Journal Backfiles in Scientific Publishing
A Marketing White Paper
Journal Backfiles in Scientific Publishing

The British Library has been working with a number of STM publishers, to help them begin or complete programmes of backfile digitisation. We wanted to further explore the market place for digitisation and assess the impacts and benefits on publishers, customers, end users and authors. We commissioned Jan Willem Wijnen to produce this marketing white paper, to give an evaluation of the current market and how it will look moving forward.

Jan Willem Wijnen (Deventer, The Netherlands) is an independent consultant in online publishing. He has worked for several years as publishing editor for chemistry at Kluwer Academic Publishers and later as business developer at KAP, later Springer. In that role he was project manager of Springer’s Online Journal Archives project. As a consultant he has worked on backfiles projects for Sage Publications and Brill.

This marketing white paper is only part of a more detailed report. For a complete copy, including a Competitor Analysis Journal Backfiles please email marketing@bl.uk.
Table of Contents

1. Introduction

2. Reasons for doing a backfiles project
   2.1. Usage of old(er) research/scientific articles
   2.2. Publisher’s perspective
   2.3. Customer’s perspective
   2.4. Author’s perspective

3. Journal backfiles project: important aspects
   3.1. Sourcing
   3.2. Production
   3.3. Hosting, back-office and fulfillment
   3.4. Setting propositions, business models and prices
   3.5. Sales and Marketing

4. Are journal backfiles a commercial success?
1. Introduction

Online publication of scholarly information is steadily becoming more important than publication in print. Online publication of books occurs on a modest scale and has not yet been introduced by all major scholarly publishers. However, all publishing houses now publish their journals online, in most cases since the late 1990’s. A more recent major development in online publishing is digitisation and online publication of backlist journal publications, in some cases back to articles published in the early 19th century.

This white paper describes various aspects of the production and publication of journal backfiles for scholarly publishers. Publication of backfiles of books falls outside of the scope of this paper, also because there are very few examples of this activity. Additionally, publication of backfiles in other types of publishing falls outside of the scope of this paper.

2. Reasons for doing a backfiles project

Why would publishers recycle content they had published decades ago? With respect to this question, a very useful comparison is a similar development in the music industry in the 1990’s. With the introduction of compact disc technology, music started to be released on CDs, besides vinyl. Initially, only new music was released (published) on CDs, but over time older records were also re-released on CDs, simply because the audience had become used to the new and different technology. They had replaced their old equipment with new technology and therefore had to replace the old medium (vinyl) with new media. The film-industry (video to DVDs) is a similar example. These developments have offered these industries tremendous opportunities, because now all Beatles records could be sold again to their fans, while at the same time, production costs are low, because the music is there, it does not have to be recorded again. Publication of backfiles is very similar.

2.1. Usage of old(er) research/scientific articles

The very first important question to answer is “are scholars/researchers/scientists interested at all in reading older academic journals articles?”.

The reading behaviour of researchers and scholars has been investigated. Carol Tenopir and Donald King showed in a study (Learned Publishing 15(4), 2002) that about 15% of articles read by scientists were five years or older. Analysis of usage statistics of STM publisher’s platforms has revealed that about 20 – 25% of the downloaded articles are at least 5 years old. Surprisingly, the percentage of old article downloads is higher in biomedical areas than in humanities.

In discussing the value of older scientific literature, a frequently used argument is the assumption that certain areas of science (biomolecular sciences is a typical example) move so fast that literature older than a few years is outdated, no longer relevant, so why would a library spend all that money? While it is true that knowledge and insights in these scientific disciplines change rapidly, this does not mean that earlier research results are no longer valid. Scientific and academic knowledge/insights normally come from existing facts, data and theories combined with recent accomplishments. Therefore, older research results must be consulted (again) to develop new insights. Sometimes that only means re-reading a paragraph or checking a table.

In conclusion, scientists and researchers are using substantial amounts of older literature. Making these articles accessible via a nearby computer instead of a distant library would certainly be attractive for them. It is very likely that older articles will be used even more extensively if additional online functionality is added to these articles, such as links.
2.2. Publisher’s perspective

For publishers, a backdigitisation project presents a sizable investment, simply because of the quantity of material to be digitised. Still, there are several good reasons for publishing houses to initiate a journal backfiles project.

2.2.1. Revenue

Despite the high investment, digitisation and selling backissues can be quite profitable. Journal backfiles are not a completely new online publication. The addition of older articles is merely an extension of a successful concept (online journals). Within the industry it is known that the introduction of journal backfiles has been profitable for the first publishing houses that have produced and introduced them.

2.2.2. Increase usage of internet platform

Online publications are gaining importance and for the major publishers, online journal revenue is now higher than print revenue. Sales models are changing as well, with usage statistics becoming an increasingly important factor in negotiations between publishers and customers, because they are a good indicator of the value of the publisher’s publications.

Adding more content to a publisher’s platform will increase the usage, not only because there are more articles end-users can find and download, but also because more inter-article links can be introduced, which allow end-users to click from one paper to another. CrossRef even enables end-users to click to articles from different publishers and back.

2.2.3. Stronger customer relations

Most publishers offer backfiles for a one-time fee, so customers own the backfiles or have perpetual access. Consequently, customers will have an ongoing relationship with the publisher, they will continue to be linked to the publisher’s platform. This relationship will greatly facilitate all sorts of up-selling activities for new content, new functionalities or new tools.

Publishers will have to define a policy for cases when a customer wants to cut links with the publisher and normally there is a policy that specifies that customers are allowed to receive backfiles on a disk and locally host them.

2.2.4. Service to authors, editors, societies

Most major publishers have backdigitised their complete journal list and consequently this is becoming a standard service to authors and editors of journals and also to societies that are associated with particular journals or perhaps hold ownership them. These days, researchers will rarely use current or recent information if it is not online, and soon the same will hold for older information. Therefore it will become increasingly difficult for publishers to explain why their backissues are not online.
2.2.5. Getting and using a complete archive

Most publishers do not own a complete print archive. Often major parts of their archives are stored in distant locations. Or it is unclear exactly what is in that archive. Having a complete and easily accessible journal archive is a valuable asset in itself, but it also offers publishers new possibilities. For example, there may be a market for a collection of breakthrough 19th century literature.

2.3. Customer’s perspective (libraries)

Online journal backfiles offer many advantages to libraries and librarians:

2.3.1. Cost reduction

Acquisition of a journal’s backfile is normally not cheap, but it is often a one-time cost and, in the long run, does offer considerable cost reduction for libraries. A substantial percentage of library budgets is spent on housing, because the immense amount of paper accumulated through the years has to be shelved or stored somewhere. Replacing all print material with an online version will lead to a substantial cost reduction. Also costs for interlibrary loans (both providing and receiving), making photo-copies, maintaining an archive (often at a different location than the main library) can be reduced considerably.

2.3.2. Completing and expanding collections

All collections in all libraries suffer from gaps, for a variety of reasons: journal issues or even volumes that were borrowed and not returned, or misplaced and never found again, or damaged. Some subscriptions may not have started from Volume One. Or subscriptions were cancelled and later continued again. It is very attractive for libraries to have complete collections, and while this is almost impossible in the print world, online collections will be complete. So, acquisition of an online journal collection offers librarians a rare opportunity to have a truly complete collection. For many librarians, the acquisition of backfiles will expand access to information that had not previously been available in their libraries.

2.3.3. Increased service to end-users

Acquisition of journal backfiles offers librarians the opportunity to make this information available to the desktops of their patrons. Information that previously was not easily available to end-users can now be disclosed. Most end-users will not understand why certain journals are only online accessible from 1999 onwards, when the journal has existed for much longer. By offering backfiles, libraries enable researchers to follow trails of citations, thereby facilitating their work. And similar to online current content, also backfiles will have additional functionality, such as linking or full-text searchability.
2.4. End-users perspective

Online backfiles are a real benefit to the end-users, the readers, researchers, scientists and scholars.

2.4.1. Using older literature

As mentioned, a substantial percentage of the literature used by researchers and scientists is older than five years. By now, most of the younger-than-5-years literature is online accessible to them and therefore they are used to searching, finding and using most of this information from the internet. Undoubtedly, end-users also want the older literature (virtually all 20th century literature) to be accessible in the same way. These days, researchers will hardly use current or recent information if it is not online, and the same holds for older information.

2.4.2. Quickly assessing need for literature

Having backfiles online, available to end-users, makes it much easier for them to assess whether a publication is really of interest to them. Formerly, end-users ordered an article (through ILL), only to discover after receipt that the article is not relevant and not worth the effort and cost to get the article.

2.4.3. Finding older literature

Besides having the information available via their desktops, it becomes a lot easier for end-users to find such information, because all sorts of databases are now able to index the backfiles. In fact, it is very likely that end-users will be able to easily locate valuable and previously unknown information, due to the full integration of older literature in the internet.

2.4.4. Linking to older literature

Having journal backfiles available on the internet allows articles to be integrated in CrossRef, both inbound and outbound links and related databases. Again, this is a very valuable tool for academics and researchers, which will certainly facilitate their work.

2.4.5. Research tool

Online backfiles dramatically strengthen the role of the internet as a research tool. Research papers are now online linked, making it a lot easier to follow the trail of scientific thinking, to connect theories to new data and to combine hypotheses with old research.

2.4.6. Online availability of own research

Scientists, scholars and researchers are happy to see their own older work published online as well. These days, it is almost as if publications do not exist, if they are not on the internet.
3. Journal backfiles project: important aspects

A complete journal backfiles project consists of the following parts:

3.1. Sourcing
3.2. Production
3.3. Hosting, back-office and fulfilment
3.4. Setting propositions, business models and prices
3.5. Sales and Marketing

3.1. Sourcing

In order to digitise old journal material, original material must be available as source material for the digitisation process. Many publishers do not have a complete archive, due to mismanagement or journal take-overs in the past, and therefore the missing issues must be found elsewhere. Getting a complete collection of source material is often the most difficult aspect of a journal backfiles project. And naturally, the older missing issues are the hardest to find.

The first step in sourcing is to get a complete overview of the journal material that the publisher owns, stored in warehouses, archives, or offices. Some of this material may be available as microfilm or in another format. There may also be material available from partners such as printers or antiquarians. It is also possible that backfiles are already available through organisations such as JSTOR (US-based organisation that makes backfiles of various publishers accessible to libraries) or subject-specific organisations (for example heinonline.org (law)). Once a complete overview of available source material exists, an overview of missing material can be compiled.

To be able to find missing source material, good bibliographic data is important. Many journals have had title and/or ISSN changes in the past. Current journals may also be the result of two former journals merging into one. Even more complicated publication histories do exist. Often research of library catalogues is necessary to get a complete picture of the publication history of a journal, in order to know which source material is missing. Most libraries provide that service.

There are several methods for retrieving these issues:

(1) Contact former editors, co-workers or societies, as they may have maintained their own archive. This is a good option, but can be quite labour-intensive, because for each single title new contacts have to be approached. Also, this option usually leads to modest successes. In addition, these contacts will normally insist their archival copies are returned intact after the digitisation process.

(2) Purchase journal material from so-called back-set traders, businesses that offer antiquarian or reprints of back issues. These back-set traders normally have substantial antiquarian material in stock and a network of libraries and other institutes that will allow them to scan holdings. Prices are published in the range GBP £50–150 per journal volume.

(3) Work with libraries. Some libraries offer a service that will locate and scan source material, from their own holdings and sometimes also from other libraries. Large national libraries or library organisations with extensive holdings are very useful. Besides providing new business to these libraries, scanning and preserving print material is also in the interest of libraries themselves. Libraries generally have access to more source material than back-set traders.
Getting the source to the digitisation vendor can be logistically challenging, because it will not be possible to ship all material at once. Material will come from different providers, at different times, in different formats and some of the material has to be returned intact.

3.2. Production

The source material, either print or scanned, needs to be further processed to become true backfiles. Metadata must be added, indexes prepared, full-text searchability enabled. Many vendors can do the actual scanning and processing of source material to backfiles. Typically these companies offer many services for production of online and print journals, such as editing, typesetting, adding metadata and conversion.

The actual digitisation normally starts with scanning the journal pages. The fastest and cheapest way is to first de-spine the journal issues and feed the resulting loose pages into a scanning machine. In case journal issues should be left complete, pages can be turned manually (and scanned using a camera). There is also scanning technology available in which pages are turned by the equipment. This so-called non-destructive scanning is more expensive. There are also different prices for black and white scanning, greyscale or colour. When publications contain many images, more expensive scanning methods need to be used.

Using OCR (Optical Character Recognition) software the scans can be “read” and converted to raw ASCII or XML-files (normally with accuracy of at least 98%). These files are hidden in the backfile, but do allow full-text search of the scans.

The next step is adding metadata. Elements such as “author”, “article-title”, “abstract” and “references” are identified and marked in the text. These tags can be used to create tables-of-content, indexes, RSS feeds, and links. Also DOI’s are added to the backfiles, so they can be registered with CrossRef and stored in databases. CrossRef charges a registration fee per article.

3.3. Hosting, back-office and fulfilment

The amount of backfiles normally exceeds existing contemporary online content and a backfiles project requires additional disk space, processing time and resources. Part of a backfiles project is making sure the hosting vendor can deal with much more content and prepares for the operation. There are also some design issues: Can the current journal website present over 100 volumes? How to deal with journals that used to have a different title and ISSN?

Similarly, customer databases must also be adjusted, because customers will now be able to purchase and have access to more annuals worth of content. Access management processes should be adjusted. The invoicing has to be adjusted. Also contracts, licensing agreements and commitments should be reviewed.
3.4. Setting propositions, business models and prices

Many publishers offer backfiles in sets or collections. There are many advantages to both publisher and customer. Collections are a relatively sizable product (in terms of number of pages), thereby enabling publishers to keep prices (per page or per volume) relatively low. Offering one or just a few collections greatly simplifies administrative procedures on both sides. It is a lot easier for libraries to be represented by a consortium, because the product is simple and there are no ongoing obligations like subscriptions. Very few publishers offer single title backfiles.

A few business models exist in the marketplace:

- One-off purchase. Generally one-off purchase prices are related to the amount of content in a collection and range between USD $0.10 and 1.00 per article. This makes the price of an online backarticle about 20-fold cheaper than a current article. A collection containing dozens of backfile titles can easily cost over GBP £10,000, but that collection can contain as many as 1 million pages. All large commercial publishers use this business model. Sometimes customers can pay in instalments.

- Subscription or lease. Typically 10% is charged on top of the regular subscription fee. Society and university publishers use this model. There are also lease-to-own offers.

3.5 Sales and Marketing

The main customers for journal backfiles, of course, are a publisher’s existing online subscribers. Backfiles are a logistical next step in selling online content to them. The majority of those customers also want access to the earlier journal volumes, but they have budgetary limitations. Using different sales models will help in this respect.

In recent years individual libraries have joined forces in consortia, thereby strengthening their negotiation position. These consortia, sometimes representing all academic libraries in a country, negotiate directly with publishers about national licenses. In some cases the outcome is the so-called “big deal”, a multi-year agreement in which all participating libraries get access to all online journal content of the publisher, for a fixed price.

Adding journal backfiles to such a consortium license is a logical next step. Publishers report sales of backfiles collections to such consortia or national organisations. Some examples:

“…significant national deals in the UK and Germany, and large consortia deals in Japan and Australia.” (see http://www.oxfordjournals.org/access_purchase/archives.html)

“The Joint Information Systems Committee (UK) and Technische Informationsbibliothek Hannover (Germany) have licensed the Institute of Physics Journal Archive (1874-1998) in perpetuity on behalf of the UK and German education community.” (http://journals.iop.org/archive/)

The business model that offers backfiles for a one-off purchase with perpetual access (which is close to really owning the files) is especially attractive to consortia and national organisations. An interesting aspect is that the one-off character of the acquisition is also attractive. The organisations just need to find money once, there is not a large ongoing financial obligation. Typically, backfiles sales are highest at the end of the calendar year, when leftover budgets must be spent.
4. Are journal backfiles a commercial success?

Elsevier was the first major publisher to set up a full journal backfiles project and nearly all other larger publishers have followed, others have announced they will do so. At industry conferences, meetings and presentations publishers have spoken about the commercial successes. Journal backfiles are popular products amongst libraries and their value is appreciated. Backfiles collections’ purchase prices are generally high and therefore often no immediate budget is available. But payment in instalments presents a solution. Purchase of journal backfiles is an obvious strategic acquisition and it seems inevitable that libraries will replace all their print holdings with online holdings, and likewise their archives, in a development very much like the music-lovers replacing all their vinyl with compact discs, and in the future with other medium formats.

As mentioned before, journal backfiles are especially interesting for the largest customers, organisations that represent many libraries (consortia) or organisations that have a central role (national libraries, foundations). Websites of national organisations announce national licenses for complete backfiles archives, for example:

- Anatolian University Libraries Consortium (Turkey): http://www.ankos.gen.tr
- NESLI (UK’s national initiative for the licensing of electronic journals on behalf of the higher and further education and research communities): http://www.nesli2.ac.uk/news_010804.htm

All in all, it is fair to say that journal backfiles are a commercial success.
The British Library – Publisher Digitisation Service

The British Library, Publisher Digitisation Service offers content auditing and PDF article provision to rightsholders who are either starting a digitisation programme or simply looking to fill gaps in an existing programme.

Why choose the British Library?

■ **Content Sourcing:** The British Library has an unrivalled collection of serial content, meaning we can save you time and money in locating material. A recent press release about Sage and their programme indicated that they had to locate 75% of the content for their programme from third parties.

■ **Logistics:** Sourcing content from multiple locations can create logistical problems. This risk is removed when working with the British Library as content is located and digitised in one place and can be delivered by FTP.

■ **Reliability:** We have already worked with a number of leading publishers to support their digitisation programmes and because all work is undertaken on site we can provide timely outputs to agreed daily targets. Projects are also managed by a dedicated Project Manager.

■ **Value:** The combination of reasons above means we can save you time and money when managing your digitisation programme.

In addition to our basic service the Library is also able to provide SGML services through partnerships to ensure your content is searchable once the digitisation programme is complete.

‘The British Library proved to be an incredibly valuable resource of journal content. The British Library were easy to work, responsive, expeditious, well organized, and provided hi-quality images of the content that was requested. It is fair to say that SAGE would have been unable to meet its ambitious goal of locating and digitizing its content without the assistance and expertise of the British Library.’

John Shaw, Director, Publishing Technologies, SAGE Publications.

To find out more about our digitisation service, to see how much of your collection we hold and to get a project quote please contact publisher-digitisation@bl.uk or go to www.bl.uk/digitisationservices.