in Science Citation Index, 63% online in Social Science Citation Index, and 34% online in Arts and Humanities Citation Index.

[1] OCLC Library and Information Center Report. Five-Year Information Format Trends. March 2003. <http://www.oclc.org/info/trends/>

The statistics are expressed as percentages for the 2002 figures. My calculations are that these represent 2755 titles for SCI, 1764 for SSCI, and 381 for AHCI, or a total of 4,900 online titles cited in these databases.

While generalisations are always fraught with danger, if we accept the data in Q4 relating to current primary storage, of 5.1 terabytes taken up by 2,525 titles, then 4,900 online journal titles might be expected to require c. 9.9 terabytes of primary storage, excluding back-up copies.

Further predictions of a steep rise in e-only journals were also made by EPS in their report for the JCVD on the impact of the extension of legal deposit. They predicted a rise of e-only serials of 122% between 2002 and 2003 and further steep rises in the short-term future[2].

Q6 What storage formats do you use? (e.g. TIFF/GIF for images, SGML/HTML for text etc.)

Summary of responses:

PDF, HTML, and XML were by far the most cited storage formats for text as was the case for TIFF and GIF for images. However there were numerous other formats referred to, pertaining to the specialised nature of the content, e.g. Chemintoch, Chemdraw, Movie Quicktime, MySQL, CAD, VR, VRML, to cite just a few. In a follow-up interview, one respondent indicated that they do not at this stage accept PDF as a format and their current policy is to convert all PDF documents to open standards but this is very time consuming. They would be very interested to hear of any work being done on preservation of PDF.

The 174 Academic Press titles hosted by Ingenta are predominantly either SGML or PDF. Ingenta estimates that it will probably be another five years before problems are likely to be experienced with changes in technology for PDF. This also reflects the fact that these journal titles, which 1996 and 1999, are likely to be relatively straightforward, compared to more sophisticated exploitation of the technology which is likely to increase.

Text formats are relatively simple to manage over time but the number of executable formats included, even in this small sample, suggests that there will be significant challenges in ensuring their continued accessibility over time. The observations of the Harvard Mellon project are relevant in this context:

> One of the great powers of digital journal articles is that they are not limited to linear text and static pictures …These supplementary materials represent a significant resource, but also a significant challenge to the Archive. In general, there is little control over the technical formats for supplementary files…[3]

[2] EPS. The Impact of the extension of legal deposit to non-print publications. Study report prepared for the Joint Committee on Voluntary Deposit. October 2002. p.59. <http://epsltd.com/ExtensionOfLegalDeposit.htm>

[3] Digital Library Federation. Report on the Planning Year Grant for the Design of an E-Journal Archive. Presented by the Harvard University Mellon Project Steering Committee and Technical Team to the Andrew W.Mellon Foundation. April 2002. <http://www.diglib.org/preserv/harvardfinal.html>

In response to a question on what represents the greatest challenge in terms of potentially high costs, Meg Bellinger also indicated that "The proliferation of digital object types represents a significant challenge. For each digital object type, one must potentially address a number of issues, including how to capture and preserve it." [4]

Q7 Do you have separate arrangements for the NESLI negotiated journal titles? (i.e. are there particular arrangements for those materials covered by NESLI archiving clauses 2.2.2, 5.4.1 and 5.4.2?)

Responses:

1. No.
2. No, all content is stored in secure storage as outlined in archiving policy.
3. Agreement with libraries means that they are given permanent access to issues subscribed to. Should the journal cease to trade the journal has an agreement with a third party to provide continuing access

[This question was not included in PA version.]

Q8 The NESLI licence includes clauses referring to continuing access to material which has been paid for during the term of the licence, after termination of the licence. Three suggested options in the NESLI licence for achieving this are:

a) Continued access from your own server
b) Supplying an archival copy to the Licensee in a mutually agreed format
c) Supplying an archival copy to a central archiving facility operating on behalf of UK HE

Which of these options do you prefer? Please comment

Responses:

There were varied responses from the publishers who answered to this question. One nominated c) as being "easy, simple, and centralised ". The same publisher also expanded on this at a post-survey interview to say that they prefer to send the files to users rather than have to continue to authenticate users. This respondent was also asked about options for contributing to the costs of a central facility undertaking archiving on their behalf. They expressed a willingness to negotiate on a range of options but suggested a reduction in the licence fee in exchange for their journals being archived might be one possibility.

Another publisher stated that a) was the only feasible option for them because of the server specific nature of much of the software which supports features in the journal. Text-only elements can be delivered to the Licensee.

One publisher indicated option a) could be supported "against an annual administration fee" and option b) would be used on termination of the licence, when a CD-ROM would be provided.

[4] The full report of the responses to a range of questions regarding OCLC's Digital Archive can be found at Appendix F

Another respondent felt that the current licence focuses on b) but plans are being developed [by the publisher] to offer different solutions.

Another respondent indicated that they would be happy to negotiate on any of the three but did not indicate a preference.

[This question was not included in the two meetings.]

Q8.1 Each of these options provides for continued free access to material which has already been paid for during the term of the licence. Do you have any comments on this?

Responses

1. "Fine with me, it's only fair"
2. No response
3. "No problem with it";
4. "Continued access would not necessarily be free, we already provide a great deal of free content online, we would negotiate".
5. "Plans are currently considered to launch a new service for continued archival access with full functionality. Details will be made available in Spring 2003."

This question was not included in individual publisher meetings, but one respondent offered an additional comment of relevance:
"We can make short to medium term commitments – possibly for as long as ten years, but can't commit to forever. There needs to be sensible discussion on who pays. We don't believe it's reasonable to ask the publisher to pay."

Q8.2 Options b or c imply negotiation between the publisher and the library/central archive regarding standard formats. Do you have any comments on this?

Responses

1. "Who has a crystal ball? It's the best we can do now. One suspects PDF and XML will be around for a long time";
2. No response.
3. No problem with html pages, but virtual reality and online databases with query screens have caused difficulties for us. Our content has been migrated three times to date and the look has changed marginally but the content has remained the same so far. Some VR require specific downloads for users to be able to view them. Upgrades of plug-ins often not backwardly compatible.
4. "This is currently being researched in the framework of our archiving agreement. In general, it would seem that SGML formats will be more complex to handle, while PDF formats are rather more standard."
5. "We feel it would be particularly useful to negotiate on c)".

This question was not included in individual publisher meetings but an additional comment of relevance offered during the meeting was: "A standard DTD (as proposed by the Harvard Mellon study) seems a sensible way to go – the problem is mainly with references. We have been involved in LOCKSS and Mellon projects but

the solution still seems some way off. Scalability and long-term costs are still issues."

Q9 Are there any journal titles for which you do not hold the rights to make an archival copy of the full content?

Responses

1. Perhaps four "third party" journals.
2. Yes, presently we don't have the archival rights for one journal but we are negotiating.
3. N/A
4. For journals owned by third parties, we generally only have archival rights for the period when we were/are the publisher.
5. For some society owned journals, we do not have the rights to make these available in electronic format.

[This question was not included in individual publisher meetings.]

Q10 Are there any parts of journals that you may not have archival rights for? (e.g. images)

Responses:

All respondents answered either "No" or "Not that we're aware".

Q11 If your organisation ceased publishing and/or was merged with another publishing organisation, how would institutions access the archival content they had paid for?

Responses:

1. We are making arrangements with 4-5 libraries.
2. As there are multiple copies of our journals stored at multiple locations, it is not envisaged that continued access will be a problem in the event of our organisation ceasing.
3. Arrangements are in place with a third party.
4. When we cease a journal we provide the new owner with all existing electronic material. In case of permanent service discontinuity, our library advisory boards would be called on to advise on the future of the electronic files. We also have other official archiving partners.
5. Many locally load our content so would have it anyway. Subscribers would also have the print version. We would make the content available on tape if possible. We are also participating in a trial with the BL for one e-only title.
6. We assisted in writing the ALPSP best practice guidelines for when a society journal moves to a new publisher.
7. We would assume the site would be carried on after take over by another publisher but we can't really guarantee this.

Appendix D

JISC LICENSING & ARCHIVING WORKSHOP

Report on the JISC e-journals Licensing & Archiving Workshop held on 17 February 2003

Background

The JISC e-journal archiving study was tasked to scope the feasibility of implementing archiving clauses in the JISC Model Licence as part of the National Electronic Site Licensing Initiative (NESLI) in the U.K. This is now widely used as the basis for negotiation and licensing between UK University and College Libraries and publishers. Clause 2.2.2 of the Model Licence assumes three possibilities, archiving by the publisher (or a third party nominated by the publisher), archiving by the Licensee (i.e. the individual libraries licensing the content), or archiving by a central archiving facility operated on behalf of UK HE. Following an analysis of NESLI licences, conducting a survey of libraries involved in e-journal deals, and researching related developments, the final event of the Study was an invitation only Workshop held in London on 17th February 2003.

The purpose of the Workshop was to explore various options for archiving e-journal content in ways which would provide the assurance libraries need before they can confidently move to e-only access, without undermining publishers' business models. This also included examining legal issues and business models. Delegates for the Workshop were selected from as wide an area of constituencies as possible, including not only library and publisher perspectives but also experience of licensing e-journals from both the operational and strategic levels. All members of the JISC Journals Working Group (JJWG), who have overseen the development of the Study, were invited to the Workshop, as were representatives of JISC Committee for Content and Services (JCCS) and JISC (Assisted) Bibliographic Data Services (JIBS). The Association of Learned and Professional Society Publishers (ALPSP) and the Publishers Association (PA) were also invited to send representatives. Tom Graham chaired the Workshop, both in his capacity as Chair of the Consortium of University Research Libraries (CURL) and Chair of the JJWG.

Workshop Report

An Executive Summary of the Workshop, with points of agreement and recommendations are provided below. The powerpoint presentations made at the workshop, the reports of three breakout sessions covering organisational models, legal issues, and business models respectively, are available to download under the Powerpoint Presentations and Breakout session reports section of the workshop report on the JISC website at:
http://www.jisc.ac.uk/index.cfm?name=pres_ejournals_report

Executive Summary

Major Points of Agreement

- Any solution *must* be collaboratively achieved between publishers and libraries (and possibly more widely than this).

- A distributed network of library and publisher repositories linked and co-ordinated by a central service(s) was seen as the most feasible implementation of a "central archiving facility operated on behalf of UK HE".
- Publisher's servers are probably the best means of gaining access to content in the short-to-medium term. Library repositories are the best means of ensuring the guarantee of access and the medium to long-term preservation of content.
- Another entity needs to be established as a central service to ensure ongoing long-term access and guarantees. A crucial part of the network would be a service to handle the rights management and ensure access is secure and abides by licence terms agreed between publishers and libraries.
- Content used by UK HE is often produced by international publishers. It was recognised pragmatically that responsibility for ensuring continued access will need to be nationally based but will need to operate in an international context.
- Responsibility for ensuring continued access independently is unlikely to be cost-effective and sustainable for individual libraries in the U.K.

Recommendation

Establish a Steering Group comprising representatives from libraries and publishers to undertake planning and advocacy for a centrally co-ordinated service. The service would operate on a not-for-profit basis and encompass the elements summarised by breakout groups 1 and 2 and consider a business model suggested by breakout group 3. It will be necessary to test the business case and model more rigorously during the planning period.

Responsibility: PALS (a joint JISC, ALPSP, and PA working group with participation from the British Library).

Timeframe: Steering Group to be established ASAP to ensure the momentum is retained. Timeframe for reporting to be negotiated within PALS but will need to be long enough to incorporate relevant progress in related developments but short enough to enable a funding proposal for 2004.

The Workshop

Twenty six invited delegates attended a one-day Workshop in London on 17th February 2003 aimed at deciding on a practical way forward in implementing the archiving clauses in the JISC Model Licence. It was recognised that this was potentially an important future issue for the proposed UK Research Library Network, for collaboration between the sector and UK copyright libraries and publishers, and with international initiatives. The morning sessions focussed on a series of presentations exploring the legal, organisational and business issues to be decided. Key issues arising out of these presentations were that compliance issues are extremely complex if handled retrospectively by an archiving service for material licensed to individual Higher Education Institutions. Organisational models considered the risks and benefits of different models, as well as the option of deferring action on the basis that it was too premature, given the early stage of archiving developments. The presentation on business models worked through a model which began with establishing the purpose and scope of the archive. The funding strategy was suggested as a combination of public funding, contribution in kind, and fees. All presentations reinforced the need for active co-operation and collaboration between libraries and publishers in finding a mutually acceptable solution.

Breakout sessions in the afternoon explored these issues in more detail and concluded that it is possible to embark on a collaborative planning process for a hybrid model. This model envisaged that a central service would provide a co-ordinating role in negotiation of guarantees and secure access and rights management mechanisms, but with delivery being distributed over a range of options, including the publishers' servers. The latter were generally seen as realistic options, at least for the short to medium term. The issues of rights management, trust, and mechanisms for achieving the guarantee of continuing access in such a system were however seen as critical. It was also recognised that this option would benefit from the rigour of a formal business modelling process, such as that outlined in the presentation on business models, together with studies to test and demonstrate the benefits and value proposition for publishers and libraries. It was recommended that a steering group be formed, under the overall direction of PALS (a joint JISC, ALPSP, and PA working group with participation from the British Library). This should develop the model to the stage where funding could be sought for the service, possibly under the aegis of a future Research Libraries Network. It was agreed that of the three options proposed by the Model Licence (Publisher archive, individual institution, and central archiving service), the publisher archive and archiving by the Licensee are not considered sustainable in the long-term. A modification of the central archiving service was proposed, which would centrally co-ordinate the service but provide distributed delivery.

Note: After the Workshop, it was interesting to note that the views of a sustainable solution for the UK seemed to closely match ideas published independently in the USA in discussions on the National Digital Information Infrastructure and Preservation Program (NDIIPP) plan and appendices. Available from: <http://www.digitalpreservation.gov>.

Neil Beagrie and Maggie Jones 28/2/03

Workshop Delegates

Toby Bainton	SCONUL
Ian Bannerman	Blackwell Publishing
Neil Beagrie	JISC
Richard Boulderstone	British Library
David Brown	British Library
Margherita Caccavale	EPS
Andy Crowther	Elsevier
Anne Foster	EPS
Emanuella Giavarra	CWG
Tom Graham	University of Newcastle
Alan Hopkinson	Middlesex University
Andy Hyde	Consultant
Simon Inger	Simon Inger & Associates

Karen Jeger	University College London
Clare Jenkins	Imperial College, London
Maggie Jones	JISC
Tony Kidd	University of Glasgow
Diana Leitch	John Rylands University, Manchester
Peter Morgan	University of Cambridge
Sally Morris	ALPSP
Tony O,Rourke	IOPP
Martin Richardson	OUP
Jean Sykes	London School of Economics
Malcolm Taggart	Keele University
Jill Taylor-Roe	University of Newcastle
Deborah Woodyard	British Library

Original Project Proposal

Archiving E-Publications: Outline of JISC Consultancy Requirements

1. Introduction

1. As a key funding body for both the provision and development of digital content for the DNER, JISC has a critical role to play in its long-term preservation and access in collaboration with a number of partners.

2. A current major gap exists in the archiving of and future access to, digital publications (primarily e-journals but e-books and abstracting and indexing services also need to be considered). The JISC has already implemented through the NESLI licence the recommendations of the JCEI interim preservation strategy to include clauses for the right to perpetual access and archiving in licensing electronic journals for the DNER. However earlier versions of the NESLI and PSLI licences do not contain these clauses. There is a pressing need to develop a strategy and evaluate services to secure practical implementation of the rights to perpetual access in these materials on behalf of the community (it is already clear that publishers themselves are not committed or cannot be expected to commit to the long-term archiving which will be necessary).

3. JISC has funded CURL through the Cedars project to address fundamental research issues in this area and Cedars has been particularly active and made a substantial contribution to developing work on preservation in digital libraries.

4. Over the last 12 months a number of e-journal archiving pilot studies have been funded by the Mellon Foundation in US universities and a number of potential business models are being explored with publishers.

5. The Digital Preservation Coalition was launched last year and copyright libraries and publishers together with JISC are involved in its activities. Publishers are becoming more aware of the issues and are increasingly involved in collaborations to explore potential solutions for long-term archiving. The Research Support Libraries Group (RSLG) has also recognised the requirement for future archiving to be a key issue.

6. The legal deposit libraries are exploring voluntary legal deposit (primarily of offline material) with publishers and legislation for electronic legal deposit is anticipated. However legal deposit of electronic materials will not provide a solution to all archiving issues. Many publications from international publishers will not be covered by legislation and continuing access by institutions in line with licensing agreements would need to be resolved. This said, there are clearly potential synergies that could be explored.

7. Input from members of the journals and e-book working groups, and Cedars management group on future developments to follow on from the Cedars Project and to address these archiving issues was solicited in a meeting during July last year and a report of discussion and recommendations prepared. With the conclusion of the Cedars project it is proposed that JISC funds a consultancy to

evaluate previous licences; explore with publishers and other stakeholders archiving and access provisions; and evaluate future options for archiving of licensed e-journals and access arrangements. In the first instance it is suggested that CURL would be invited to utilise existing expertise built up in the Cedars project and extend the contract for the project manager for 12 months to complete the consultancy.

8. This is not an area in which the UK HE/FE community is the sole player and collaboration with other partners particularly the national libraries and publishers is essential. A key part of the consultancy would be to scope opportunities for sharing development risks and costs with other funders.

2. Requirements

1. To compile and evaluate past NESLI and PSLI licences and identify where archiving provision has not been made.

2. To sample other institutional and consortial agreements for electronic publications made in HE/FE and their archiving provisions.

3. To provide an overall risk assessment in terms of rights, volumes and formats of material.

4. To identify legal issues and seek advice as required.

5. To evaluate lessons from the Mellon e-journal archiving pilots, the Cedars Project, and other national and international projects for potential models and solutions in archiving of e-publications.

6. To explore and identify with other stakeholders including publishers, aggregators, libraries, and the JISC Content team key requirements, options for addressing them, and potential collaborations/shared development risk. This should include both long-term solutions, financial models and any transitional arrangements which may be required, e.g. for NESLI.

7. Prepare a report with key findings and recommendations and draft implementation plan for discussion with JISC working groups and committees.

8. Oversee implementation of the report in conjunction with the JISC Executive.

9. A workplan will be agreed with the consultant and schedule for delivery of the report and implementation plan over 12 months. It is anticipated work would commence in May 2002 and would complete in April 2003.