

# Preservation, Trust and Continuing Access for e-Journals

Neil Beagrie

DPC Technology Watch Report 13-04 September 2013

Series editors on behalf of the DPC  
Charles Beagrie Ltd.



Principal Investigator for the Series  
Neil Beagrie



Digital **Preservation** Coalition

DPC Technology Watch Series

© Digital Preservation Coalition 2013 and Charles Beagrie Ltd 2013  
Published in association with Charles Beagrie Ltd.

**ISSN: 2048-7916**

**DOI: <http://dx.doi.org/10.7207/twr13-04>**

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, without the prior permission in writing from the publisher.

The moral right of the author has been asserted.

First published in Great Britain in 2013 by the Digital Preservation Coalition.



LIGUE DES BIBLIOTHÈQUES EUROPÉENNES DE RECHERCHE  
ASSOCIATION OF EUROPEAN RESEARCH LIBRARIES

This report has been endorsed by the LIBER Forum on Digital Cultural Heritage

## Foreword

The Digital Preservation Coalition (DPC) is an advocate and catalyst for digital preservation, ensuring our members can deliver resilient long-term access to digital content and services. It is a not-for-profit membership organization whose primary objective is to raise awareness of the importance of the preservation of digital material and the attendant strategic, cultural and technological issues. It supports its members through knowledge exchange, capacity building, assurance, advocacy and partnership. The DPC's vision is to make our digital memory accessible tomorrow.

The *DPC Technology Watch Reports* identify, delineate, monitor and address topics that have a major bearing on ensuring our collected digital memory will be available tomorrow. They provide an advanced introduction in order to support those charged with ensuring a robust digital memory, and they are of general interest to a wide and international audience with interests in computing, information management, collections management and technology. The reports are commissioned after consultation among DPC members about shared priorities and challenges; they are commissioned from experts; and they are thoroughly scrutinized by peers before being released. The authors are asked to provide reports that are informed, current, concise and balanced; that lower the barriers to participation in digital preservation; and that they are of wide utility. The reports are a distinctive and lasting contribution to the dissemination of good practice in digital preservation.

The report is published by the DPC in association with Charles Beagrie Ltd. Neil Beagrie, Director of Consultancy at Charles Beagrie Ltd, was commissioned to act as principal investigator for, and managing editor of, this Series in 2011 and has authored this report. He has been further supported by an Editorial Board drawn from DPC members and peer reviewers who comment on text prior to release: William Kilbride (Chair), Janet Delve (University of Portsmouth), Sarah Higgins (University of Aberystwyth), Marc Fresko (Inforesight), Tim Keefe (Trinity College Dublin), Andrew McHugh (University of Glasgow) and Dave Thompson (Wellcome Library).

## Acknowledgements

The author is grateful to William Kilbride for taking on the role of acting managing editor for this report, to Daphne Charles for reading drafts, and to attendees at the DPC workshop on e-journal archiving and colleagues who gave their time commenting and helping to shape the outline of the report's contents. Particular and substantial thanks are due to David Rosenthal and Vicky Reich (LOCKSS), Adam Rusbridge (UK LOCKSS Alliance), Liam Earney and Lorraine Estelle (Jisc Collections), Neil Grindley (Jisc), Kate Wittenberg (Portico), Marcel Ras and Barbara Sierman (KB), and Randy Kiefer (CLOCKSS). Over the years I have been privileged to work with a number of colleagues on e-journal archiving studies including Maggie Jones, Julia Chruszcz, Najla Rettberg, Mary Auckland, Terry Morrow, and Matthias Hemmje and staff at Globale Informationstechnik GmbH. This report has benefited from the experience and insights

gained on previous work with them funded by Jisc or the DFG. My thanks go to all of them for productive exchanges and insights on e-journal archiving. I wish to note that all opinions expressed in this report are ultimately the responsibility of the author, and do not necessarily represent those of funders, or colleagues.

## Contents

Abstract.....	1
Executive Summary.....	1
1. Introduction .....	3
1.1. Background .....	3
1.2. Key terminology and definitions.....	3
1.3. Key players .....	5
1.4. Lessons learnt .....	5
2. Issues.....	6
2.1. Print to digital transition .....	6
2.2. Structure of the publishing industry .....	6
2.3. Preservation .....	7
2.4. Licensing.....	9
2.5. ‘Trigger events’ .....	9
2.6. Journal transfers .....	10
2.7. Trust .....	10
2.8. Economic considerations .....	10
3. Standards and Best Practices .....	13
3.1. Object structures and metadata .....	13
3.2. Identifiers and identifier services .....	15
3.3. Best Practices .....	16
4. Services .....	18
4.1. Overview .....	18
4.2. British Library and UK and Ireland copyright libraries .....	18
4.3. CLOCKSS .....	18
4.4. KB e-Depot .....	19
4.5. LOCKSS .....	20
4.6. Portico .....	20
4.7. Consortial hosting .....	21
4.8. Keepers Registry .....	21
5. Use Cases .....	23
5.1 Use case 1: library cancels subscription to journal .....	23
5.2 Use case 2: library cancels subscription to a ‘big deal’ .....	23
5.3 Use case 3: e-journal or its past issues are no longer available from the publisher .....	24

5.4	Use case 4: publisher ceases operation .....	24
5.5	Use case 5: catastrophic failure of publisher's operations/servers.....	25
5.6	Use case 6: de-accessioning print .....	25
6.	Trust .....	27
6.1.	Background .....	27
6.2.	Trust in digital repositories .....	27
6.3.	Lessons learnt .....	28
7.	Horizon Scanning .....	30
8.	Conclusions .....	32
9.	Recommended Actions .....	33
9.1.	For academic libraries, library associations and library consortia.....	33
9.2.	For publishers.....	33
9.3.	For e-journal archives .....	34
10.	Glossary.....	35
11.	References and Further Reading.....	39

## Abstract

This report discusses current developments and issues which libraries, publishers, intermediaries and service providers are facing in the area of digital preservation, trust and continuing access for e-journals. It also includes generic lessons and recommendations on outsourcing and trust learnt in this field of interest to the wider digital preservation community. It is not solely focused on technology, and covers relevant legal, economic and service issues.

## Executive Summary

The report discusses the subjects of continuing access, trust and preservation for e-journals. These issues have become increasingly important for research libraries as published journals and articles have shifted from print to electronic formats; and as traditional publishing business models and relationships have undergone major transformations as a result of that shift.

It is important for readers to understand the significant implications for preservation and access of the different requirements (and terminology) that apply for e-journals: in particular the distinction made between continuing (sometimes also called post-cancellation or perpetual) access that applies only to subscription journals and securing long-term access for their subscribers; and long-term preservation that applies to both open and subscribed content. These differences lead to different types of service for e-journal archiving.

Libraries and publishers represent the key stakeholders although their professional organizations and scholarly societies also have a role, particularly in developing model contractual agreements and legislation. International initiatives and organizations such as LOCKSS, CLOCKSS, Portico and the e-Depot at the Koninklijke Bibliotheek (the Dutch national library) that have pioneered new forms of collaboration and explored issues of trust in third-party digital preservation solutions or in access services to digital journals and articles provided to libraries by publishers have also emerged.

Establishing assessment criteria and objective measures for trust in digital repositories has been the focus of a number of initiatives in recent years. These initiatives can help libraries, scholars, publishers and others to judge the reliability of the repositories and digital preservation services they may choose to rely upon or support. Two have been particularly prominent: the Trusted Repository Audit Checklist (TRAC) and its successor, the Trusted Digital Repository Checklist (TDR).

However, the archive repositories are only one element in the provision of long-term access to e-journals. Many of the same requirements for transparency and defined procedures for enacting agreements also apply between authors, libraries and publishers. Other initiatives such as the Transfer Code of Practice, the Keepers Registry, and the development of model clauses and enactment mechanisms in licences are also key components in trust in e-journal archiving. For trust to be established there need to be clear agreements for long-term archiving, and clear procedures and mechanisms for those agreements to be implemented and validated when necessary across all elements of the supply chain.

Throughout, this report seeks to place developments in e-journal archiving in their current context and to highlight emerging and on-going issues that may have an impact for many years to come. Two issues are worthy of particular iteration in thinking about likely trends impacting on e-journal archiving in the future.

Firstly, the transformation of e-journals continues and they are diverging from the 'electronic mirror copy of print' model prevalent in the early years. The atomic and static PDF files of the early e-journals days are steadily being replaced by bundles of dynamic and interdependent resources that are distributed across the web. There are more complex links to underlying data and other related articles and more functionality provided for accessing and interpreting the content itself. This trend to embrace fully the web as infrastructure for scholarly communication will require us to adapt future preservation efforts accordingly.

The second issue to note is the growing trend towards open access in e-journals and its implications. Challenges that arise include ensuring appropriate archiving permissions for open-access articles and funding of the preservation of open-access journals.

Continuing access and preservation of e-journals has involved initiatives in organizing multi-institutional collaboration, developing third-party services, establishing trust in long-term access and preservation between different stakeholders. The issues it has had to address go well beyond technology. Legal, economic and service developments are equally critical to its success.

In order to illustrate the potential challenges and issues discussed, six e-journal use cases are described in this report.

The report concludes with draft criteria for libraries for the selection of archiving solutions that can address their e-journal archiving needs and a set of recommendations for the different stakeholders (academic libraries, library associations and library consortia; publishers; and e-journal archives respectively).

The lessons learnt are of wider interest than the specific domains affected. For this reason we hope this *Technology Watch* report will be of value and general interest to a wide range of DPC members and non-members.



## 1. Introduction

### 1.1. Background

In January 2012 the DPC held a forum on Trust and e-Journals and this *Technology Watch Report* was commissioned and influenced in light of discussion by DPC members and others at that meeting.<sup>1</sup>

This *Report* discusses the subjects of continuing access, trust and preservation for e-journals. These issues have become increasingly important for research libraries as published journals and articles have shifted from print to electronic formats; and as traditional publishing business models and relationships have undergone major transformations as a result of that shift.

There have been many significant changes. These include moving from libraries purchasing and physically holding (and preserving) a paper journal locally (with multiple redundancy of copies between libraries), to renting (licensing) remote access to an electronic journal held on publishers' platforms that are often based internationally in other jurisdictions.

Alongside these changes has been a growing open-access movement for e-journal articles that seeks to remove the subscription charges for access. Subscription journals, open-access journals and hybrids of the two (either a mixture of open-access and subscription articles in a journal or a 'moving wall' to open access after a fixed period of time) provide a complex landscape against which preservation of, and long-term access to, e-journals needs to be considered.

This e-journal landscape continues to evolve as e-publishing itself begins to shift from static to dynamic content, and the importance of data and supplementary material linked to articles increases in major disciplines.

All these changes in turn have made preservation of e-journals more challenging, more international and dependent on others, and brought issues of trust to the fore. Trust in this context is not solely of technology for preservation, but negotiating rights (and retaining a record of them for future use), and having transparent information on what is being archived, how it is preserved, and how and when it can be accessed.

This makes e-journals one of the most dynamic and challenging areas of digital preservation, particularly in terms of business models and trust mechanisms for shared or out-sourced preservation services.

This report is therefore likely to be of wider interest in the digital preservation community, and not just to those involved in producing, accessing and preserving e-journals.

### 1.2. Key terminology and definitions

As this report is intended for a wide readership, it is important for readers to understand the significant implications for preservation and access of the different requirements (and terminology) that apply for e-journals: in particular the distinction made between continuing access (sometimes also called post-cancellation or perpetual access) that applies only to subscription journals and securing long-term access for their subscribers; and long-term preservation that applies to both

---

<sup>1</sup><http://www.dpconline.org/events/previous-events/837-trust-and-e-journals>

open and subscribed content. These terms were discussed in a Jisc *Briefing Paper* on e-journals archiving and preservation published in 2007 (Jones, 2007b). They are defined in this report as follows:

### **Continuing access**

Continuing access is most commonly associated with e-journal licence clauses designed to provide assurance of perpetual access to subscribed material in certain circumstances, including post-cancellation of subscriptions.

### **Long-term preservation**

Long-term preservation refers to the processes and procedures required to ensure content remains accessible well into the future.

Effectively, continuing or perpetual access is an attempt to replicate the situation with subscription paper journals where a library purchases issues of the journal for on-going reference, regardless of whether or not in the future the subscription is continued or the publisher changes. In an electronic environment this normally involves the publisher providing current and former subscribers with access to journal content hosted on their platforms and e-journal archives providing an option to guarantee that access and preservation of that content will continue in the long term.

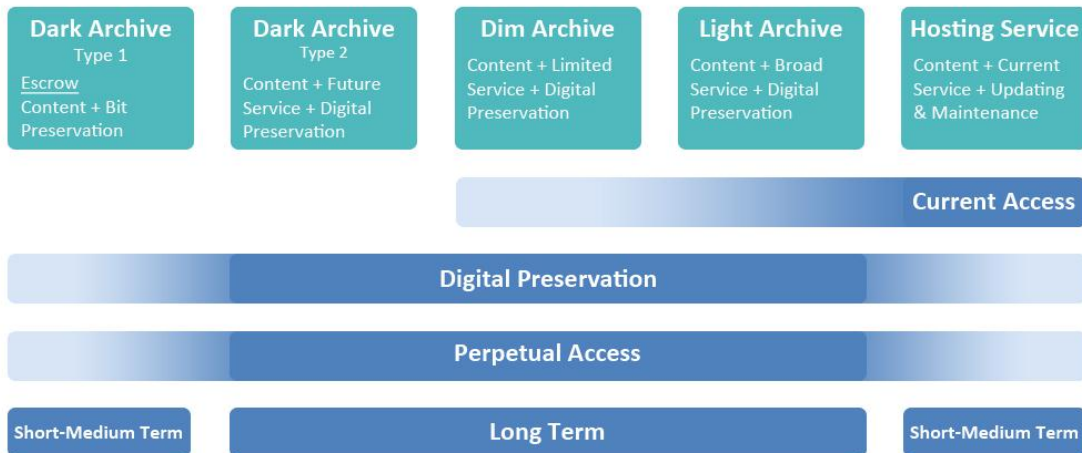
Long-term preservation of the content, on the other hand, is an issue affecting both open and subscribed journals. It is an issue not just for the subscribers, but for society as a whole (and particularly for national libraries), ensuring that the scholarly record continues to be accessible to future generations of scholars. This becomes a significant challenge as storage and delivery technologies evolve and develop.

The distinction is important as a commitment to preserve subscription journals for the long-term (as national deposit libraries do) usually involves restricted access rights for subscribed content. It will normally only be available within the deposit library. It does not necessarily provide mechanisms for protecting and enacting subscribers' rights for remote online access to that restricted content.

These changes in the roles and relationships of the stakeholders and the location of the electronic journal have had profound implications for long-term management of e-journals, requiring the establishment of new types of collaboration, and sometimes new types of organization, for preservation between libraries, publishers and academic societies.

A range of terminology has been used to describe preservation and access roles in this new landscape, from 'dark', 'dim' and 'light' e-journal archives to 'hosting service'. These and other key terms are defined in the Glossary (Section 10).

Various combinations of these options can be tailored to provide services for specific content, time horizons and functional requirements. Over time roles can transition, and may be strong or relatively weak at different points. Figure 1 aims to illustrate core roles (dark blue) and weaker roles (fading blue), and timescales for these different types of archive and services.



**Figure 1:** Concepts and properties of archives and hosting and their relationships. ©Charles Beagrie Ltd 2009. Creative Commons Attribution-Share Alike3.0

### 1.3. Key players

Libraries and publishers represent the key players, although their professional organizations, and scholarly societies, also have a role, particularly in developing model contractual agreements and legislation. International initiatives and organizations such as LOCKSS, CLOCKSS, Portico and the e-Depot at the KB (the Dutch national library) that have pioneered new forms of collaboration and explored issues of trust in third-party digital preservation solutions or in access services to digital journals and articles provided to libraries by publishers have also emerged.

### 1.4. Lessons learnt

The lessons learnt from these early initiatives in organizing multi-institutional collaboration, developing third-party services, establishing trust in long-term access and preservation between different stakeholders in e-journals, and enacting them in contractual agreements and procedures are of wider interest than the specific domains affected. For this reason we hope this *Technology Watch Report* will be of value and general interest to a wide-range of DPC members and non-members. These issues and lessons are explored in more detail in the subsequent chapters.

## 2. Issues

### 2.1. Print to digital transition

E-journals have become increasingly popular and have largely replaced paper journals as a mechanism for delivering research articles to the academic community. Although there are many obvious benefits that accrue from publishing and accessing academic journals electronically, there are costs and challenges which need to be addressed in maintaining access to them over the long term.

The transformation of e-journals continues, and they are diverging from the ‘electronic mirror copy of print’ model prevalent in the early years. There are more complex links to underlying data and other related articles, and more functionality provided for accessing and interpreting the content itself. Increasingly the electronic version should be seen as the ‘version of record’, richer and more complete than any paper copy.

E-journals perhaps more than any other content are evolving rapidly, raising new challenges in digital preservation and requiring on-going research and development to address them. They also stretch beyond national boundaries in terms of ownership and the location of the publishing platforms from which they are accessed, posing challenges for the drafting of legal deposit legislation and its transition to cover both electronic and print materials.

The transition also has implications for re-engineering library workflows. In the print world, selection, collection maintenance, and preservation were overseen by different units (each with distinct and complementary roles). In the electronic world, a much more integrated decision-making framework is required, for example, to factor in preservation issues at the point of selection (2CUL, 2011).

### 2.2. Structure of the publishing industry

The move to electronic publishing has also accelerated consolidation and a process of globalization in the publishing industry, particularly amongst the STM (science, technology and medical) publishers. Academic journal publishing currently encompasses a very large number of titles and publishers. In 2009 Swets estimated there were approximately 52,730 commercially licensed academic journals worldwide: of these 24,369 academic journals were available from large STM publishers and approximately 28,361 academic journals were from small STM publishers whose lists ranged from a few to one journal (Charles Beagrie Ltd and Globale Informationstechnik GmbH, 2010).

One of the challenges in digital preservation of e-journals is the wide difference in standardized workflows and formats, technical competence and innovation between the small group of very large publishers and the remainder produced by a long tail of over 700 publishers, many of whom are single title society publishers. This polarization has been observed as being a developing trend (Rightscom, 2008), and means that in relative terms, large publishers (or the large hosting platforms for publishers) are cheaper for services to archive than smaller publishers, due to easier set-up of archiving procedures and the economies of scale which apply.

Large publishers also have been able to offer 'big deals' in which e-journal subscriptions are bundled across their range of titles. Although access may be extended by a big deal, post-cancellation rights in the bundle are not treated equally: typically there is a core of subscribed titles which can have post-cancellation access rights, and the remainder of the bundled titles have current subscription access but no post-cancellation access rights.

There are also aggregators who offer current journal titles from many different publishers, often at attractive rates. However it is important to note aggregators cannot offer post-cancellation access: they are only suitable for current access requirements.

Finally, Open Access, the practice of providing unrestricted access via the Internet to peer-reviewed scholarly journal articles, is having a significant impact on the publishing industry. This in turn affects preservation of e-journals and the scholarly articles in them. Challenges that arise include ensuring appropriate archiving permissions for open-access articles and funding of the preservation, or selection for preservation of, open-access journals.

## 2.3. Preservation

There are perhaps three major issues to note: preservation methods; coverage; and business models. Business models are largely dealt with later under economic consideration (Section 2.7) but legal deposit is considered here.

### 2.3.1 Preservation methods

There are two broad techniques for preserving e-journals:

- The publisher supplies to the archive the source files from which the presentation files of a journal have been generated;
- The archive captures the presentation files from the publisher's website, including all the ancillary files that enable a browser to render the content.

There are advantages and disadvantages to each approach.

The advantages of source file preservation are that: it can include more content than appears in the journal presentation files (but note in some cases the delivery format and metadata agreed between the publisher and archive may vary from the actual source files and this may not be the case); the archive will retain these original source files but in addition may normalize these files to a standard archival format which it can subsequently manage and migrate or emulate over time; it does not require manifests and crawlers that need to be updated to accurately describe content boundaries for web content. The disadvantages are that: it may be incomplete if some material is missing in error; it requires a large upfront investment; and the presentation will almost certainly differ from that of the publisher.

The advantages of capturing presentation files are that it is possible to retain the look and feel of the publication, and initial costs are likely to be lower. The disadvantages of this technique are that: the archive must adapt to the publisher's website to ensure that the relevant content is captured; publishers typically change their websites more frequently than they change their source file

delivery processes; the archive must crawl the publisher's website carefully, limiting the rate of ingest at times of high usage, and avoiding crawler traps, etc.

These techniques also share some potential limitations: as content becomes more dynamic and web-like, each may struggle in different ways to keep up; both may require future format migration or emulations which are potentially challenging (de Boer and van Otegem, 2012; Morrow *et al.*, 2008, Rosenthal *et al.*, 2005).

### 2.3.2 Coverage

An important consideration for any library wishing to use a third-party service or to establish its own archive will be the coverage of titles it uses and how new titles will be selected and prioritized for archiving. Publishers and titles can be prioritized on the basis of size and value (in terms of subscriptions invested and number of users) or degree of risk of loss – typically these are diametrically opposite and the smallest or niche titles will be those at greatest risk (but least in demand). Coverage between the main archives has been shown to be quite variable. One recent study for the university libraries of Cornell and Columbia found that LOCKSS and Portico preserved only 15% of their e-journal titles. They noted, however, that although there is overlap in coverage of journals between LOCKSS and Portico, both services archive titles that are not preserved by the other (2CUL, 2011).

Another recent study compared the content of the LOCKSS, CLOCKSS and Portico archiving systems and found a significant overlap in participating publishers and journals. LOCKSS does however archive more small and arguably endangered publishers and it is suggested LOCKSS may be the only economically viable choice for them (Seadle, 2011a).

### 2.3.3 Legal deposit

The role of a national library is to ensure that the published heritage of its country is preserved and made accessible. An important vehicle in many countries for achieving this is legal deposit (UNESCO, 2000). There is a global trend towards extending legal deposit from the print environment to cover e-journals and other electronic publications. However, there are many practical challenges. Concepts, principles and practices that are accepted and understood in the print environment, such as publication, publisher, and place of publication, may have new meanings or no longer be appropriate in a networked environment (Muir, 2001). National deposit libraries in partnership with other libraries, publishers and technology vendors have been actively exploring these challenges and the technology issues involved. Legal deposit regulations for electronic publications were introduced in April 2013 in the UK (HMSO, 2013) after long periods of consultation and delays (DCMS, 2012).

Legal deposit legislation (or similar voluntary deposit arrangements) normally involves those subscription e-journals considered part of the national published heritage of that country. In addition it restricts to a significant degree off-site access to preserved electronic material for a substantial period of time to protect the commercial interests of the publisher. Typically this makes the national legal deposit collection unsuitable for meeting the international range of, and

‘perpetual access’ rights for, subscription e-journals licensed by other libraries and their users. One national library, the KB in the Netherlands, has taken the policy decision to archive e-journals that are within its national mandate and additionally a range of e-journals (including open-access titles in DOAJ –the *Directory of Open-Access Journals*) published beyond its borders.

## 2.4. Licensing

As noted in the Introduction, libraries paying subscriptions for licensing access to e-journals is a key component of the business model for many commercial publishers. Consortia representing libraries often play a very important role in negotiating model licences and terms. Being able to retain a record of these terms and entitlements for post-cancellation access can be very important for future access, particularly as ‘big deals’ may have different rights within them. A recent study by Jisc Collections has highlighted some of the challenges in maintaining records of such rights over time for both publishers and libraries (Jisc Collections, 2012).

The archiving and continuing or post-cancellation access clauses and detailing specific mechanisms for achieving them are also important. For one example of this, see the relevant clauses in the UK NESLi2 2013 Model Licence.<sup>2</sup>

Publishers may provide for post-cancellation access via their own hosting services and platforms, but subscribers remain concerned about their ability to maintain that access over the very long term. To address the specific issues of long-term preservation and continuing access for e-journals, and for commercial subscription e-journals in particular, a number of new archives have emerged. These seek to implement relevant clauses in licence agreements developed by publishers, libraries, and their professional organizations or consortia.

## 2.5. ‘Trigger events’

It is important for the publishers’ business models that the designated archives do not compete directly with the publishers’ own platforms. Similarly it is important for libraries to ensure continuity of access, regardless of events that may impact on the publishers over time. Significant effort has therefore been given to exploring ‘trigger events’ or transfers of ownership that may arise and procedures for them that are incorporated in the agreements between the publisher and the archive. The term ‘trigger event’ is used when specific conditions relating to an electronic journal title and its continued delivery are met. If the journal is no longer available to users from the publisher or any other source for a variety of reasons, then a trigger event is said to have occurred. This can set in motion access for users via an archive where the electronic journal may be digitally preserved.

<sup>2</sup> The clauses dealing with archiving and perpetual access are 5.1.5–5.1.9 inclusive. In addition in Schedule 3 Industry Standards the publisher agrees to archive the licensed material to ensure that it is preserved for future scholarship in at least one of the following archiving solutions: Portico, CLOCKSS or LOCKSS, and inform the institution in which of the archiving solutions the licensed material may be found. See <http://www.jisc-collections.ac.uk/Help-and-information/How-Model-Licences-work/NESLi2-Model-Licence/>



## 2.6. Journal transfers

In the electronic publishing world 'transfer' refers to the change of publisher or owner and the consequent roles and responsibilities which must be adopted by the 'transferring publisher' and the 'receiving publisher' respectively, to ensure continued access to content, both currently and in perpetuity, for the subscriber to that journal and their users. A Transfer Code of Practice has been created (see Section 3.3.1), now in Version 2 (UKSG, 2008), with a consultation underway during 2013 on an updated and extended draft version 3,<sup>3</sup> which seeks to set acceptable operational standards. Publishers are encouraged to sign up to this code and become Transfer Compliant.

## 2.7. Trust

From the above it can be seen that there is a complex set of inter-dependencies between different sectors and organizations in ensuring preservation of e-journals, and as a result achieving many of the benefits of electronic access to articles. That dependency is only viable when a significant level of trust can be established between the parties involved. The level of awareness and knowledge of digital preservation amongst library staff is also important (2CUL, 2011). Trust in digital preservation of e-journals is therefore multi-layered. It involves trust in different stakeholders as well as in the robustness of agreements and implementation procedures between them. Efforts to define trust in repositories and provide metrics for it are discussed further in Section 6.

## 2.8. Economic considerations

### 2.8.1. Economic considerations for research libraries

Libraries are facing increasing space pressures and funding constraints. There is a growing interest in moving more rapidly to e-only provision wherever possible, to help alleviate these pressures as well as to provide new electronic services to users. One of the most cited barriers and concerns both from library and faculty staff to moving to e-only has been sustaining and assuring long-term access to electronic content.

A major incentive for institutions to invest in e-journal archiving has been the cost and service benefits which arise from moving to electronic copies of journals. Broadly these fall into two complementary parallel areas:

- The benefits arising from transitioning from print or print + electronic to electronic-only for current journal licensing; and
- Benefits arising from the purchase or licensing of past electronic issues and/or retro-digitized versions of historic print journals.

For libraries an electronic copy of a journal offers significant potential cost savings over its paper equivalent as a result of:

---

<sup>3</sup> See <http://www.uksg.org/transfer/code>



- Reducing library staff resources and cash budgets required to process print copies of journals on receipt, such as making changes to current issue displays, spine labelling, bar coding, inserting and applying bookplates, binding, and shelving;
- Reducing library staff resources required to carry out stack maintenance, such as shelf reading and maintenance, collection shifting, collection weeding and cleaning;
- Reducing library staff resources required to assist users to access journals by providing direct access through local tools such as online public access catalogues;
- Reducing library space rental costs by eliminating the need to store print journals;
- Freeing-up core library space used for low-access print runs of journals for other library needs (and avoiding the capital and recurrent running costs of new library extensions);
- E-only providing better opportunities for collaborative purchasing than print + electronic.

These savings are very substantial but do require some off-setting of costs for them to be realized:

- Investment in e-journal archiving solutions to provide additional guarantees and mechanisms for continuing/perpetual access to licensed e-journals. Typically, annual costs for a solution will be relatively modest;
- Continuing access rights may have an associated cost detailed in the licences: for example, that an institution must remain a subscriber to the service in order to maintain access to journals that it previously subscribed to; or that there will be an access cost or a 'minimal' cost to maintain continuing access on cancellation (but often no indication is given of what this will be);
- Possible additional IT hardware and software, online user management, and maintenance;
- Within the European Union, higher Value Added Tax (VAT) charges incurred by libraries arising from a movement away from print journals (which incur lower VAT charges);
- Higher costs arising from implementing and monitoring more complex local access terms and conditions in electronic licensed products than print;
- Printing costs incurred by users in libraries and/or at home.

Other studies have estimated potential cost savings from moving to e-only and from rationalization of print collections (CEPA, 2008; Charles Beagrie and GlobaleInformationstechnik GmbH, 2010). They suggest that the costs of e-journal archiving solutions can normally be absorbed into core cash budgets from savings in print journal costs such as binding. There are then significant cost reductions or cost containment arising from released/reduced print storage that underpin the economic case for moving to e-only and investment in e-journal archiving solutions as part of this.

### 2.8.2. Economic considerations for deposit libraries/service providers

Publishers don't all present their electronic offerings in the same way. For a British Library e-journal mapping study (Rightscom, 2008), the authors examined the publisher outputs and metadata that one aggregator (Ingenta) had to deal with. They found that a number of different processes were involved, 'depending on the technical sophistication of the publisher and the level of service being purchased'.

In summary, any aggregation process for preservation of e-journals is potentially complex and labour-intensive. This effort is not short-term and one-off, but on going as journal issues are published and processed and publishers merge or sell titles.

As noted in Section 2.3.2 Coverage, technical complexity and cost cause serious dilemmas as it is difficult to obtain economies of scale or a critical mass of content on constrained budgets.

Any complexity in the copyright of articles also means relevant preservation activity scales badly. Even for open-access articles and journals preservation will not scale, unless there is a consistent granting of rights that allows archiving without negotiation (such as use of Creative Commons licences), because the cost of negotiating individually with copyright holders is prohibitive.

Finally, long-term economic sustainability remains a fundamental challenge for e-journal archiving services. As noted by the Blue Ribbon Taskforce (BRTF) on Sustainable Digital Preservation and Access, the collective interest of higher education means that the scholarly record benefits everybody. But with digital information such as e-journals there is little incentive for an individual institution to preserve this record, and strong incentives to wait for some other institution to do it. Thus is born the free-rider problem (BRTF, 2010, 53). The problem is compounded by lack of awareness and understanding. Archive services can face a significant challenge in encouraging a wide base of library membership that will enable their sustainability.

### 3. Standards and Best Practices

This section discusses some of the key standards and best practices that are influential in the field of e-journal preservation. It is selective and not intended to be comprehensive of all standards and best practices in digital preservation more generally, many of which are discussed in other reports in the *DPC Technology Watch Report* series. In particular, readers are referred to the report on web archiving (Pennock, 2013) for discussion of harvesting and other standards and best practices relevant to web content and the presentation files for e-journals from publishers' websites.

Discussion is focused here on three areas: object structures and metadata; identifiers and services; and formally defined best practices. Standards for trust and repositories and how these have been implemented for e-journal repositories are discussed in Section 6.

#### 3.1. Object structures and metadata

As journals moved from paper to electronic formats they closely mirrored the look and structure of each other. Publishing and preservation efforts towards maintaining that static visual appearance of e-journals have focused on the use of Portable Document Formats (PDF). As e-journals have evolved and become far more numerous and dynamic there has been increasing use of Extensible Markup Language (XML) for both publishing and preservation of e-journals. XML is a markup language that defines a set of rules for encoding documents in a format that is both human-readable and machine-readable. That machine-readability of XML over PDF is becoming a significant advantage for many users in science, technology, engineering and medicine<sup>4</sup> and an important issue for e-journal archives.

##### 3.1.1. PDF and PDF/A

Adobe's Portable Document Format (PDF) was first released in the early 1990s as a way to publish and share documents in a manner that maintains their static visual appearance independent of the tools and systems used for rendering the files. At first solely a proprietary format, it was released as an ISO open standard in 2008. Versions of the PDF standard known as PDF/A have defined a PDF file format standard for the long-term archiving of electronic documents. They are intended as a subset of PDF, leaving out features not suited to long-term archiving. The most recent version of PDF/A version 3 (ISO, 2012a) however, has been received with some caution by the library preservation community.<sup>5</sup> A forthcoming *DPC Technology Watch Report* devoted to PDF/A is expected to be published in 2014.

<sup>4</sup> e.g. <http://blogs.ch.cam.ac.uk/pmr/2008/05/03/the-merits-and-demerits-of-pdf/>

<sup>5</sup> e.g. Butch Lazorchak <http://blogs.loc.gov/digitalpreservation/2012/11/all-in-embedded-files-in-pdf/a/>

### 3.1.2. XML and the Journal Article Tag Suite (JATS)

XML-based formats have become the default for many office-productivity tools, including Microsoft Office (Office Open XML).

The XML specification defines an XML document as a well-formed text - meaning that it satisfies a list of syntax rules provided in the specification. In addition to being well formed, an XML document should be valid. This means that it contains a reference to, and conforms to, a schema. At first this was a Document Type Definition (DTD). A newer schema language is XML Schema, often referred to as XSD (XML Schema Definition). XSDs are far more powerful than DTDs in describing XML and allow for more detailed constraints on an XML document's logical structure.

The Journal Article Tag Suite (JATS) provides a set of XML schema modules that defines elements and attributes for describing the textual and graphical content of journal articles, as well as some non-article material such as letters, editorials, and book and product reviews. The standard consists of element and attribute descriptions, and three journal article tag sets: the Archiving and Interchange Tag Set; the Journal Publishing Tag Set; and the Article Authoring Tag Set (Beck, 2011; NISO, 2012). It has evolved out of the US National Library of Medicine (NLM) DTD Suite, created when the NLM found it needed a common format with a single Document Type Definition (DTD) in PubMed Central for exchanging data between publishers and archives. The first version of the NLM DTD was made available to the public in early 2003, and in 2006 the British Library and the US Library of Congress also announced their support for the migration of journal content to this standard. It quickly became the de facto standard for tagging journal articles in XML, even outside the NLM. As usage grew, users and potential users started asking about formalizing the article models as a standard with the National Information Standards Organization (NISO). Work on the NISO standard began in late 2009 and was finalized and released in August 2012 as NISO JATS: Journal Article Tag Suite, version 1.0 (NISO 2012).

### 3.1.3. ONIX for Preservation Holdings (ONIX-PH)

ONIX (ONline Information eXchange) is an international standard for representing book and serial product information in electronic form.

ONIX for Preservation Holdings (ONIX-PH) is an XML message structure conveying information about holdings of e-resources that have been preserved or that are the object of preservation initiatives. ONIX-PH communicates snapshot holdings statements and indicates the status of preservation work on each release of the e-resource covered. Structurally, the message is a specialized development of the ONIX for Serials SOH (serial online holdings) format and it utilizes the ONIX for Serials Coverage Statement to permit complex, multilevel communications of coverage and enumeration.

ONIX-PH may be used between publishers, preservation agencies, those maintaining registries of preserved resources, and other intermediaries in the e-resource supply chain. Partners in the Jisc-funded PEPRS (Piloting an E-journals Preservation Registry Service) project were instrumental in helping define the business requirements for ONIX-PH, and pilot exchanges focus initially on messages from preservation agencies to populate the Keepers Registry (see Section 4.8), hosted by EDINA and developed in partnership with the ISSN International Centre. ONIX-PH is currently

available in a draft v0.21 version and is undergoing pilot exchanges. On conclusion of the pilot, the message will be modified as necessary and released in a formal v1.0.<sup>6</sup>

### 3.1.4. Ajax and HTML

Hyper Text Markup Language (HTML) for displaying text and images in a web browser, has long been the standard for web-based interaction. Ajax (Asynchronous JavaScript and XML)<sup>7</sup> has become an increasingly popular design approach and a set of techniques for delivering a more interactive, desktop-like user experience for web applications in popular HTML browsers. The current version of the HTML standard (HTML4) is expected to be formally replaced by HTML5 (currently going through the official candidate adoption process) in 2014. HTML5 aims to improve HTML with support for the latest multimedia and devices. It extends and rationalizes the markup available for documents; it can handle multimedia and graphical content without having to resort to proprietary plugins; and it is being built to run on low-powered devices such as smartphones and tablets (W3C, 2012). These developments are symptomatic of how the web and its technologies, and e-journals along with them, are moving away from static to more dynamic content and raising new preservation challenges as they do so.<sup>8</sup> For wider discussion of web archiving and related technical issues see the companion *DPC Technology Watch Report* on web archiving (Pennock, 2013).

## 3.2. Identifiers and identifier services

The concept of the Persistent Identifier has been developed to provide a way of directing access attempts via a lookup process so that the reference for electronically published articles or datasets remains constant, even if their location changes over time. For e-journals the most widely adopted persistent identifier has been the Digital Object Identifier (DOI, see Section 3.2.1). DOIs are also becoming increasingly common for citation of research datasets. Another critical identifier for e-journal preservation services has been the role of ISSNs (see Section 3.2.2) as unique identifiers for serials and different media versions of them.

### 3.2.1. Digital Object Identifier (DOI)

The Digital Object Identifier (DOI) is a managed service for persistent identification of digital content. The identifiers (DOI names) resolve to locations specified by the registrant. An extensible metadata model is used to associate descriptive and other metadata elements with the DOI name. The DOI system is implemented through a federation of registration agencies, under policies and common infrastructure (Crossref) provided by the International DOI Foundation (IDF).<sup>9</sup>

<sup>6</sup> <http://www.editeur.org/127/ONIX-PH/>

<sup>7</sup> <http://www.openajax.org/index.php>

<sup>8</sup> For example, see David Rosenthal <http://blog.dshr.org/2011/08/moonalice-plays-palo-alto.html>

<sup>9</sup> <http://www.doi.org/>

### 3.2.2. International Standard Serial Number (ISSN)

An International Standard Serial Number (ISSN) is a unique eight-digit number used to identify a print or electronic periodical publication. Periodicals published in both print and electronic form may have two ISSNs, a print ISSN (p-ISSN) and an electronic ISSN (e-ISSN or eISSN). The linking ISSN or ISSN-L enables collocation or linking among different media versions of a continuing resource. ISSN codes are assigned by a network of ISSN National Centres, usually located at national libraries and coordinated by the ISSN International Centre based in Paris. The International Centre is an intergovernmental organization created in 1974 through an agreement between UNESCO and the French government. The International Centre maintains the ISDS Register (International Serials Data System), otherwise known as the ISSN Register, a database of all ISSNs assigned worldwide.<sup>10</sup> ISSNs and the ISSN Register have proved to be critical elements of the Keepers Registry (see Section 4.8), and their role in preservation activities are discussed in detail by Burnhill (Burnhill, 2013).

## 3.3. Best Practices

### 3.3.1. Transfer Code of Practice

The Transfer Code,<sup>11</sup> currently in version 2.0 (with a consultation underway during 2013 on an updated and extended draft version 3), establishes a set of standards that apply whenever a journal is transferred from one publisher to another. Publishers who publicly sign up to the Code and who apply it in practice are considered 'Transfer Compliant'.

As a voluntary 'best practices' code for industry participants, the Code of Practice does not supplant contractual terms, or intellectual property rights. In all journal transfers, contract terms negotiated between the parties involved in the transfer will govern. However, publishers that sign up to this Code are electing to use commercially reasonable efforts to ensure that the contracts governing journal transfers to which they are a party are consistent with the Code.

The UK Serials Group (UKSG) operates a free Enhanced Transfer Alerting Service that can be used by publishers and libraries. It is designed as a basic alerting mechanism to facilitate communication to the serials community of journal transfers that are planned or underway.<sup>12</sup> More recently, to overcome possible delays in notifying transfers and also the effort required for manual tracking, the KB and a group of research partners have been investigating automated preservation watch for monitoring of publishers and journals.<sup>13</sup>

### 3.3.2. Recommended practices for online supplemental journal article materials

<sup>10</sup> <http://www.issn.org/>

<sup>11</sup> <http://www.uksg.org/Transfer/Code>

<sup>12</sup> <http://www.uksg.org/transfer/notifications>

<sup>13</sup> Marcel Ras pers comm 6<sup>th</sup> June 2013. A paper *Automatic Preservation Watch using Information Extraction on the Web* will be presented at the iPRES 2013 conference.

Supplemental materials are increasingly being added to journal articles, but until now there has been no recognized set of practices to guide in the selection, delivery, discovery, and preservation of these materials. Ensuring effective access, use, and long-term preservation of supplemental materials to journal articles requires up-front planning about persistent identifiers, metadata, file formats, and packaging. This joint project from NISO and NFAIS (the National Federation of Advanced Information Services) developed a Recommended Practice for publisher inclusion, handling, display, and preservation of supplemental journal article materials. The Recommended Practice on Online Supplemental Journal Article Materials, a metadata schema, a tag library, and tagged examples are available from the NISO website.<sup>14</sup>

### 3.3.3. KBART and OpenURLs

Link resolvers based on the OpenURL standard are tools that help library users connect to their institution's electronic resources. The data that drives these tools are stored in knowledge bases. Libraries build, purchase or license link resolver knowledge bases in order to maximize efficient access for their users to an appropriate copy of all the electronic resources they license. Link resolvers play a critical role in the delivery of the appropriate copy to end-users of content in a networked environment, particularly in situations such as post-cancellation of an e-journal subscription or content no longer being available from the original publisher. In 2006, UKSG commissioned a research report that identified and described a range of problems affecting the efficiency of OpenURL linking in existing knowledge bases. The report (SIS, 2007) recommended the creation of a group that would determine and promote 'best practice' solutions for the community. The KBART (Knowledge Bases And Related Tools) working group established by NISO and UKSG to implement the report comprises representatives from publishers, libraries, link resolver and ERM vendors, subscription agents, and e-journal archives. It published its first recommendations on best practices in 2010 (KBART, 2010).

---

<sup>14</sup> <http://www.niso.org/workrooms/supplemental>



## 4. Services

### 4.1. Overview

This section provides a concise overview of the main preservation and continuing access services available for e-journals. Details of the British Library and UK copyright libraries are included for a UK audience.

### 4.2. British Library and UK and Ireland copyright libraries

The British Library is one of the world's largest research libraries, active within the UK and internationally. It is part of a network of six legal deposit libraries in the UK and Ireland (British Library, National Library of Scotland, National Library of Wales, University of Cambridge, University of Oxford, and Trinity College Dublin).<sup>15</sup>

These libraries have built a secure, resilient and scalable storage system for digital content with four mirrored storage nodes at the British Library (London and Boston Spa), the National Library of Scotland and the National Library of Wales. It is expected access to stored material will be possible from all four nodes and from an additional three 'access nodes' at Trinity College Dublin, and the Universities of Cambridge and Oxford.

The British Library has undertaken studies and pilots with ingesting e-journals from a range of publishers as part of a voluntary legal deposit scheme (Rightscom, 2008), and recently entered into an agreement with Portico to provide an e-journal processing service to support the Library's legal deposit requirements.<sup>16</sup> As noted in Section 2.3.3, the UK government introduced legal deposit for e-journals and other electronic publications via the Legal Deposit Libraries (Non-print works) Regulations in April 2013. Legal deposit restricts public access to deposited commercial works to the reading rooms of the deposit libraries, and therefore these will be 'dim archives' designed to support long-term preservation rather than continuing access for e-journal subscribers.

### 4.3. CLOCKSS

CLOCKSS (Controlled LOCKSS)<sup>17</sup> was launched in 2006 and is a not-for-profit collaboration between libraries and publishers. It is a dark archive based on the LOCKSS software (see section below on LOCKSS) in which a limited number of libraries take on an archival role on behalf of a broader community. It provides insurance to libraries that the e-journal and other content they have subscribed to will be preserved for the long term. It is described as a 'private LOCKSS network'. CLOCKSS is now a standalone not-for-profit company owned by librarians and publishers. The archive ingests and saves either source or presentation files, as the publisher chooses. Allowing access to the archive requires approval from a CLOCKSS board consisting of representatives from

<sup>15</sup> <http://www.bl.uk/aboutus/stratpolprog/legaldep/>

<sup>16</sup> <http://pressandpolicy.bl.uk/Press-Releases/Path-breaking-partnership-helps-national-library-preserve-thousands-of-e-journals-for-future-generat-63e.aspx>

<sup>17</sup> <http://www.clockss.org/clockss/Home>



publishers and libraries. Access to preserved content is only sanctioned when it is no longer available from any publisher; the Board of Directors then agrees to make it freely available to everyone (not just previous subscribers) under a Creative Commons licence. CLOCKSS is a solution to the problem of long term preservation. However, it doesn't provide post-cancellation access. As a 'dark archive' it has been able to recruit major publishers such as Elsevier, Taylor & Francis, and Wiley-Blackwell. Twelve libraries act as archive nodes in eight different countries.

CLOCKSS is supported financially by publishers and libraries, but the Board is working to raise an endowment in order to make the service less dependent on subscription income. Supporting libraries are currently asked to contribute on a sliding scale of between \$450 and \$15,000 per annum depending on the size of their library materials budget. UK libraries have access to a consortial discount with CLOCKSS brokered by Jisc Collections. Publishers are currently asked to contribute on a sliding scale of between \$200 and \$25,000 depending on the size of their total publishing revenue.

A TRAC preservation audit of CLOCKSS is being undertaken between September 2013 and May 2014.<sup>18</sup>

#### 4.4. KB e-Depot

The Koninklijke Bibliotheek (KB) is the national library of the Netherlands and operates the e-Depot,<sup>19</sup> its archive for the Dutch national deposit collection of electronic publications and other e-content (e.g. Dutch websites or master image files from KB digitization projects). The e-Depot became operational in January 2003 and focused initially on Dutch material. Recognizing the international nature of journal publishing (approximately 50–60% of all Science, Technology and Medical publishing is based in the Netherlands), this was extended during 2003 to international e-journals from the major publishers.

The KB intends to conclude archiving agreements for all the journals from 80 of the world's largest publishers and around 80% of open access publishers. Publishers wishing to make use of the services provided by the e-Depot are required to conclude an archiving agreement with the KB and to deliver bulk content and specified metadata. The primary deposit file formats are PDF and XML.

The e-Depot could be considered a 'dim archive'. Generally, end-user access is restricted to on-site perusal at the KB for reasons of private research only and online access is denied. The archiving agreement specifies a number of trigger events when wider access would be permitted. The e-Depot does not currently provide for post-cancellation continuing access by licensees of the content. Note however, full online access is already granted to publications by open-access publishers such as Biomed Central and the content of publishers listed in the DOAJ.

The e-Depot is an intrinsic part of the Dutch national library, and therefore the Dutch government is the major funder of both the e-Depot and the R&D efforts for digital preservation at the KB. However, the KB intends to develop a sustainable business model for the international e-Depot which will reflect both public and private responsibility for digital scholarly and cultural heritage. The new business model is based on annual contributions from publishers and libraries, calculated on the publisher's annual revenues or the library's materials budget. The KB will introduce this

<sup>18</sup> [http://www.researchinformation.info/news/news\\_story.php?news\\_id=1336](http://www.researchinformation.info/news/news_story.php?news_id=1336)

<sup>19</sup> <http://www.kb.nl/en/expertise/e-depot-and-digital-preservation>

model and contribution system in two phases: first starting with the publishers' contributions; when coverage is high enough, the KB will then start subscriptions for libraries.<sup>20</sup>

## 4.5. LOCKSS

LOCKSS (Lots of Copies Keep Stuff Safe)<sup>21</sup> provides libraries with open-source tools and support so they can take local custody of a wide variety of materials, including subscription and open-access scholarly assets (books, journals, etc.). LOCKSS-participating libraries harvest the web-published files from participating publishers. Readers access LOCKSS preserved content whenever (and for whatever reason) the material cannot be viewed on the publisher's (or intermediary's) servers. This includes, for example, short-term network problems. The highly distributed nature of this approach aims to ensure that there is sufficient replication to safeguard content despite any potential disasters which might befall individual LOCKSS institutions. The LOCKSS Program commenced in 1999.

Users of LOCKSS have the option of joining the LOCKSS Alliance which has a scale of fees, depending on the size and nature of the institution. US academic library fees currently range from \$2,236 to \$11,180 per annum. Participant fees support on-going technical development as well as regular monitoring and tuning of LOCKSS Alliance boxes. In the UK a two-year pilot has led to a UK LOCKSS Alliance with membership fees and a locally based support service. Current UK LOCKSS Alliance fees range from £1,800 to £5,000 per annum.

LOCKSS is currently being used to preserve content in two distinct types of environments. The Global LOCKSS Network holds and preserves material of general interest to a wide community. Publishers are not charged for participation in the Global LOCKSS Network. Private LOCKSS Networks hold and preserve materials of interest to targeted communities.

A pilot Trusted Repository Audit Checklist (TRAC) preservation audit of LOCKSS was completed in 2006–7 and the audit report is publicly available, although somewhat out of date (CRL, 2007).

## 4.6. Portico

Portico,<sup>22</sup> launched in 2005, is designed specifically as a third-party service for scholarly literature published in electronic form and provides three specific preservation services for e-journals, e-books and digitized historical collections respectively. Portico is a part of Ithaka (which is also responsible for JSTOR). It provides insurance to libraries that the e-journal and other content they have subscribed to will be preserved for the long term.

Portico preserves publisher source and presentation files and delivers content in renditions appropriate to the current technology.

Portico only provides access to the e-journals they have preserved after specified 'trigger events'. When e-journal titles have 'triggered', they are available to all participants in the Portico E-Journal Preservation Service, regardless of whether the participating institution has previously licensed the

<sup>20</sup> Marcel Ras personal communication, 6 June 2013.

<sup>21</sup> <http://www.lockss.org/>

<sup>22</sup> <http://www.portico.org/digital-preservation/>

content. In addition, if a publisher has designated Portico as such, it can also serve as a potential mechanism for post-cancellation access. Currently 88% of e-journals titles in Portico have been designated for post-cancellation access. This post-cancellation mechanism operates as follows:

- 1) a participating institution makes a claim to Portico, which includes documentation such as past invoices of their former subscriber status (see Section 2.4 for further discussion of the importance of retaining documentation);
- 2) Portico notifies the publisher of the claim and allows 30 days for the publisher to request any additional information it may need and confirm the claim;
- 3) upon publisher response, or at the end of 30 days, campus-wide access is provided by Portico to the requesting participating library.

Portico is funded by archive support fees from libraries and publishers. The annual fee to the e-journal preservation service for a library is set on a sliding scale of a library's materials budget (approximately 1%) with a 5% discount for consortia. UK libraries have access to a consortial discount with Portico brokered by Jisc Collections.<sup>23</sup> Annual financial contributions by participating publishers for the support of the Portico archive for e-journals are tiered on a scale from \$250 to \$81,960 based on a publisher's total journals revenues (print and electronic subscriptions, licensing, and advertising).

A TRAC preservation audit of Portico was completed in 2009 and the audit report is publicly available (CRL, 2010).

#### 4.7. Consortial hosting

A small number of regional consortia also organize and provide their own hosting services for access and preservation of e-journals. Notable examples are OhioLink,<sup>24</sup> operated by the Ohio Library and Information Network, and the Scholars Portal,<sup>25</sup> operated by the Ontario Council of University Libraries.

A TRAC preservation audit of the Scholars Portal was completed in 2012 and the audit report is publicly available (CRL, 2013).

#### 4.8. Keepers Registry

The Keepers Registry<sup>26</sup> is a pilot service to provide easily accessible information about inclusion of e-journals in preservation services and highlight those e-journals for which no archiving arrangements exist.

Establishing such a service was originally recommended in 2006 (Kenney *et al.*, 2006, 73–4) but it has taken many years to initiate. Jisc commissioned a report (Loughborough University and Rightscom, 2008) that reviewed the issues involved and recommended a registry should be established. The Keepers Registry service has been developed from these recommendations,

<sup>23</sup> <http://www.portico.org/digital-preservation/jisc-collections-agreement-for-portico>

<sup>24</sup> <http://www.ohiolink.edu/>

<sup>25</sup> <http://www.ocul.on.ca/node/135>

<sup>26</sup> <http://thekeepers.org>

initially as an output of the Jisc-funded Piloting an E-journals Preservation Registry Service (PEPRS) project.<sup>27</sup>

EDINA, a national data centre based at the University of Edinburgh, has developed the service along with its partner in the project, the ISSN International Centre in Paris. The ISSN IC has made available the ISSN Register, which contains bibliographic metadata for all journals that have been assigned an ISSN. E-journal archiving agencies are making metadata on the journals in their programmes available to the Keepers Registry. The data supplied by the agencies is linked to the authoritative bibliographic information obtained from the ISSN Register (Burnhill and Guy, 2010; Burnhill 2013).

---

<sup>27</sup> <http://edina.ac.uk/projects/peprs/index.html>

## 5. Use Cases

In order to illustrate the potential challenges and issues discussed in this report, six e-journal use cases have been identified and described in more detail below.

### 5.1 Use case 1: library cancels subscription to journal

Few libraries have remained unaffected by the serials pricing crisis of the past several years, and journal subscription cancellations have been one outcome of this. Unfortunately the current economic environment is likely to lead to an increase in cancellations as library budgets remain under pressure.

A library may have invested significant funds over a number of years to purchase the journal before finding that cancellation of the current journal issue is necessary. Before taking this step, the library will need to have developed its own risk assessment and a contingency plan for continuing access to past subscriptions. This will include having negotiated with publishers at the outset, where possible, post-cancellation access clauses in licences, any additional fees and the mechanism for access (e.g. via publisher's servers, trusted third-party services, or copies for loading on local systems). If copies are provided for local hosting, the nature of those copies needs to be agreed. However, it may in practice be very difficult for the institution alone to put in place or afford the services needed for the access and preservation of that content.

If post-cancellation access is provided via the publisher's server there could be long-term risks to access from commercial fluctuations and changes in the publisher's business. This has led to the emergence of trusted third-party (Portico) or collaborative (LOCKSS) archiving solutions for e-journals that can provide a potential safety net for continuing access. Coverage of different journals and journal issues and participation of publishers in these solutions vary considerably and will need to be assessed. In all cases, another major factor in selecting a solution will be the types and levels of service the library's users require to access the content, not solely its preservation. Electronic access and preservation services may involve a significant level of investment and cost and lead even large libraries to seek collaborative solutions to post-cancellation access.

### 5.2 Use case 2: library cancels subscription to a 'big deal'

The 'big deal' was first introduced by Academic Press as part of the UK National Pilot Site Licence Initiative in 1996, and has since become widespread in academic journal publishing. It is an offer in which a publisher sells online subscriptions to a bundled list of electronic journals (often the publisher's entire journals list) to individual members of a library consortium for a set number of years. The licence fee is normally based on the cost of the member library's historical print subscriptions with an electronic premium added. Typically, the library selects a core set of titles and its continuing access rights are confined to titles in that core set. If the library cancels its subscription (or amends the content of the core set), 'Use case 1' applies only to the core set of journal titles in the subscribed years: there are no continuing access rights to other titles in the bundle purchased via the electronic premium.

### 5.3 Use case 3: e-journal or its past issues are no longer available from the publisher

This is a highly likely use case as publishers merge or change their business models, as larger publishers review and adjust their portfolio of titles, or as learned societies move publication contracts for their journals from one publisher to another. Journal titles are also sometimes traded between publishers, which may mean that access to past issues is no longer supported by the previous owner.

The UKSG Transfer Code of Practice initiative has produced a Code of Practice aimed at easing the problems created when journal titles move between publishers. Of relevance are the following paragraphs contained in version 2 of the code (UKSG, 2008):

‘It is very common for journal content to be included in one or more archiving services. The Receiving Publisher will not remove content that was previously deposited in an archive, or archives, even if the Receiving Publisher will not be continuing to deposit content in the archive, or archives. The Receiving Publisher is encouraged to continue the existing, or equivalent, archiving arrangements for a journal after the Effective Transfer Date.

Customers that have been granted perpetual access rights to previously published content with the authority of the journal owner must have those rights honoured. Either the Transferring or the Receiving Publisher, or both, could fulfil perpetual access obligations.’

The decision of the publisher Sage to no longer offer its publication *Graft* provided a real-life example of triggered access from three archiving solutions – Portico, KB e-Depot, and CLOCKSS. In this case all were able to continue to offer access to the issues they held, either as open access (CLOCKSS and KB e-Depot) or else as a service to members (Portico). While it cannot be guaranteed that the archive will include all back issues of the title (as with *Graft*), participation in an archiving solution which covers at least some issues will significantly reduce the risk of disruption to continuity of service.

### 5.4 Use case 4: publisher ceases operation

In this use case, the publisher is no longer in business and therefore unable to support a service providing access to their collection of previously published journal issues. Suitable strategies for coping with this event depend on risk management techniques. The probability of this trigger event for large publishers is arguably low. Between 2006–2013 Portico lists seven trigger events as a result of publishers ceasing to operate (Use case 4) or a journal no longer being available (Use case 3), roughly an average of one per year over its operations so far.<sup>28</sup> The number of trigger events for CLOCKSS is similar with eight having been listed up to April 2013.<sup>29</sup>

<sup>28</sup> <http://www.portico.org/digital-preservation/the-archive-content-access/access-to-archived-content>

<sup>29</sup> [http://www.clockss.org/clockss/Triggered\\_Content](http://www.clockss.org/clockss/Triggered_Content)

## 5.5 Use case 5: catastrophic failure of publisher's operations/servers

In this use case, access to a publisher's e-journals suffers a major disruption. Examples might include fire, flood, explosion, lightning strike or terrorist action, which result in the total destruction of a publisher's (or its service provider's) servers and, perhaps, damaged or destroyed backup materials. It should be added that major publishers and service providers are likely to take steps to protect themselves against such events by running mirror sites at various locations around the world, but smaller publishers might not have the resources to do this. Off-site or secure fire-safe storage of backup media is also common practice, but perhaps not universal.

Rebuilding a service following such an event might take many months. An agreement between a publisher and an archiving solution that this is an acceptable trigger to open up access to the archive would enable subscribers to continue to access back copies. For example, Portico specifies the following condition for opening access:

'Licensor has stopped publishing or providing access to the Publication for a period longer than ninety (90) days due to technical difficulties or any business interruption, bankruptcy, insolvency, receivership or business failure.'

Note, however, that if it is assumed that the publisher will be able to resume operations at a future date, such a service would only open up access to the material on the preservation service's servers until such time as the publisher was able to resume its own service. The normal legal protections against downloading bulk data for access on other servers would still have to apply.

## 5.6 Use case 6: de-accessioning print

This use case differs from the others in illustrating a few of the issues in realizing some of the potential cost savings from e-journals, particularly space savings. Increasingly, academic libraries are investing heavily in e-journals which duplicate their print back-runs. For libraries facing acute pressures on space, one solution to their problem is to dispose of or relegate print back-runs which overlap with their electronic holdings.

In the USA, Ithaka has published a *What to Withdraw Tool* and accompanying *What to Withdraw Decision Support Tool Instructional Guide* to apply the logic of the *What to Withdraw: Print Collections Management in the Wake of Digitization* report (Schonfeld and Housewright, 2009). This can be used by librarians to determine which JSTOR-digitized journals meet their 'What to Withdraw' criteria.

A published UK case study focuses on work at Imperial College London Library in providing a database and toolkit for staff making such de-selection decisions (Cooper and Norris, 2007). Imperial established three criteria to determine the sustainability of their e-journals for de-accessioning of print. Their electronic access was classified as sustainable when at least one of the following applied:

- Imperial had perpetual access rights to the content, via the web. Imperial's perpetual access rights were nowhere near as ubiquitous as they would have wished; they estimated that less than 50% of their content was covered. In addition, some of their licences specified an unsuitable delivery method for post-termination access. As they were no

longer supporting networked CD-ROMs and did not have the resources to mount journal content locally, they considered a journal sustainable only if perpetual access is provided via the web.

- The journal was permanently open access for all years or certain years. Hybrid open-access journals were not included in this category, as the project was not interested in sustainability at the article level. Finding open-access journals which fulfilled their criteria proved harder than anticipated. The main stumbling block was their need for assurance on the permanency of open access. Although the Bethesda<sup>30</sup> and Berlin Declarations on Open Access<sup>31</sup> include perpetual access in their definitions, Imperial discovered that not all 'open-access journals' met this criterion of permanency.
- The content was in one of Imperial's trusted services such as JSTOR, the ACM digital archive or a Jisc-funded archive. Imperial noted that of their three sustainability criteria, this one, covering services that did not offer perpetual access rights, was the hardest to pin down. The services falling into this category all shared two characteristics: the first was a good track record of stability, i.e., they had demonstrated continuity of titles from one year to another for as long as they had subscribed; the second was a history of, and reputation for, affordability and value for money.

Twenty-one months into the project Imperial had identified 700 shelf-metres of sustainable stock for disposal from one site, and planned to rollout the de-selection exercise to other sites. Although it was still early days, they felt their sustainability criteria seemed to be working. The only sustainable content that they had lost was four journals from the same publisher, and they were in the process of challenging that loss. This proved to be an added benefit of the entitlements database they had created for the project; without it they would not have been aware that content over which they had perpetual access rights had been lost.

<sup>30</sup> <http://legacy.earlham.edu/~peters/fos/bethesda.htm>

<sup>31</sup> <http://oa.mpg.de/lang/en-uk/berlin-prozess/berliner-erklarung/>



## 6. Trust

### 6.1. Background

The requirements for trust in digital preservation were discussed at the DPC forum preceding this report.<sup>32</sup> Participants felt that trust required the archives to:

- Have a clear preservation definition and mission;
- Communicate what DP means in their services. The governance and operation need to back this up – preservation should be a priority and not secondary to access;
- Be transparent about what has been preserved and how. The preservation status of a journal must always be visible. Machine tools must facilitate this and it must be done in consultation with stakeholders;
- Extend this openness to auditing by third parties and have this presented to stakeholders;
- Have reliable delivery of service over time and effective responses to events requiring a preservation solution;
- Say what its aims are and follow up. This means responding to content in jeopardy and being able to demonstrate speedy effectiveness in a crisis;
- Show commitment to, and capacity for, research into future needs;
- Demonstrate effective and sympathetic solutions to new sorts of content and new ways in which research is done. This needs to take place through collaboration with research institutions.

Trust can be seen as a matter of both belief and reality (with hopefully only a narrow gap between the two).

### 6.2. Trust in digital repositories

Discussion at the forum focused on trust in archives, and establishing assessment criteria and objective measures for trust in digital repositories has been the focus of a number of initiatives in recent years. These initiatives can help libraries, scholars, publishers, and others to judge the reliability of the repositories and digital preservation services they may choose to rely upon or support. Two have been particularly prominent: the Trusted Repository Audit Checklist (TRAC) and its successor, The Trusted Digital Repository Checklist (TDR, or ISO 16363).

TRAC contains metrics that help in judging repositories in the areas of Administration and Policy, Object Handling and Technology. It was developed by a taskforce under the auspices of OCLC's Research Libraries Group (RLG) and the US National Archives and Records Administration (NARA). It also draws upon the experience and findings of test audits of repositories by the Center for Research Libraries (CRL) in 2005–2006. These test audits included a number of e-journal archiving services, specifically LOCKSS, the KB, and Portico. The LOCKSS test audit of 2006–7 (CRL, 2007) and

<sup>32</sup> <http://www.dpconline.org/events/previous-events/837-trust-and-e-journals>

a more recent 2010 second audit of Portico (CRL, 2010), and for the Scholars Portal (CRL, 2013) are publicly available. The final TRAC checklist was published in 2007 (RLG and NARA Taskforce, 2007).

The TDR is ISO Standard 16363 and was published in February 2012 (ISO, 2012b). It is a revision of the TRAC Checklist. Many of the changes were structural, and it continues to address the same core areas. A supplementary standard (ISO 16919), on requirements for bodies providing audit and certification of candidate trustworthy digital repositories, led by a Consultative Committee for Space Data Systems (CCSDS) working group is in preparation (ISO, 2011).

Other formal initiatives in this area of archive certification have been the Data Seal of Approval (DSA),<sup>33</sup> and the German Standard on Trustworthy Archive Certification DIN 31644 (DIN, 2012). There are also several toolkits for internal self-evaluation and assessment by repositories, including the Digital Repository Audit Method Based on Risk Assessment (DRAMBORA),<sup>34</sup> the Planning Tool for Trusted Electronic Repositories (PLATTER)<sup>35</sup> and the Long Term Digital Preservation (LTDP) Automated Assessment tool developed by IBM.<sup>36</sup>

In 2010, the European Framework for Audit and Certification of Digital Repositories was established as a collaboration between the Data Seal of Approval, the Repository Audit and Certification Working Group of the CCSDS, and the German Standards (DIN) Working Group on Trustworthy Archives Certification. It aims to support an integrated framework for auditing and certifying digital repositories consisting of a sequence of three levels, in increasing trustworthiness:

- Basic Certification is granted to repositories which obtain DSA certification;
- Extended Certification is granted to Basic Certification repositories which in addition perform a structured, externally reviewed and publicly available self-audit based on ISO 16363 or DIN 31644;
- Formal Certification is granted to repositories which in addition to Basic Certification obtain full external audit and certification based on ISO 16363 or equivalent DIN 31644.

Alongside these efforts there have been a number of projects that have focused on automating aspects of the repository audit process and/or a 'trust engineering' approach. Notable amongst these is the SafeArchive System designed to create a virtual overlay network on top of the LOCKSS peer-to-peer replication network, to support provisioning, monitoring, and TRAC-based auditing. It automates compliance with formal replication and storage policies by identifying specific TRAC criteria that can be verified automatically, and additional criteria that can be supported through integrated documentation (Altman and Crabtree, 2011).

### 6.3. Lessons learnt

Lessons learnt to date from the process of repository certification have been usefully summarized by the APARSEN project (APARSEN, 2012). Early efforts at developing repository certification were often dependent on significant grant funding for subsidy and/or involved significant costs to the repositories. Experience from the various pilot projects and current efforts towards tiered levels of certification and streamlining and standardizing approaches are reducing this. However, although

<sup>33</sup> Data Seal of Approval <http://www.datasealofapproval.org/>

<sup>34</sup> DRAMBORA <http://www.repositoryaudit.eu/>

<sup>35</sup> PLATTER Repository Planning Checklist and Guidance <http://www.digitalpreservationeurope.eu/platter.pdf>

<sup>36</sup> LTDP Assessment Tool <https://www.research.ibm.com/haifa/projects/storage/datastores/ltdp.html>

there has been considerable progress, arguably audit procedures are not yet fully bedded down and some issues remain for both auditors and repositories.

Repository audit tools and procedures have been helpful in raising standards of digital preservation and clarifying best practices. Publicly available audit reports have provided information and reassurance to users and funders on digital preservation practice in specific repositories at specific times.

Audit reports and tools can be produced or applied in a number of different internal or public contexts. Seadle has suggested a useful set of definitions that may be useful in categorizing them:

**Public test:** a public test is one that a) people not involved in the original can replicate; and b) whose results have been published openly in public journals or websites.

**Private test:** a private test is one that a) cannot be replicated without the assistance of the original partners; and b) whose results are available only to a closed audience.

**Semi-public test:** a semi-public test is one that a) can only be replicated with the assistance of the original partners; and b) whose results have been published openly in public journals or websites (Seadle, 2011b).

As noted in Section 6.1, discussion of trust in e-journal archiving has focused largely on trust in the archive repositories. However, as set out in previous sections, the archive repositories are only one element in the provision of long-term access to e-journals. Many of the same requirements for transparency and defined procedures for enacting agreements also apply between authors, libraries and publishers. Other initiatives, such as the Transfer Code of Practice (see Sections 2.6 and 3.3.1), the Keepers Registry (see Section 4.8), and the development of model clauses and enactment mechanisms in licences (see Section 2.4) are also key components in trust in e-journal archiving. For trust to be established there need to be clear agreements for long-term archiving, and clear procedures and mechanisms for those agreements to be implemented and validated when necessary across all elements of the supply chain.

## 7. Horizon Scanning

Throughout, this report has sought to place developments in e-journal archiving in their current context and to highlight emerging and on-going issues that may continue to have an impact for many years to come. Two issues are worthy of particular iteration in thinking about likely trends impacting on e-journal archiving in the future.

The atomic and static PDF files of the early e-journals days are steadily being replaced by bundles of dynamic and interdependent resources that are distributed across the web. Now complex entities mentioned in e-journals may be automatically expanded to show diagrams or pictures that facilitate understanding; and ancillary material that augments the text may be linked to or distributed with the article. This may include research data (from experiments), images and videos, sound recordings, and presentational materials. Extracts and discussions of that scholarly work on social media such as blogs, online discussion groups and Twitter may greatly broaden its visibility and enable it to be better evaluated and cross-linked to other information sources. In the longer term this may see greater emphasis being placed on ‘research objects’, bundles of linked content (Force 11, 2011).

These changes present technical challenges regarding information interoperability and long-term preservation; and new challenges related to stewardship, the delineation of the scholarly record, and the very notion of the version of record itself.<sup>37</sup>

This trend to embrace fully the web as infrastructure for scholarly communication will require us to adapt future preservation efforts accordingly. It underlines why attendees at the DPC e-journal archiving workshop emphasized the importance placed on on-going R&D in securing trust in e-journal archiving (see section 6.1).

The second issue to note is the growing trend towards open access in e-journals and its implications. Challenges that arise include ensuring appropriate archiving permissions for open-access articles and funding of the preservation of open-access journals. A recent study looked at the archiving of journals listed in the Directory of Open Access Journals (DOAJ). It found that only 8% of the DOAJ titles are in LOCKSS/CLOCKSS and only 5% in Portico. An agreement with KB, the Dutch national library, will mean that the 50% of DOAJ titles that have full text will be archived by the KB e-Depot; however, it is estimated that it could take eight years to complete this process. It notes the implication that most open-access titles listed in DOAJ currently have no effective long-term archiving (Seadle, 2011c).

In 2009, the composition of the DOAJ collection (then 4,000 journals) was characterized by a very large number of publishers (2,000+), each publishing a very small number of journals on different platforms, in different formats and in more than 50 different languages. Many of these publishers were and are – with a number of exceptions – fragile when it comes to financial, technical and administrative sustainability.<sup>38</sup> Doing more to address the preservation of these open-access journals, and at the same time maintaining the business case and business models underpinning preservation repositories, is likely to be another major challenge over coming years.

<sup>37</sup> For an excellent discussion of this see the plenary by Herbert Van De Sompel at the Spring 2013 CNI meeting <http://www.cni.org/news/video-van-de-sompels-plenary-from-the-version-of-record-to-a-version-of-the-record/>

<sup>38</sup> DOAJ press release 1 April 2009 <http://www.doaj.org/doaj?func=news&nId=22&uiLanguage=en>. It would be valuable to update this analysis of OA titles and publishers and assess current archiving of them.

Preservation is likely to be further complicated by the increasing primacy of the article level for access and rights, e.g. in hybrid journals mixing open-access and subscriber-only articles.

## 8. Conclusions

Continuing access and preservation of e-journals has involved initiatives in organizing multi-institutional collaboration, developing third-party services, and establishing trust in long-term access and preservation between different stakeholders. The issues it has had to address go well beyond technology, and legal, economic and service developments are equally critical to its success.

e-journal archiving will need to evolve further in future to address on-going changes in modes of scholarly communication and business models for its publishing.

For libraries it is possible to suggest draft criteria for the selection of archiving solutions that can address their e-journal archiving needs:

### CONTENT

- Number and type of publishers, titles, and year ranges included in the archive.
- Degree of fit of the archive's coverage to required subscribed journal content.
- Content is quality assured at ingest to ensure there are no missing issues, or corrupt content.

### SERVICE PROVISION AND QUALITY

- Degree of support for required continuing access provisions in subscribed journal content.
- Acceptable terms and conditions for continuing access in the event of subscription cancellation or another trigger event. Minimal time delays and clear certainty of outcomes.
- An explicit commitment by the archive to long-term digital archiving for scholarly peer-reviewed journals.
- Open to formal certification of its digital archive and meets broadly accepted standards of best practice.
- There should be appropriate levels of support. Any required level of IT or curatorial support expected from the library or its institution must be acceptable.

### GOVERNANCE AND BUSINESS MODEL

- Stakeholders are involved in the governance of the archive.
- Well-founded robust organization, funded for long-term sustainability, with sufficient staff resources and range of expertise.
- Affordable and value for money.

Many challenges remain in e-journal archiving, but there have been significant successes and lessons learnt of interest to the wider digital preservation community as well as to libraries and publishers.

## 9. Recommended Actions

### 9.1. For academic libraries, library associations and library consortia

- If archiving e-journal content locally or collaboratively develop:
  - (a) An appropriate digital preservation strategy and implementation procedures for the archive;
  - (b) Integrated access to archived digital content from library IT systems and tools such as link resolver knowledge bases;
- If outsourcing, assess how e-journal archiving solutions fit against your needs for coverage, and service terms and conditions;
- Evaluate potential benefits and savings from e-only and off-setting the cost of e-journal archiving against identified savings in cash budgets in areas such as binding;
- Implement clauses for sustainable content in your library collection and development strategies and policies;
- Assign staff effort to this activity and ensure the local knowledge and workflows are in place to carry out your strategies and policies;
- Engage on e-journal archiving with your peers, and provide input on selection and service issues to service providers and your membership organizations;
- Work to develop licence provisions for continuing access and relevant e-journal archiving solutions;
- Collect data on your continuing/perpetual access rights in e-journal licences.

### 9.2. For publishers

- Cooperate with one or more external e-journal archiving solutions;
- Develop and publicize your policies on perpetual access and long-term preservation;
- Post-cancellation access conditions should be defined in licensing agreements between libraries and publishers;
- When titles are sold on to other publishers, the Transfer Code of Practice should be followed;
- If you are an open-access publisher, use the CC-BY licence, which is the most user-friendly and preservation-friendly licence, allowing among other things for long-term preservation and text- and data mining. Simple and consistent licensing substantially reduces the cost of preservation of open-access content.

### 9.3. For e-journal archives

- Commit to providing the latest information on archive coverage and new content (e.g. via the Keepers Registry);
- Provide a means for the community to nominate new content to the archive;
- Consult your stakeholders on a Collection Development Policy which sets out criteria against which new acquisitions are evaluated and prioritized;
- Maintain formal relationships with publishers that include the right to ingest and manage a significant number of journal titles over time;
- A digital archive of journals of international importance to scholarship in perpetuity should consist of at least a minimum of two international nodes (mirrored services). These nodes should be in different geographical continents and jurisdictions to minimize long-term threats to any archive;
- There should be publicly accessible policies and procedural documents. These should include clear statements and definitions of access conditions, including post-cancellation access and trigger events;
- Seek to include more small and medium-sized publishers and open-access titles in your programmes;
- Maintain an active involvement in R&D efforts to address future changes in e-journal publishing and scholarly communication and preservation challenges that arise.



## 10. Glossary

### ACCESS

There are varying degrees of access to journal articles in the digital world. Subscriber access refers to the ability of a subscribing organization or a consortium of organizations and their patrons to make use of content which has been licensed from a commercial publisher for the specific use of the registered members of that organization during the licensing period. Guest access refers to the ability of users other than those belonging to a subscribing organization to have restricted access to the content. Open Access (OA) is e-journal article content that is digital, openly available online, free of charge to the user and free of most copyright and licensing restrictions. OA is now practised in a variety of forms including Gold OA, and Green OA. Mixed models of subscriber and open access include hybrid journals containing both open and subscriber-only articles, and Delayed OA where access is only available first to subscribers for a fixed period before being available free of charge.

### APPROPRIATE COPY

One or more versions of a journal article, among many, which are most appropriate for a specific user in a specific situation at a given institution. This is likely to be a version of which they are entitled to access the full text, probably because of a subscription paid for by the library.

### ARCHIVE

An archive is a repository created to preserve material no longer in heavy current usage but nevertheless still needing to be accessed on occasions or for specific reasons. Three distinct terms to describe grades of digital archive for e-journals are in common use. A DARK ARCHIVE is an archive that cannot be accessed by any current users but may be accessible at future dates subject to the occurrence of specific pre-defined events ('trigger event'). Access to the data is either limited to a few set individuals or completely restricted to all. Typically these dark archives can be divided into two main types: type 1 – those only providing a form of escrow or 'bit preservation' of content that is suitable as a short-medium term solution for guaranteeing access; and type 2 – those providing the bit preservation of the content plus some degree of associated services for future access (this may include a 'back-up' access service should a primary access service fail for any reason, and digital preservation planning and preservation action services such as file format migration which will keep the content accessible in the future). These provide for and can help guarantee long-term perpetual access. A DIM ARCHIVE provides bit preservation for the content plus digital preservation planning and actions for long-term perpetual access, and also limited current access (perhaps limited to on-site users or previous subscribers post-cancellation, etc.). A crucial concept behind many archives is agreeing to restrict access over a period of time or until specific events occur in the future. At such points 'dark archives' and 'dim archives' may transform to a LIGHT ARCHIVE that can be accessed by many authorized users. Access to the information is open to all members of the 'community' that has a need for the information. Access may be subject to access restrictions agreed upon by the publisher of the material and the archive. Current access under some circumstances would always be presumed and an access system would be maintained.

## BACKFILE

That portion of an electronic journal title which is not included in the subscription to the current year of that journal and a number of designated years prior to the current year is known as the backfile. A specific cut-off date is used by many publishers to designate the difference between backfile and current content. Frequently the backfile can be purchased as a one-off but still under licence and can contain all the digitized volumes from the start of the journal to the cut-off date. Cumulative backfiles for a group of electronic journal titles by the same publisher may be offered with a common cut-off date. Supplementary backfiles may be offered subsequently. Both may be subject to an on-going access fee if accessed from a publisher's server.

## CONSORTIUM

A group of individuals, libraries or organizations with common interests formed to undertake an enterprise or activity that would be beyond the capabilities of the individuals, libraries or organizations on their own. In the context of electronic materials it normally refers to a group (regional, national or international) that undertakes the negotiations for the purchasing and licensing of such materials on behalf of the constituent members of that group.

## CONTINUING ACCESS

Continuing access refers to the right of the subscriber and their users to have on-going permanent access to electronic materials which have already been leased and paid for by the subscriber from a publisher. It is a term used, along with its synonyms perpetual access and post-cancellation access, in the information industry to describe the ability to retain access to electronic materials by the subscriber/licensee after the contractual licensing agreement with the publisher/licensor for those materials has ended, whatever the reason for the cessation. It may also cover as appropriate arrangements for digital preservation needed to guarantee some elements of continuing access. The subscriber/licensee and the publisher/licensor, both of whom are party to the licence, need to agree terms for the granting of continuing access rights to the subscriber. It may refer to both the leasing of current content and the purchase of backfiles. For some publishers the payment of an on-going annual access fee to cover remote hosting costs is required. An alternative to on-going access via the publisher's servers post cancellation may be supplying a copy of the content to the licensee for local hosting or a third-party archive that also provides this service.

Dark Archive – see ARCHIVE

## DIGITAL PRESERVATION

Digital preservation is a term used to cover the series of managed activities necessary to ensure reliable access to digital materials for as long as necessary and beyond the limits of storage media failure or technological change.

Dim Archive – see ARCHIVE

## ESCROW

Escrow is a widespread legal practice of the deposit of content or software source code with a third party. Escrow is typically requested by a licensee, to ensure access to a version of the licensed material under certain agreed conditions, e.g. if the licensor files for bankruptcy or otherwise fails to meet its obligations. Escrow takes place in a contractual relationship, formalized in an escrow agreement, between at least three parties: the licensor(s), licensee(s), and the third party providing the escrow service.

## HOSTING

Hosting refers to the service used for primary access to content. Local hosting of electronic materials is the holding of data files provided by a publisher under an agreement with the subscriber on a local server under the control of the subscriber themselves or a designated organization, other than the publisher, working in partnership with the subscriber. It is the alternative to access/hosting via the publisher's own servers. Local hosting provides archival protection and rights to the resources under local control. Control of the local server is under the subscriber who must provide the labour to maintain it and also maintain access controls to the content. The service may encompass elements of digital preservation and an archive but this is usually not its primary function and it may be less comprehensive in these areas than dedicated services. In some cases institutions may co-operate to federate local hosting to regional or potentially national scale.

## HOSTING SERVICE

Hosting service(s) provide the point(s) of full current access to authorized users and on-going maintenance, updating, and security of that content (including routine back-ups and disaster recovery planning). These may provide for and help guarantee short- to medium-term horizons for access but are not focused on digital preservation or organized for the long term, and would need to partner with or be combined with other services to achieve or guarantee longer-term perpetual access. A hosting service may be run in-house or via an outsourced service provider.

## LICENCE

A licence is a legal document giving official permission to undertake an activity and is granted by one party (the licensor) to another party (the licensee) as an element of an agreement between these parties. In the digital world it is the authorization for the licensee to provide, and their users to use, licensed electronic materials supplied by the licensor to the subscribing organization. It is governed by contract law and may include specific elements determined by the laws of the issuing country. It is normally time limited to the subscription period, and needs to be renewed at given periods. A site licence is a particular type of licensing agreement that permits access to and use of digital information at a specific geographical site or location agreed by both parties and under specific conditions and groups of users. Licences may contain clauses covering a designated archive, hosting, perpetual access, digital preservation and transfer.

Light Archive – see ARCHIVE

Local Hosting – see HOSTING

Open Access – see ACCESS

Perpetual Access – see CONTINUING ACCESS

Post-Cancellation Access – see CONTINUING ACCESS

## TRANSFER

In the electronic publishing world, 'transfer' refers to the change of publisher or owner which can occur for an electronic journal title and the consequent roles and responsibilities which must be adopted by the 'transferring publisher' and the 'receiving publisher' respectively, to ensure the continued access to content, both currently and in perpetuity, for the subscriber to that journal and their users. A Transfer Code of Practice has been created, which seeks to set acceptable operational standards and publishers are encouraged to sign up to this code and become Transfer Compliant.

## TRIGGER EVENT

This terminology is used when specific conditions relating to an electronic journal title and its continued delivery to users are met. If the journal is no longer available to users from the publisher or any other source for a variety of reasons then a trigger event is said to have occurred. They can set in motion access for users via an archive where the electronic journal may be digitally preserved. Generic examples of trigger events include four from the publisher's side and one from the subscriber's side – a publisher ceases operations, a publisher ceases to publish a title, a publisher no longer offers back issues or there is a catastrophic and sustained failure of a publisher's delivery platform, or the subscriber/licensee for economic or scholarly reasons cancels their subscription to the journal/s. An archive will have its specific definitions of trigger events defined in its licences and contracts.

## 11. References and Further Reading

All web links cited in references were last accessed on 14 May 2013.

2CUL 2011, *Final Report of the 2CUL LOCKSS Assessment Team*, Cornell University Library & Columbia University Library. Available:  
<http://2cul.org/sites/default/files/2CULLOCKSSFinalReport.pdf>

Altman, M and Crabtree, J 2011, Using the SafeArchiveSystem: TRAC-Based Auditing of LOCKSS. *Proceedings of Archiving 2011*, Society for Imaging Science and Technology. May 2011, 7, 165–170. Available:  
[http://thedata.org/files/thedata\\_new2/files/archiving2011\\_altman\\_crabtree.pdf](http://thedata.org/files/thedata_new2/files/archiving2011_altman_crabtree.pdf)

APARSEN 2012, *Report on Peer Review of Digital Repositories*, APARSEN-REP-D33\_1B-01-1\_0. Available: [http://www.alliancepermanentaccess.org/wp-content/uploads/downloads/2012/04/APARSEN-REP-D33\\_1B-01-1\\_0.pdf](http://www.alliancepermanentaccess.org/wp-content/uploads/downloads/2012/04/APARSEN-REP-D33_1B-01-1_0.pdf)

Beagrie, N 2010, *E-journal archiving for UK HE libraries: a draft white paper* (JISC Consultation Draft 1 Oct 2010). Available:  
<http://www.jisc.ac.uk/whatwedo/programmes/preservation/2010ejournalwhitepaper.htm>

Beck, J 2011, NISO Z39.96 The Journal Article Tag Suite (JATS): What Happened to the NLM DTDs? , *The Journal of Electronic Publishing*, 14 (1). Available:  
<http://dx.doi.org/10.3998/3336451.0014.106>

BRTF, Blue Ribbon Task Force on Sustainable Digital Preservation and Access 2010, *Sustainable Economics for a Digital Planet: Ensuring Long-Term Access to Digital Information* . Available: [http://brtf.sdsc.edu/biblio/BRTF\\_Final\\_Report.pdf](http://brtf.sdsc.edu/biblio/BRTF_Final_Report.pdf)

de Boer, T and van Otegem, M 2012, *Moving to new digital storage: migrating and reloading collections*, August 2012 78th IFLA General Conference and Assembly  
<http://conference.ifla.org/past/ifla78/102-boer-en.pdf>

Burnhill, P 2013, *Tales from The Keepers Registry: Serial Issues About Archiving & the Web*, *Serials Review*, 39 (1), 3–20.  
 Available: <http://dx.doi.org/10.1016/j.serrev.2013.02.003>

Burnhill, P and Guy, F 2010, Piloting an E-journals Preservation Registry Service (PEPRS), *The Serials Librarian*, 58 (1), 117–126. Available: <http://dx.doi.org/10.1080/03615261003622742>

CEPA (Cambridge Economic Policy Associates) 2008, *Activities, costs and funding flows in the scholarly communications system in the UK*, Research Information Network, May 2008. Available: <http://www.rin.ac.uk/system/files/attachments/Activites-costs-flows-report.pdf>

CRL, Center for Research Libraries 2007, *LOCKSS Audit Report*. Available:  
[http://www.crl.edu/sites/default/files/attachments/pages/LOCKSS\\_Audit\\_Report\\_11-07.pdf](http://www.crl.edu/sites/default/files/attachments/pages/LOCKSS_Audit_Report_11-07.pdf)

- CRL, Center for Research Libraries 2010, *Report on Portico Audit Findings*. Available:  
<http://www.crl.edu/sites/default/files/attachments/pages/CRL%20Report%20on%20Portico%20Audit%202010.pdf>
- CRL, Center for Research Libraries 2013, *Report on Scholars Portal Audit*. Available:  
[http://www.crl.edu/sites/default/files/attachments/pages/ScholarsPortal\\_Report\\_2013\\_%C6%92.pdf](http://www.crl.edu/sites/default/files/attachments/pages/ScholarsPortal_Report_2013_%C6%92.pdf)
- Charles Beagrie Ltd and Globale Informationstechnik GmbH 2010, *Ensuring Perpetual Access: establishing a federated strategy on perpetual access and hosting of electronic resources for Germany*, edited final report, February 2010. Available:  
[http://www.allianzinitiative.de/fileadmin/hosting\\_studie\\_e.pdf](http://www.allianzinitiative.de/fileadmin/hosting_studie_e.pdf)
- Cooper, R and Norris, D 2007, To bin or not to bin? Deselecting print back-runs available electronically at Imperial College London Library, *Serials* 20 (3), 208–214. Available:  
<http://serials.uksg.org/openurl.asp?genre=article&issn=0953-0460&volume=20&issue=3&spage=208>
- DCMS 2012, *Consultation on the Draft Legal Deposit Libraries (non-print works)*, Regulations 2013. Available: <http://www.culture.gov.uk/consultations/8878.aspx>
- DIN 2012, *DIN 31644 Information and documentation – Criteria for Trusted Digital Repositories*  
<http://www.nabd.din.de/cmd?level=tpl-art-detailansicht&committeeid=54738855&artid=147058907&languageid=de&bcrumblevel=3&subcommitteeid=112656173>
- Force 11 2011, *Force 11 Manifesto – Improving Future Research Communication and e-Scholarship*. Available: [http://force11.org/white\\_paper](http://force11.org/white_paper)
- HMSO 2013, *The Legal Deposit Libraries (Non-Print Works) Regulations 2013*, Statutory Instruments 2013 No. 777, 5th April 2013. Available:  
<http://www.legislation.gov.uk/ukxi/2013/777/contents/made>
- ISO 2011, *ISO 16919 Space data and information transfer systems - Requirements for bodies providing audit and certification of candidate trustworthy digital repositories*  
[http://www.iso.org/iso/iso\\_catalogue/catalogue\\_tc/catalogue\\_detail.htm?csnumber=57950](http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=57950)
- ISO 2012a, *ISO 19005-3:2012 Document management – Electronic document file format for long-term preservation – Part 3: Use of ISO 32000-1 with support for embedded files (PDF/A-3)*. Available:  
[http://www.iso.org/iso/home/store/catalogue\\_ics/catalogue\\_detail\\_ics.htm?ics1=37&ics2=100&ics3=99&csnumber=57229](http://www.iso.org/iso/home/store/catalogue_ics/catalogue_detail_ics.htm?ics1=37&ics2=100&ics3=99&csnumber=57229)
- ISO 2012b, *ISO 16363: Space data and information transfer systems – Audit and certification of trustworthy digital repositories*. Available:

[http://www.iso.org/iso/iso\\_catalogue/catalogue\\_tc/catalogue\\_detail.htm?csnumber=56510](http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=56510)

Jisc Collections 2010, *Ensuring that 'e' doesn't mean ephemeral: A practical guide to e-journal solutions*, version 1.1, February 2010. Available: [http://www.jisc-collections.ac.uk/Documents/practical\\_guide\\_to\\_ejournal\\_archiving.pdf](http://www.jisc-collections.ac.uk/Documents/practical_guide_to_ejournal_archiving.pdf)

Jisc Collections 2012, *Post-Cancellation Entitlement Registry Scoping Project*. Available: <http://www.jisc-collections.ac.uk/Reports/entitlementregistryreport/>

Jones, M 2003, *Archiving E-Journals Consultancy*. October 2003. Available: [http://www.jisc.ac.uk/uploaded\\_documents/ejournalsfinal.pdf](http://www.jisc.ac.uk/uploaded_documents/ejournalsfinal.pdf)

Jones, M 2007a, *E-Journal Archiving – Review and Analysis of the CLIR report E-Journal Archiving Metes and Bounds: A Survey of the Landscape*. Available: [http://www.jisc.ac.uk/media/documents/programmes/preservation/jisc\\_ej\\_reportfinal2\\_16feb07001.pdf](http://www.jisc.ac.uk/media/documents/programmes/preservation/jisc_ej_reportfinal2_16feb07001.pdf)

Jones M 2007b, *JISC Briefing Paper: e-journals Archiving and Preservation*. Available: [http://www.jisc.ac.uk/publications/publications/pub\\_ejournalspreservationbp.aspx](http://www.jisc.ac.uk/publications/publications/pub_ejournalspreservationbp.aspx)

KBART Working Group 2010, *KBART: Knowledge Bases and Related Tools A Recommended Practice of the National Information Standards Organization (NISO) and UKSG*. Available: [http://www.uksg.org/sites/uksg.org/files/KBART\\_Phase\\_I\\_Recommended\\_Practice.pdf](http://www.uksg.org/sites/uksg.org/files/KBART_Phase_I_Recommended_Practice.pdf)

Kenney, AR, Entlich, R, Hirtle, PB, McGovern, NY and Buckley, EL 2006, *E-Journal Archiving Metes and Bounds: A Survey of the Landscape*. Available: <http://www.clir.org/pubs/reports/pub138/contents.html>

Loughborough University and Rightscom 2008, *Scoping study for a registry of electronic journals that indicates where they are archived*. Available: <http://www.jisc.ac.uk/media/documents/programmes/preservation/ejournalregstudy.pdf>

Morrow, T, Beagrie, N, Chruszcz, J, and Jones, M 2008, *A Comparative Study of e-Journal Archiving Solutions*, Joint Information Systems Committee 2008 . Available: <http://www.jisc-collections.ac.uk/Reports/e-journal-archiving-comparative-study/>

Muir, A 2001, Legal deposit and preservation of digital publications: a review of research and development activity, *Journal of Documentation*, 57 (5), 652–682. Available: <http://dx.doi.org/10.1108/EUM0000000007097>

NISO 2012, *Journal Article Tag Suite*, version 1.0 (ANSI/NISO Z39.96-2012). Available: <http://www.niso.org/workrooms/journalmarkup>

Pennock, M 2012, *Web Archiving, DPC Technology Watch Report 13-01* March 2013. Available: <http://dx.doi.org/10.7207/twr13-01>



- Rightscom 2008, *E-Journal Mapping Study*. Final Report to British Library Legal Deposit Advisory Panel 29 August 2008.
- RLG-NARA Digital Repository Certification Task Force 2007, *Trustworthy repositories audit & certification: Criteria and checklist*. Available: <http://www.crl.edu/PDF/trac.pdf>
- Rosenthal, D S H, Lipkis, T, Robertson, T S and Morabito, S 2005, Transparent Format Migration of Preserved Web Content, in *D-Lib Magazine*, 11 (1). Available: <http://www.dlib.org/dlib/january05/rosenthal/01rosenthal.html>
- Rosenthal, DSH and Reich, V 2010. Archiving Supplemental Materials, *Information Standards Quarterly*, 22 ( 3), 16-21. Available: <http://dx.doi.org/10.3789/isqv22n3.2010.04>
- SIS, Scholarly Information Strategies) 2007, *Link Resolvers and the Serials Supply Chain*. Available: <http://www.uksg.org/projects/linkfinal>
- Schonfeld, RC and Housewright, R 2009, *What to Withdraw? Print Collections Management in the Wake of Digitization*, Ithaka S+R September 29, 2009. Available: <http://www.sr.ithaka.org/research-publications/what-withdraw-print-collections-management-wake-digitization>
- Seadle, M 2011a, Archiving in the networked world: by the numbers, *Library Hi Tech*, 29 (1), 189–197. Available: <http://dx.doi.org/10.1108/07378831111117001>
- Seadle, M, 2011b, Archiving in the networked world: metrics for testing, *Library Hi Tech*, 29 (3), 557–564. Available: <http://dx.doi.org/10.1108/07378831111174486>
- Seadle, M 2011c, Archiving in the networked world: open access journals, *Library Hi Tech*, 29 (2), 394–404. Available: <http://dx.doi.org/10.1108/07378831111138251>
- UKSG (UK Serials Group) 2008, *TRANSFER Code of Practice: Version 2.0*, September 2008. Available: <http://www.uksg.org/Transfer/Code>
- UNESCO 2000, *Guidelines for Legal Deposit Legislation*. Available: <http://unesdoc.unesco.org/images/0012/001214/121413eo.pdf>
- Van de Sompel, H, Payette, S, Erickson, J, Lagoze, C, and Warner, S 2004, Rethinking scholarly communication: Building the System that Scholars Deserve, *D-Lib Magazine*, 10 (9). Available: <http://dx.doi.org/10.1045/september2004-vandesompel>
- W3C 2012, *HTML 5.1 A vocabulary and associated APIs for HTML and XHTML*, W3C Working Draft 17 December 2012 Available: <http://www.w3.org/TR/html51/>
- Waters, D 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*. Available: <http://www.clir.org/pubs/reports/pub107/waters.html>



Waters, D 2005, *Urgent Action Needed to Preserve Scholarly Electronic Journals*. Available:  
<http://www.diglib.org/pubs/waters051015.htm>