[Link to publication](#)

**University of Bath**

# Online serials: preservation issues

Michael Day[1]

UKOLN: The UK Office for Library and Information Networking,
University of Bath, Bath BA2 7AY, United Kingdom
http://www.ukoln.ac.uk/
m.day@ukoln.ac.uk

**Abstract.** The paper consists of a preliminary investigation of preservation issues related to scholarly online electronic serials. Some background issues are discussed relating to how preservation should be defined, types of electronic serials, the current and future development of scholarly communication and formats currently in use. The discussion of preservation that follows looks at media longevity, hardware and software dependence and obsolescence, data migration, authentication, copyright and ownership issues, legal deposit and the related problems of who should preserve and what should be preserved in the digital age.

## 1. Introduction

Digital preservation has interested the library and information communities since the viability of electronic publication was first noted in the 1970s. Serious problems have consistently been raised with ensuring the continued existence of electronic information, and these concerns led to the publication of a report on *Preserving digital information* by a Task Force on Archiving of Digital Information commissioned by the Commission on Preservation and Access and the Research Libraries Group in May 1996 [1]. The recent growth in the production and use of online serials - electronic serials available over networks - demonstrates that the development of sensible preservation strategies for these items are essential to ensure the future of scholarly communication. This paper will look at the preservation implications of scholarly online serials with a discussion of some background issues and an outline of the most important current digital preservation issues.

---

[1]  Preprint version of an article of the same name appearing simultaneously in: *E-Serials: publishers, libraries, users, standards*, ed. Wayne Jones. Binghamton, N.Y.: Haworth Press, 1998, pp. 199-221; and *The Serials Librarian*, Vol. 33, No. 3/4, 1998, pp. 199-221, doi:10.1300/J123v33n03_01

## Background issues

### Defining preservation

The concept of preservation, when used with regard to electronic or digital information, can be difficult to define adequately. Some people prefer to use the word 'archiving', which is used in a computing science context to mean the creation of a secure backup copy for a fixed period of time, but the term preservation will be used in this paper as it avoids confusion with the work of the archives community. The objective of preservation itself has been defined by John Feather as: "to ensure that information survives in a usable form for as long as it is wanted" [2]. Preservation, therefore, is not just concerned with the conservation or restoration of artefacts, but includes all strategic considerations related to the survival of information over time. A distinction is often made between the preservation of the information embodied in a document (the informational content) and the conservation of the information carrier itself (the physical object) [3]. This is especially relevant for digital information, including networked online serials where the user does not even have to be aware of what particular physical object is being accessed. There may be a case for retaining physical objects - for example, in a museum of technology - but the preservation of informational content is much more important.

How long information should be kept is also another important, and often emotional, issue. It is often assumed that preservation should be permanent, often defined with reference to loaded terms like 'in perpetuity' and 'indefinitely'. In the context of archives, David Bearman has pointed out the absurdity of using concepts of permanence with regard to preservation and instead has proposed a more realistic concept of "retention for period of continuing value" [4]. On similar lines, James O'Toole has pointed out that an acute conservation consciousness has meant that archivists have been lulled into a false sense of security about their collections, and have thus "lost sight of the larger purposes of their work - preserving over time information that is of benefit and use to society - and have restricted the available options for approaching that goal" [5]. In consequence, this paper will assume that preservation is irrevocably linked to access and use, not to nebulous concepts of permanence.

Furthermore, a distinction has also to be made between digital preservation and digitisation for preservation, as the two are often confused. In the context of serials, back issues of important or rare serials are often digitised and made available over networks both to improve access to the information contained therein and sometimes to aid the preservation of the original item [6]. Examples of this type of operation are the Journal Storage Project (JSTOR) funded by the Andrew W. Mellon Foundation and the Internet Library of Early Journals (ILEJ) funded by the UK Electronic Libraries Programme (eLib) [7]. Although these digitised versions of serials will need preservation themselves, they will not form the primary focus of this paper.

## Types of online serial

While printed serials are a familiar part of the research library, there is more than one recognisable type of electronic serial. A preliminary survey of electronic information and serials collection management by Hazel Woodward in 1994 identified three major categories of electronic serial [8].

- **Online** - full-text electronic serials available through online hosts like DIALOG, usually available on a cost-per-access basis. Woodward has elsewhere noted that these types of online electronic journals are mostly "the electronic versions of existing printed journals" [9].
- **CD-ROM** - electronic serials distributed on CD-ROM, e.g. full-text newspapers or the ADONIS document delivery system [10]. Many of these are also electronic versions of printed serials. Serials published exclusively on CD-ROM will not, however, be discussed in this paper.
- **Networked electronic journals** - electronic serials available via networks, usually the Internet. These, as defined by Woodward, are normally only available in electronic form and include only those serials which have been peer-reviewed. They also tend to be non-commercial in nature. Serials of this type are listed in the *Directory of electronic journals, newsletters and academic discussion lists* published by the Association of Research Libraries (ARL) [11].

For the purpose of this paper, there is no reason to maintain a fixed distinction between online serials and networked electronic journals as both are available online. The major differences are the type of publisher involved - commercial and learned society publishers on one side and non-commercial 'network publishers' on the other - and whether the serial is an electronic version of a journal that already exists in printed form or is only available in electronic form.

## The development of the online serial

The electronic serial has been in development for about twenty years. Research projects in the 1980s first proved that electronic serials were technically feasible. The Electronic Information Exchange System (EIES) and the Birmingham and Loughborough Electronic Network Development (BLEND) demonstrated that systems could be developed for all stages of the production of an electronic serial, including article submission, peer-review processes, editing and distribution [12]. However, there was no major take-up of electronic serials as a result of these experiments largely because the necessary technological infrastructure - widespread access to computers and robust international networks - was not in place at that particular time [13].

At about the same time as these technological shortcomings were being remedied, concern was also being raised over two significant problems of scholarly communication: the relatively long time it takes for a paper to be published in scholarly serials - particularly significant in the fields of Science, Technology and Medicine (STM) - and the "serials crisis" - a rapid growth in serial numbers and prices at the same time as research institutions and their libraries had entered an extended period of financial stringency [14]. It was felt that electronic publication would help speed up the publication time for serial articles and - crucially - might

enable financial savings to be made to offset the effect of the serials crisis. Indeed, self publishing through electronic networks began to be seen in a new, more radical light - returning the responsibility of ownership and distribution of scholarship to its creators [15]. Widespread adoption of self publishing through electronic networks would have a significant impact on the future viability of serials, whether available online or traditionally, and needs to be considered here with reference to preservation.

Andrew Odlyzko, a mathematician, has argued in an important article that traditional printed scholarly serials are likely to disappear within the next ten to twenty years [16]. He suggests that there are two main factors for this: the rapid growth of the scholarly literature; and the increasing availability and capability of electronic technology. Odlyzko, thus, predicts that scholarly publishing will soon move to electronic delivery mechanisms because of the "economic push of having to cope with increasing costs of the present system and the attractive pull of the new features that electronic publishing offers" [17]. He is a supporter of recent developments in informal scholarly communication like bulletin boards and pre-print archives, but argues that there is also a need to build-in reliability mechanisms, like peer-review, so that scholars are able to build upon the accumulated knowledge.

The academic psychologist, Stevan Harnad, has similarly argued, with reference to what he calls 'esoteric' literature (i.e. that originating in specialised scholarly and scientific research and with no realistic market), that in the post-Gutenberg era there is no need to perpetuate the 'Faustian bargain' made between authors and commercial publishers whereby authors trade the copyright of works in exchange for having them published [18]. He argues that this type of bargain made sense when publishing remained a exclusive and expensive domain, but has no relevance in the electronic era when scholars can publish their own papers at little or no personal cost. In addition to the benefits of improved accessibility, increased speed of publication and possible financial savings, Harnad suggests that network publication would enable authors to interact with their peers, for example published articles could be open to immediate comment and response. This is what has been characterised by the term 'scholarly skywriting' [19].

In order to facilitate the post-Gutenberg era, Harnad, Odlyzko and others have formulated what they refer to as a 'subversive proposal' to bring down the 'paper house of cards' [20]. They suggest that all authors of 'esoteric' works should make available the texts of all current papers on the Internet and readers would rapidly form the habit of accessing the free electronic version of a paper rather than a more expensive paper version published much later [21]. Harnad considers that publishers could respond to this challenge in one of three ways [22]. Firstly, they could invoke copyright law to attempt to force the removal of these papers from the Internet. Secondly, they could give up altogether the publication of 'esoteric' serials - leaving this task to the research communities themselves who would have to implement the relevant quality control mechanisms. Thirdly, publishers could adopt an economic model where the publication cost and profit could be recovered by means of page-charges, paid for as part of a research grant or by the author's institution. Distribution would then be electronic and free of charge to the user.

The most commonly cited example of the 'subversive proposal' in action is Paul Ginsparg's 'e-print archive' established at the Los Alamos National Laboratory [23]. The original service, which went online in August 1991, gave electronic access to pre-

prints in Ginsparg's own subject area: high-energy physics. The most remarkable thing about this service was that it very quickly became the primary means of scholarly communication in its subject area. An academic physicist was quoted in 1994 as saying that the archive had completely changed the way people in the field exchanged information: "the only time I look at the published journals is for articles that predate the Los Alamos physics databases" [24].

Odlyzko, Harnad and their supporters are correct in supposing that scholarly communication methods are being changed significantly by network publishing. The use of electronic bulletin boards, mailing-lists and pre-print servers has already had a profound effect on the way research is carried out. The evidence of successive editions of the *ARL Directory* also demonstrates that scholarly electronic serials are increasingly available on the Internet. In addition, there are signs that quality control mechanisms are being applied to these serials. Many of the first scholarly serials available online were newsletters or non-refereed journals but there is today an ever increasing number of refereed journals [25]. Despite this, however, there is no evidence that commercial scholarly serials are in terminal decline. Indeed, many papers deposited in pre-print archives will eventually find their way through peer-review and into the printed journal literature. The reason for this is that online serials have not been able to replicate all of the functions currently carried out by traditional printed serials.

Cliff McKnight has pointed out that for electronic serials to be to be completely acceptable to users, they must allow them to do at least what can be done with traditional paper serials, and preferably more. Fytton Rowland, summarising the work of others, has described the main functions of traditional scholarly serials as follows [27]:

- The dissemination of information
- Quality Control
- The canonical archive
- Recognition of authors

Network publication already fulfils the important function of dissemination, indeed it could be claimed to be a better (faster and cheaper) dissemination medium than print. Quality control processes like peer-review can be added to electronic publications, as can strategies for the recognition of authors' contributions to them. This is why online serials are increasingly being used for the publication of scholarly output. There is not the same level of confidence, however, that any electronic publication - including network publication - will contribute much to the canonical archive.

While Harnad and others have been arguing for the widespread adoption of network publication of 'esoteric' works, professional publishers from the commercial sector and learned societies themselves have not neglected to investigate the potential of electronic publication of their own serials [28]. The publisher-led ADONIS experiment in document delivery was an early example of this, and this has been followed by initiatives like the Chemical Journals Online (CJO) service whereby the American Chemical Society (ACS) has made its scholarly journals available over the international Scientific and Technical Network (STN). Professional publishers have also showed themselves willing to co-operate with libraries in experiments like The University Licensing Program (TULIP) where journals in the subject area of materials science from the Elsevier/North Holland/Pergamon group were delivered to

participating US research libraries to investigate the technical, legal and economic issues associated with electronic serials and user behaviour [29]. Another example of publisher and library co-operation is the ELVYN project involving the Institute of Physics Publishing, a team at Loughborough University and others [30]. In the UK a major reason for the recent increased visibility of the online serial is the Higher Education Funding Council for England (HEFCE) pilot site initiative which has given universities online access to serial articles originating from professional publishers like Academic Press, Blackwell Science and Institute of Physics Publishing [31]. The UK eLib Programme, funded by the Joint Information Systems Committee of the higher education funding councils, has also funded some projects relating to electronic serials. The result of all this activity is a growing number of scholarly online serials, distributed in several different ways and in several different formats.

**Distribution methods and formats currently in use**

In 1994 David Pullinger suggested that there are three possible models for the network publishing of serials [32]. In the first, the 'publisher' distributes copies of contents pages, abstracts, articles or whole issues to subscribers over the Internet usually in the form of e-mail. Many of the earliest non-commercial network serials were distributed in this way, based on electronic mailing-list software like LISTSERV where serial contents pages and abstracts could be sent out to a centrally-held list of subscribers who then could request the system to deliver particular articles [33]. Pullinger's second model relies on serials being made available through local networks like Campus Wide Information Systems - as used in the ELVYN project. In the third model, serial issues/articles are held on a central host where users can browse and download relevant items. This seems to be the currently favoured model - largely because of the influence of the World Wide Web - and is used for both network publication and for the commercial distribution of scholarly serials.

   Electronic serials are also available in a variety of formats. Until comparatively recently the most popular format was plain ASCII text or bit-mapped page images. ASCII is fine for articles comprised largely of text, but is inadequate for representing visual complexity of STM literature. ADONIS, with this in mind, scanned the paper copy of the journal and distributed the pages as bit-maps on CD-ROM [34]. ASCII and bit-mapped images are still used to distribute electronic journal articles, but a variety of other formats are emerging and growing in popularity

   STM serials have been described as the most difficult (and expensive) to produce in any format because, according to Hitchcock, Carr and Hall, they contain "specialised terminology, they frequently include detailed mathematics and often have complex artwork and tabular data" [35]. For these reasons online STM serials tend to use formats which retain the features of print journals. The most popular of these formats are PostScript and its more flexible relation, Adobe's Portable Document Format (PDF) [36]. PDF's popularity has been enhanced by the free distribution of Adobe's Acrobat viewer and Acrobat's integration with Web browsers like Netscape Navigator. PDF is particularly good for use in situations where electronic versions of printed serials are being made available online as the pages will look the same as in the printed version. Its strong position in the field of online serials is attested by the

fact that all the serials included in the UK HEFCE's pilot site licence initiative are all currently made available in PDF.

The other popular format currently used for the distribution of online serials is the HyperText Markup Language (HTML). Online serials use HTML because they want to take advantage of the hypertextual and multimedia features of publishing on the Web rather than just replicate print. For example, the UK eLib funded journal Internet Archaeology chose amongst its first papers one on Roman amphorae found in Britain with 'clickable' maps and timelines [37]. Many online serials previously distributed in ASCII form by LISTSERV are now also available in HTML on the Web. HTML is not always an ideal format for STM serials as it has limitations encoding some special characters and relies on inline graphics or helper applications for the full display of illustrations [38]. Accordingly, HTML is often used to create an interface for the viewing of other formats. The International Digital Electronic Access Library (IDEAL) service from Academic Press contains contents page information and abstracts in HTML while the full-text of the article itself is available in PDF. A similar approach is taken in giving online access to the journals published by the Johns Hopkins University Press in Project Muse.

Naturally, there is interest in other formats - SGML being the obvious example. Project ELVYN, for example, had first considered PostScript to be the best delivery format for the journal Modelling and Simulation in Materials Science and Engineering, but it was eventually delivered in SGML so that it could be converted to HTML for viewing on a Web browser [39]. The Chemistry Online Retrieval Experiment (CORE), which gave access to American Chemical Society (ACS) serials, converted data from the format used by the ACS to a variant of the SGML Document Type Definition (DTD) produced by the American Association of Publishers for their Electronic Manuscript Standard [40].

Online serials differ from printed serials in their publication and use. For example, there is no essential requirement for serials to be published in regular issues as this is a constraint of printing processes which makes it cheaper and more convenient to bundle articles together for distribution. For this reason, online serials often issue individual articles as soon as they have cleared peer-review. In a similar way, the use of online serials, is going to be more orientated towards individual articles rather than issues or volumes - meaning that organisations giving access to such serials are going to have much more in common with document delivery services than they do at the moment. There will be more emphasis upon resource discovery through searching rather than browsing. In addition, scholarly online serials are partly moving away from the position where electronic serials merely replicate the functionality - and sometimes the appearance - of printed serials. An electronic serial can be a dynamic document including multimedia, active links to related publications or data and can be regularly updated to take account of comments made by scholars in reviews or other publications. Experiments with dynamic types of online journal include the eLib funded Internet Archaeology and the CLIC [41] electronic journal project. For example, CLIC has attempted to include ways in which users can acquire three-dimensional molecular data in digital form through electronic journals to act as a starting point for their own exploration of the content [43]. These new publication models will present a severe challenge for preservation. As the CLIC researchers themselves ask: "how long should any given data be expected to reside in

automatically accessible form on the Internet?". Also, can data be preserved in such a way that it is able to be retrieved in the future using mechanisms developed at the present time?

## Preservation issues

### Technological preservation issues

The root of the digital preservation conundrum is technological. Digital information has to be interpreted by a machine before it becomes intelligible to human users. There are problems associated with three aspects of digital information technology: the digital medium itself and its associated hardware and software.

### Media longevity

Ten years ago, the digital preservation literature had a lot to say about media longevity [44]. In the nineteen-eighties this was often seen as the key technological problem with what was then called machine-readable information. It is still important. Digital media, both magnetic and optical, have short lifetimes in comparison with media like paper and microfilm. Margaret Hedstom argues that the threat posed by magnetic and optical media is "qualitatively different" in that the media is reusable and they deteriorate in a matter of years, not decades [45]. The technological response to media longevity is known as 'digital refreshing', the periodic recopying of the data onto a new medium. The focus has in recent years, however, moved away from media longevity issues, not because the problems have been solved to any extent, but because there is an greater awareness of significant technological problems associated with hardware and software obsolescence.

### Hardware obsolescence and software dependence

John Mallinson noted in 1987 that one of the most serious problems with preserving electronic information is the rapid obsolescence of electronic hardware [46]. Brichford and Maher sum up this problem when they say that a "twenty-year life for the plastic backing material used for computer tapes and disks is irrelevant if the tape or disk drives on which they were recorded become obsolete and unavailable after ten years" [47]. In addition, digital information is increasingly stored in formats which are dependent upon particular software to interpret them correctly. This is known as software dependence. We have already noted that online serials are currently made available in several different ways and in a large number of different formats, and this is likely to get worse rather than better in the future as delivery mechanisms and formats change and increase in complexity. One possible answer to the problem of software dependence is to preserve all software together with the information itself, although it is debatable whether this would be a long-term solution. As an alternative, Jeff Rothenberg has suggested preserving metadata that would adequately describe systems so that future generations could build emulators to mimic the behaviour of obsolete hardware and software [48]. A more realistic answer, in the short term at least, is the concept that has replaced refreshing - data migration.

**Migration**

The Task Force on the Archiving of Digital Information defined migration as "the periodic transfer of digital materials from one hardware/software configuration to another, or from one generation of computer technology to a subsequent generation" [49]. Migration differs from refreshing in that it takes account of the hardware/software obsolescence problem. There is no point in making an exact copy on to a new medium if the software and hardware necessary to interpret it no longer exist. The point of migration is to transfer to new formats while, wherever possible, preserving the integrity of the information.

The simplest migration strategies would involve transfer to universal formats like ASCII text or flat-file data which are (relatively) software independent. Indeed, this might be the best solution for online serials primarily consisting of textual information, but would probably result in a considerable loss of functionality for most STM online serials or those publications which have tried to incorporate dynamic features. An alternative strategy would be to migrate records to a small number of 'standard' formats like PDF or particular applications of SGML. This strategy would simplify the migration process itself while helping to maintain some of the important characteristics of the original [50]. In any case, migration strategies used should be recorded as metadata and preserved together with the original item so that future users are aware of significant changes made to a document during the preservation process.

Despite their intractability, technological problems are probably not the most significant factor in the preservation of online serials. Digital information can be preserved if it is identified early enough, although, as Lesk notes, "preservation means copying, not physical preservation" [51].

**Intellectual preservation issues**

**Authenticity**

Digital information is easy to change and update. Indeed this is one of its major advantages over paper and print. However, with preservation in mind, this malleability becomes problematic. It becomes very difficult to prove that digital information has not been accidentally or deliberately corrupted at some time. This is as true of online serials as any other digital information. Corruption is not the only issue, online serial articles could be frequently updated to take account of new research and the comments of other scholars. This is not in itself a problem, but it is important that users of electronic serials are sure that the version that they are referring to is the version that they want to see. This is taken for granted in the print world where citations refer to the 'canonical archive' of publications. This concern with data integrity is characterised by Peter Graham as intellectual preservation [52].

Archivists also have an interest in data integrity and a project based at the University of British Columbia has stressed the importance of the concepts of 'reliability' and 'authenticity'. In an archival context, Luciana Duranti has defined both of these terms as follows [53]:

- Reliability refers to the authority and trustworthiness of a record as evidence.
- Authenticity proves that the document is what it claims to be.

In archives, reliability is exclusively linked to record creation. Reliability would similarly be guaranteed for an online serial by editing, peer-review and other quality control processes. Authenticity, on the other hand, ensures that a record "is protected and guaranteed through the adoption of methods that ensure that the record is not manipulated, altered, or otherwise falsified after its creation" [54]. With relation to online serials, authenticity could be promoted by the adoption of techniques based on cryptographic theory. Graham himself has suggested adopting something like digital time-stamping (DTS) which uses one-way cryptographic hashing techniques [57], but recognises that there are likely to be other solutions [56]. Other relevant initiatives might include the Digital Signature Initiative (DSig) of the World Wide Web Consortium (W3C) which is concerned with the development of trust mechanisms for the Web [57]. The importance of intellectual preservation for the future of scholarly communication cannot be over-emphasised. Without some way of ensuring authenticity over-time, it is possible that networks will ultimately be unable to support serious scholarly communication.

## Administrative preservation issues

There are technical solutions to the digital preservation of online serials, and there is some interest in problems concerning intellectual preservation. However, there remain a large number of unresolved issues. These relate to administrative or legal matters, and may actually be the most difficult problems to solve. The questions regarding copyright and ownership are the most intractable, so will be dealt with first.

## Copyright and Ownership

The growth in provision and use of electronic information resources (including online serials) has resulted in fundamental changes in the way information is owned [58]. The information embodied in traditional printed serials is normally purchased by subscription, either directly with publishers or through subscription agents. This remains the case, even when most users of this information will obtain it through an intermediary like a library or document supply service. The organisation or individual that purchases a serial will then retain physical custody of an artefact - a volume or issue - for as long as it is required. Assuming that this artefact is kept in a reasonably good environment and safeguarded against disaster (fire and flood), it should last a long time. Given the fact that printed serials will normally be subscribed to by more than one organisation, a distributed 'canonical archive' of scholarly serials will be built-up in this way. Long-term preservation (and access) is essentially a by-product of this process and does not require specific initiation.

The situation of electronic information resources is quite different. The 'purchaser' of an electronic resource, unless it is an artefact like a CD-ROM, does not necessarily retain 'physical custody' over it. Concerns over copyright mean that the current practice is for commercial publishers to licence the use of information to customers, thus ensuring that the use of this information is governed by contract law rather than copyright law [59]. A licence for an commercial online serial might only give the 'subscriber' specific rights over use of a particular serial or group of serials for a limited amount of time. This is where the position becomes problematic. What

happens when the subscribing institution decides to cancel its subscription? Will all access rights to that journal, including 'volumes' already 'paid-for', be removed? What would happen if the serial ceases publication or the publisher goes out of business? Research organisations and libraries might find that they have no direct control over which particular online serials would be preserved as part of the 'canonical archive'. Publishers, especially network publishers, currently seem to be happy to actively encourage browsing, downloading and copying of online serial articles for personal use and also permit the distribution of copies for research purposes [60], but may have to be persuaded that ensuring long term access and preservation of their online serials is a desirable objective and that co-operation with other stakeholders (e.g. research libraries or data archives) may be the best way to do this. After all, authors may be reluctant to contribute articles to serials that may not exist very far into the future. In the short term, at least, solutions to these problems will have to be worked out with co-operation between publishers and libraries [61].

**Management issues**

**Who should preserve?**
In the past, preservation has been the responsibility of organisations (like libraries and archives) that give access to information rather than the producers of that information themselves. Two broad approaches are possible, firstly a centralised repository model, the other a radical decentralised 'non-custodial' model.

   National libraries have traditionally conformed to the centralised model using legal deposit legislation to ensure that all relevant published works are collected and preserved. A few have successfully extended legal deposit legislation to cover electronic publications, but not this does not usually include online publications. The British Library, for example, has recently requested the extension of legal deposit to electronic media, but specifically excluded online publications because of their problematic nature [62]. One exception is the National Library of Canada's experimental Electronic Publications Pilot Project which has been identifying and making copies of Canadian online serials, texts, etc. with the co-operation of publishers. Despite this, it is clear that the physical transfer of online publications to central repositories is not going to be a cost-effective solution. It is likely that online publications will require a more decentralised model.

   This model originates from ideas promoted by David Bearman (and others) for archives [63]. They have argued that record creating organisations should retain responsibility for their electronic records while centralised archival repositories would take these over only as a last resort to ensure their preservation. The Task Force on the Archiving of Digital Information, following this, proposed that digital information creators, providers and owners should have the initial responsibility for archiving while certified digital archives should be given the right and duty to exercise an aggressive fail-safe rescue function [64]. This sort of approach is attractive for online serials because most publishers will have an interest in maintaining access over time and will migrate the information themselves until such time as the serial is no longer of interest or economic value. Hardware will be periodically upgraded, sometimes formats will change, but the serial need not go 'out of print'. The crucial time will

come when publishers no longer have any interest in keeping a serial going or cease to exist. It is at that time that fail-safe mechanisms should be activated. The role of central repositories would, thus, be reduced but it would also emphasise the need for leadership and co-ordination from information professionals [65]. The role of organisations like the Commission on Preservation and Access in the US and the National Preservation Office in the UK will be crucial in co-ordinating activity in this general area.

**What should be preserved?**
Deciding what needs to be preserved is another problem. The temptation will be, with digital storage devices becoming cheaper and more compact, to keep everything. However, preservation of digital information is still likely to be expensive - continuous migration will be time consuming and requires technical expertise - and as the amount of digital information grows, is likely to remain so. Therefore, some kind of selection before preservation will be necessary.

For scholarly online serials, this process will be made easier by the reliability mechanisms that exist. Peer-reviewed serials will be obvious candidates for preservation as will less formal, newsletter-type, publications issued by learned societies and research organisations. It may be harder to assess articles 'published' in electronic pre-print archives or the products of scholarly 'skywriting'. If Odlyzko's vision of an integrated system combining informal methods of communication (like electronic bulletin boards) with formal methods (like peer reviewed online serials) comes to fruition, there may be a case for preserving some information that has not been peer-reviewed or subject to any other quality control processes. This could build either upon appraisal techniques developed for records selection by the archives community or upon the quality selection criteria being adopted by some Internet subject services [66].

The nature of digital information also means that such appraisal will have to take place very early in its life-cycle. It will not be possible to wait until the information has become unavailable or 'out-of-print' before preservation or no copies may be in existence to be preserved. Instead, the identification of relevant items will have to be made almost at the time of 'publication' and rights for its long term preservation negotiated. It would be an additional advantage if these processes of selection and appraisal were continuous, indeed with migration as the preservation strategy there will be regular opportunities to reassess the value of the information being migrated.

# Conclusion

This paper has attempted to outline some of the problems which will need to be confronted to ensure the continued existence of and accessibility of the information embodied in the scholarly online serial. Other important issues exist which have not been discussed here, most notably the economic implications of digital preservation Many of these issues are currently being investigated by librarians, publishers, computer scientists and other stakeholders in digital preservation, but it is important to realise that many of the problems will only be solved with practical experience of ensuring the preservation of digital information and with practical co-operation.

It is easy to become despondent when considering the magnitude of the challenge that digital preservation poses to scholarly communication and those organisations who currently support it, but there are grounds for a cautious confidence in our ability to make the information embodied in online serials available for future generations of scholars.

One more point needs to be considered. In the past, preservation was an activity considered only by specialists within the library and information professions. With the advent of digital preservation, it might become the primary function of the digital research library. All other activities, including resource discovery and access, may ultimately become dependent upon digital preservation.

## Notes and references

1. Task Force on Archiving of Digital Information, *Preserving digital information* (Washington, D.C.: Commission on Preservation and Access, 1996).
2. John Feather, *Preservation and the management of library collections* (London: Library Association Publishing, 1991), p. 2.
3. Charles M. Dollar, *Archival theory and information technologies: the impact of information technologies on archival principles and methods* (Ancona, Italy: University of Macerata, 1992), p. 66.
4. David Bearman, *Archival methods*, Technical Report, vol. 3, no. 1 (Pittsburgh, Pa.: Archives and Museum Informatics, 1989), pp. 17-27.
5. James M. O'Toole, "On the idea of permanence," *American Archivist* 52, no.1 (Winter 1989): 10-25.
6. Michael Alexander, "Virtual stacks: storing and using electronic journals," *Serials* 10, no. 2 (July 1997): 173-178.
7. Michael Breaks, "The digitisation of journal literature: towards sustainable development," *Serials* 10, no.2 (July 1997):164-172.
8. Hazel Woodward, "The impact of electronic information on serials collection management," *IFLA Journal* 20, no. 1 (1994): 35-45.
9. Hazel Woodward, "Electronic journals in libraries," in *Project ELVYN: an experiment in electronic journal delivery*, ed. Fytton Rowland, Cliff McKnight and Jack Meadows (London: Bowker-Saur, 1995): 49-64, p. 51.
10. Robert M. Campbell and Barrie T. Stern, "ADONIS: a new approach to document delivery," *Microcomputers for Information Management* 4, no. 2 (1987): 87-107.
11. *ARL Directory of electronic journals, newsletters and academic discussion lists*, 6th ed. (Washington, D.C.: Association of Research Libraries, 1996).
12. Brian Shackel, *BLEND-9: overview and appraisal*, British Library research paper, 82, (London: British Library, 1991).
13. Fytton Rowland, "Recent and current electronic journal projects," in *Project ELVYN: an experiment in electronic journal delivery*, ed. Fytton Rowland, Cliff McKnight and Jack Meadows (London: Bowker-Saur, 1995): 15-36.
14. Dennis P. Carrigan, "Research libraries evolving response to the 'serials crisis'," *Scholarly Publishing* 23, no. 3 (April 1992): 138-151.
15. Ann Okerson, "Publishing through the network: the 1990s debutante," *Scholarly Publishing* 23, no. 3 (April 1992): 170-177.
16. Andrew M. Odlyzko, "Tragic loss or good riddance? The impending demise of traditional scholarly journals," *International Journal of Human-Computer Studies* 42 (1995): 71-122.
17. Odlyzko, "Tragic loss," p. 83.

18. Stevan Harnad and Jessie Hey, "Esoteric knowledge: the scholar and scholarly publishing on the Net," in *Networking and the future of libraries 2: managing the intellectual record*, ed. Lorcan Dempsey, Derek Law and Ian Mowat (London: Library Association Publishing, 1995): 110-116.

19. Stevan Harnad, "Scholarly skywriting and the prepublication continuum of scientific inquiry," *Psychological Science* 1 (1990): 324-343.

20. Ann Okerson and James O'Donnell, eds., *Scholarly journals at the crossroads: a subversive proposal for electronic publishing* (Washington, D.C.: Association of Research Libraries, 1995).

21. Harnad and Hey, "Esoteric knowledge," pp. 114-115.

22. Stevan Harnad, "Electronic scholarly publication: quo vadis?," *Serials Review* 21, no. 1 (1995): 70-72.

23. Paul Ginsparg, "First steps towards electronic research communication," *Computers in Physics* 8, no. 4, (July/August 1994): 390-396.

24. Steven B. Giddings, quoted in: Gary Stix, "The speed of write," *Scientific American* 271, no.6 (December 1994): 72-77.

25. Jack Meadows, David Pullinger and Peter Such, "The cost of implementing an electronic journal," *Journal of Scholarly Publishing* 26, no. 4 (July 1995): 227-233.

26. Cliff McKnight, "The human factors of electronic journals," in *Project ELVYN: an experiment in electronic journal delivery*, ed. Fytton Rowland, Cliff McKnight and Jack Meadows (London: Bowker-Saur, 1995): 37-47.

27. Fytton Rowland, "Print journals: fit for the future?," *Ariadne* 7 (January 1997): 6-7. Available: <URL:http://www.ariadne.ac.uk/issue7/fytton/>

28. Fytton Rowland, "Recent and current electronic journal projects," in *Project ELVYN: an experiment in electronic journal delivery*, ed. Fytton Rowland, Cliff McKnight and Jack Meadows (London: Bowker-Saur, 1995): 15-36.

29. Jaco Zijlstra, "The University Licensing Program (TULIP): a large scale experiment in bringing electronic journals to the desktop," *Serials* 7, no. 2 (1994): 169-172.

30 Fytton Rowland, Cliff McKnight and Jack Meadows, eds., *Project ELVYN: an experiment in electronic journal delivery* (London: Bowker-Saur, 1995).

31. Bahram Bekhradnia, "Pilot national site licence initiative for academic journals," *Serials* 8, no. 3 (November 1995): 247-250.

32. David Pullinger, "Journals published on the Net," *Serials* 7, no. 3 (November 1994): 243-248.

33. Cliff McKnight and John Richardson, "The impact of new publishing media," in *The international serials industry*, ed. Hazel Woodward and Stella Pilling (Aldershot: Gower, 1993): 89-105.

34. Barrie T. Stern and Henk C. J. Compier, "ADONIS: document delivery in the CD-ROM age," *Interlending and Document Supply* 18, no. 3 (1990): 79-87.

35. Steve Hitchcock, Leslie Carr and Wendy Hall, "A survey of STM online journals, 1990-95: the calm before the storm," in *ARL Directory of electronic journals, newsletters and academic discussion lists*, 6th ed. (Washington, D.C.: Association of Research Libraries, 1996): 7-32. Available: <URL:http://journals.ecs.soton.ac.uk/survey/survey.html>

36. Judith Wusteman, "Electronic journal formats," *Program* 30, no. 4 (October 1996): 319-343.

37. Paul Tyers, "Roman amphoras in Britain," *Internet Archaeology* 1 (1996). Available: <URL:http://intarch.ac.uk/journal/issue1/tyers_index.html>

38. Richard Entlich, Lorrin Garson, Michael Lesk, Lorraine Normore, Jan Olsen and Stuart Weibel, "Testing a digital library: user response to the CORE project," *Library Hi Tech* 14, no. 4 (1996): 99-118.

39. Cliff McKnight, Jack Meadows, David Pullinger and Fytton Rowland, "ELVYN: publisher and library working towards the electronic distribution and use of journals," [in Digital

Libraries '94: Proceedings of the First Annual Conference on the Theory and Practice of Digital Libraries, June 19-21, 1994 - College Station, Texas]. Available: <URL:http://csdl.tamu.edu/DL94/paper/mcknight.html>

40. Michael E. Lesk, "Electronic chemical journals," *Analytical Chemistry* 66, no. 14 (15 July 1994): 747A-755A.

41. CLIC is the acronym for the four organisations involved in the project: Cambridge site of the Royal Society of Chemistry; Leeds University; Imperial College of Science, Technology and Medicine, University of London; Cambridge University.

42. David James, Benjamin J. Whitaker, Christopher Hildyard, Henry S. Rzepa, Omer Casher, Jonathan M. Goodman, David Riddick and Peter Murray-Rust, "The case for content integrity in electronic chemistry journals: the CLIC project," *New Review of Information Networking* (December 1995). Available: <URL:http://www.ch.ic.ac.uk/clic/video.html>

43. Omer Casher, Gudge K. Chandramohan, Martin J. Hargreaves, Christopher Leach, Peter Murray-Rust, Henry S. Rzepa, Roger Sayle and Benjamin J. Whitaker, "Hyperactive molecules and the World-Wide-Web information system," *Journal of the Chemical Society: Perkin Transactions* 2, no 1 (January 1995): 7-11.

44. Michael W. Day, *Preservation problems of electronic text and data* (Loughborough: EMBLA Publications, 1990).

45. Margaret Hedstrom, "Preserving the intellectual record: a view from the archives," in *Networking and the future of libraries 2: managing the intellectual record*, ed. Lorcan Dempsey, Derek Law and Ian Mowat (London: Library Association Publishing, 1995), p. 180.

46. John C. Mallinson, "On the preservation of human- and machine readable records," *Information Technology and Libraries* 7 (1988): 19-23.

47. Maynard Brichford and William Maher, "Archival issues in network electronic publications," *Library Trends* 43, no. 4 (Spring 1995): 701-712.

48. Jeff Rothenberg, "Ensuring the longevity of digital documents," *Scientific American* 272, no. 1 (January 1995): 24-29.

49. Task Force on the Archiving of Digital Information, *Preserving digital information*, p. 6.

50. Margaret Hedstrom, "Preserving the intellectual record," p. 185.

51. Michael Lesk, *Preservation of new technology: a report of the Technology Assessment Advisory Committee of the Commission on Preservation and Access* (Washington, D.C.: Commission on Preservation and Access, 1992).

52. Peter S. Graham, *Intellectual preservation: electronic preservation of the third kind*. (Washington, D.C.: Commission on Preservation and Access, 1994).

53. Luciana Duranti, "Reliability and authenticity: the concepts and their implications," *Archivaria* 39 (Spring 1995): 5-10.

54. Luciana Duranti and Heather MacNeil, "The protection of the integrity of electronic records: an overview of the UBC-MAS Research Project," *Archivaria* 42 (Fall 1995): 46-67.

55. Hashing techniques use algorithms which convert the arrangement of all characters, symbols, graphics, etc., within a particular document into a unique hash value which can be stored and retrieved as metadata. Any change (however small) to the document will produce a different hash value when it is converted using the same algorithm. The process is described as "one-way" because there is no means of recreating the original document from its hash value.

56. Peter S. Graham, "Long-term intellectual preservation," in *Digital imaging technology for preservation*, ed. Nancy E. Elkington (Mountain View, Calif.: Research Libraries Group, 1994): 41-57.

57. World Wide Web Consortium, *Digital Signature Initiative*. Available: <URL:http://www.w3.org/Security/DSig/>

58. Ann Okerson, "What academic libraries need in electronic content licenses," *Serials Review* 22, no. 4 (Winter 1996): 65-69.
59. Okerson, "What academic libraries need," p. 65.
60. Pamela Pavliscak, "Trends in copyright practices of scholarly electronic journals," *Serials Review* 22, no. 3 (Fall 1996): 39-47.
61. Ann Okerson, "Some economic challenges in building electronic libraries: a librarian's view," [Paper presented at IFLA Congress, Beijing, August 1996]. Available: <URL:http://www.library.yale.edu/~okerson/ifla.html>
62. British Library Research and Innovation Centre, *Proposal for the legal deposit of non-print publications* (London: British Library, 1996), Section 2D. Available: <URL:http://www.bl.uk/services/ric/legal/legalpro.html>
63. David Bearman, "An indefensible bastion: archives as repositories in the electronic age," in *Archival management of electronic records*, ed. David Bearman, Technical report, no. 13 (Pittsburgh, Pa.: Archives and Museum Informatics, 1991): 14-24.
64. Task Force on the Archiving of Digital Information, *Preserving digital information*, p. 21.
65. Margaret Hedstrom, "Electronic archives: integrity and access in the network environment," in *Networking in the humanities*, ed. Stephanie Kenna and Seamus Ross (London: Bowker-Saur, 1995): 77-95.
66. Gregory F. Pratt, Patrick Flannery and Cassandra L. D. Perkins, "Guidelines for Internet resource selection," *College and Research Libraries News* 57, no. 3 (March 1996): 134-135.

## Acknowledgements