Revisiting ‘obsolescence’ and journal article ‘decay’ through usage data: an analysis of digital journal use by year of publication

David Nicholas a,*, Paul Huntington a,1, Tom Dobrowolski a,2, Ian Rowlands b,3, Hamid R. Jamali M. a,1, Panayiota Polydoratou b,4

a SLAIS, University College London, Gower Street, London WC1E 6BT, UK
b Department of Information Science, City University, London EC1V 0HB, UK

Received 27 September 2004; accepted 3 March 2005

1. Introduction

The publication age or date of documents used (or not used) has long fascinated researchers and practitioners alike. Much of this fascination can be attributed to the weeding opportunities the data is thought to provide for libraries in their never-ending battle to find the space to accommodate their expanding collections. In general journal article age studies have shown an initial increase in use/citation, then a gradual or sharp decline, depending on the discipline concerned. This characteristic has been termed obsolescence or decay and was largely measured, in the absence of accurate journal usage/borrowing data, by citations. In the sciences the decay rate was shown to be the greatest. This was largely put down to the rapid obsolescence of much scientific content. New research findings, methods or ensuing events rendered the material obsolescent. Of course, when reviewing the data we need to be reminded of the fact that citation studies reveal “use” by authors, whereas library loans or downloads represent actual use by readers, and it is readers that libraries and digital libraries principally target.

Clearly the fall of in use with time must have also been a function of the way that libraries arranged their material (in reverse chronological order); a lack of time and patience will inevitably result in readers aborting their searches after a few years and those few years will be the most recent ones. Similarly, it must also have been a function of the difficulties of searching hard-copy back volumes/issues in libraries over time.
In a hard-copy, university library this took time, it was laborious and was also hit and miss because of missing issues, journals being bound, etc.

Even when searching moved to commercial online hosts, like Dialog, the default hit list was furnished in reverse chronological order, so reinforcing the established order. However, now when you search vast digital full-text libraries like Emerald Insight⁵ or Blackwell Synergy⁶ the default search is a relevance one and is conducted on the whole dataset—in Emerald’s case that is back 10 years or more. Such an approach clearly uncovers material that would otherwise not have come to light and is proving popular with users. Thus even when users are offered the opportunity of changing the default so that items are displayed in date order they choose not to do so. Take the case of Emerald, where users can opt for a newest first or oldest first ranking; respectively, only 4.1% and 0.2% chose to do so.⁷ This, together with improved access to journal archives, which can be searched from work desk or home, provides a level playing field for ‘old’ documents and increases their visibility.

Thus information systems no longer load article selection/retrieval in favour of the new, each article now has a more even chance of being selected and selection is more likely to be determined by information need. We need therefore to look again at the usage of articles by their date of publication, employing the very latest methods for monitoring digital use, to determine what difference this has all made. Thus, it might be expected that, as Odlyzko (2000) has discovered, with improved access to journal archives back numbers might increasingly be used, and increased visibility and accessibility are part of the explanation.

Log usage data, gives us many more opportunities to investigate the age of material used. Thus in addition to the standard individual title and journal subject analyses, we can, as we demonstrate here, discover the differences in age distribution over day of the week and between different types of item displayed (e.g. abstract, full-text). Also, when related to subscriber data, as shown in this article, the data can tell us about differences in the age of material used by type of user, for instance, according to status (e.g. student, academic, and practitioner) or type of subscription (e.g. Big Deal and non-Deal subscriber).

By investigating the usage logs of two important and contrasting digital libraries—EmeraldInsight and Blackwell Synergy, the first a largely business and information studies database and the latter much more scientific in orientation (although still containing social science and humanities material) we hope to shed further light on the topic of decay and mark out areas for further research.

We believe that the research we have conducted is probably one of the largest date of publication studies conducted on usage data, covering as it does around 1 million users and 13 million user transactions. Our results are of special importance to those people charged with archiving digital copy, determining what value the archive has, to whom and what price should be charged for access to archival material. At present this important area is lacking critical information upon which policies can be developed. This has resulted in the emergence of a variety of practices regarding archival or older material, with some publishers selling it off to contractors, others providing access to current material for free and charging for access to archival material and others doing the very opposite.

2. Aims and objectives

The main aims were to:

• determine the age/publication date profile of material viewed in two popular digital scholarly journal libraries;

---
⁷ Analysis based on nearly 590,000 searches conducted in March 2003.
illustrate, by means of deep log analysis techniques developed by the authors, how a detailed analysis of the date distribution of documents used can be obtained, particularly in regard to establishing whether there are differences between:
- type of journal content viewed (table of contents, abstract, full-text article);
- individual journal titles;
- different subject fields;
- types of user, defined by occupation/status and type of subscription/access;

Secondary aims were to:
- determine whether the age distribution of items viewed was a function of article availability (i.e. whether the fall-off in usage is a reflection of the fewer number of older articles available on the system);
- compare our findings to previous work especially that conducted in a hard-copy environment and based upon citations.

It should be stressed that the paper is largely concerned with showing what a particular methodology—deep log analysis (relating usage logs to subscriber demographic data—see methods section for a full explanation) can disclose about the age of journal articles used by digital scholarly users. It is not our intention, nor is there the space, to extend our investigation into a consideration of what our data means in connection with the information needs of scholarly journal users. Thus, for instance, the question why readers want older material has to be addressed at a later stage in the research, employing qualitative methods.

The analyses were conducted on two digital libraries, Blackwell Synergy (covered about 700 scholarly journals at the time of investigation) and Emerald Insight (150 journals). However, we do concentrate on Synergy because of its greater coverage and the enhanced opportunities it offered for age analysis. The intention was not to compare the two but really to show the variety of decay analyses that could be undertaken using transactional log and subscriber data; and the two digital libraries offered different opportunities in this regard. Admittedly the analyses are somewhat limited by the relative recency of the archives of the two libraries (In the case of Synergy they went back about 7 years and for EmeraldInsight it was 8–10 years). However, the purpose of the research was to mainly show what can be done and what kinds of analysis we should pursue in future. Future work will see the analyses being conducted on longer datasets and new analyses being introduced for other types of user group—those defined by geographical location, how they arrived at the site, and whether they belonged to a research or teaching orientated university.

3. Literature review

The term “obsolescence” appeared for the first time in the work of Gross and Gross (1927). Since then hundreds of studies have investigated the issue (Egghe, 1994). Burton and Kebler (1960) introduced the term half-life to quantitatively describe the rate of obsolescence of the scientific literature. They defined it as “the time during which one-half of the currently active literature was published”. Brookes (1970) established some other significant obsolescence concepts including utility, utility factor and annual aging factor.

Citation studies have disclosed that papers in the hard science age more quickly than in the social sciences and as we move through the social sciences to the humanities aging slows down. The half-life data from Cunningham and Bocock (1995) is illustrative: Metallurgical engineering, 3.9; Genetics, 5.0; Physics, 5.2; Archaeology 9.5; Botany 10.0; Geology 11.8; Music education 15.0; Biblical criticism 21.6.

Several studies have confirmed that there is a correlation between literature growth and obsolescence. The faster a given journal grows (in terms of number of articles published per year) the more rapidly it ages.
(Tabah, 1999). Sangam (1999) concurred that the higher growth of a journal, the higher obsolescence and the higher half life. Wallace (1986) defined productivity as “the number of references to a particular journal in the given database” and by citation analysis of the desalination literature showed that those journals that were most productive would, on the average, have relatively short active lives and as journal productivity decreased, the average active lives of the articles contributed by journal would increase.

It has to be said that Gapen and Milner (1981) have poured cold water over many citation-based obsolescence studies and pointed out in their classic review article that all such studies had been of little value or use in library operations. They posed the question “who is to say that published research is the only use to which information can be validly put?” It is self-evident that authors do not cite all articles they read and not all readers are authors.

While citation data are the norm, circulation and re-shelving statistics have been used in a few obsolescence studies, particularly for the study of the aging of book collections (e.g. Burrell, 1985, 1986; Jager, 1994). A study by Sullivan, Vadeboncoeur, Shiotani, and Stangl (1980) is one of the few research studies on journal obsolescence which uses re-shelving data. The study tested the hypothesis that newer volumes of biomedical journals were more likely to be consulted than older volumes as they (the newer ones), contained more articles. The results showed that the hypothesis was not true. Another study that used journal usage data was that by Tsay (1998) who compared in-house use (re-shelving) data and citation data in a medical library and found a correlation between frequency of use and citation frequency.

Analyses of document delivery data again do not conform to the picture produced by citation data, or usage data for that matter. Thus an analysis conducted by the British Library Document Supply Centre (BLDSC) on requests made in 2003 showed that the most recent two years only accounted for only 27% of requests (Brown, 2004). Brown feels that this is because it represents derivative demand—the main demand being felt by publisher’s digital library site, with BLDSC taking the demand for the more esoteric or non-digital journals which the publishers cannot yet provide.

King and Tenopir (1999) in an extensive review article on using and reading scholarly literature demonstrate that the age distribution of articles has remained much the same over the past 30 years, as Table 1 demonstrates.

Odlyzko (2000) writing at the same time comes to a different conclusion. He puts forward a strong case for the slow but sure move to electronic formats in scholarly journal usage which he believes will lead to much wider usage of older materials, attributable to easy online access. Apparently, on average, an electronic article gets downloaded most frequently within one year of its submission, with the largest number of downloads to articles from the current month’s issue; in subsequent months accesses drop, but whereas for print journals usage continues to decrease with time, for electronic information it appears to increase. This says Odlyzko, “supports the thesis that easy online access leads to much wider usage of older materials”.

<table>
<thead>
<tr>
<th>Age (years since publication)</th>
<th>Case Institute of Technology 1960 (%)</th>
<th>Tenopir and King (2000) 1993–1998 (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>61.5</td>
<td>64.4</td>
</tr>
<tr>
<td>2</td>
<td>13.3</td>
<td>14.3</td>
</tr>
<tr>
<td>3</td>
<td>2.6</td>
<td>3.1</td>
</tr>
<tr>
<td>4–5</td>
<td>8.4</td>
<td>6.0</td>
</tr>
<tr>
<td>6–10</td>
<td>10.2</td>
<td>4.8</td>
</tr>
<tr>
<td>11–15</td>
<td>1.7</td>
<td>2.3</td>
</tr>
<tr>
<td>&gt;15</td>
<td>2.3</td>
<td>5.0</td>
</tr>
</tbody>
</table>
Tenopir and King (2000) however do acknowledge that, although they found that there were fewer readings of older materials, they were rated more importantly than new articles and more time was spent reading them: “When university scientists said they needed an article for research or teaching purposes, they were asked: ‘How important is this article to achieving your research or teaching objectives?’ They rated importance from 1 (not at all important) to 7 (absolutely essential). The average importance rating for research articles less than a year old was 4.78; for those articles over a year old the rating was 5.26. For teaching, the averages were 4.76 and 5.00 respectively. The average time spent reading articles less than one year old was 43 min per article; it was 61 min for articles more than one year old and slightly more than that for those over five years of age. Outside the universities, the average reading times were 35 min for articles equal to or less than one year old, 49 min for those older than one year, and 53 min for those over five years old. These findings were consistent with the fact that a higher proportion of older articles were consulted to prepare a formal publication, such as an article or a book. In other words, approximately 26 percent of new articles were reported to be read for that purpose compared with 42 percent of older articles” (Tenopir & King, 2000).

Guthrie (2000) provides some support for Odlyzko (2000) in a report on the use of JSTOR, a database that provides access to backfiles of journals from 15 diverse disciplines, that in their assessment of the age of useful articles, “… [in the] the field of economics as an example, a surprising number of older articles have emerged as the most heavily used. The average age of the articles in the top ten most printed and viewed articles in the economics cluster is 13 years. When put alongside the publication date, it was apparent that some older articles have truly lasting value. In most of our major fields, older articles were well represented among the ‘top ten.’” Although, he notes that caution should be taken on drawing conclusions from this data—due to limited numbers—he wonders “… if the number of citations to older articles in JSTOR will increase as the older articles become more conveniently accessible”.

Finally, Herman (2004a, 2004b), employing depth interviews, found that, when scholars talk of the importance accorded to older material in research work they seem to think of the cornerstone articles of a given field, the classics, rather than the run of the mill publications: quite a few of her interviewees stated categorically that they usually go back no more than three to five years in their literature searches, even in the humanities for if you rely on 30, 40, 50 year old information you are not considered ‘with it’, and you will have trouble getting your work published. Obviously, if they want the seminal research of the past, they have little reason to search for more than a few years old information, for the more central findings of any given field are cited in subsequent works.

4. Methods

All digital information platforms have a facility by which logs are generated that provides an automatic and real-time record of use by everyone who accesses information services on these platforms. They represent the digital information footprints of the users and by analysing them you can track and map their information seeking behaviour, and, when enhanced—something we call deep log analysis, they can tell us something about the kinds of people that use the services. The attraction of logs is that they provide abundant and fairly robust evidence of use, in the particular case of this paper use by year of publication. With log analysis it is possible to monitor the use of a system by millions of people, around the world. Logs record use by everyone who happens to engage with the system—there is no need to take a sample. Thus we provide evidence of year of publication use based upon millions of user transactions. The great advantages of the logs are not simply their size and reach, although the dividend here is indeed a rich and unparalleled one. Most importantly, they are a direct and immediately available record of what people have done: not what they say they might, or would, do; not what they were prompted to say, not what they thought they did.
4.1. Data collection and definitions

Our analyses are based mainly upon three sets of raw server transaction logs obtained from the Emer-
aldInsight and Blackwell Synergy journal libraries over different time periods. The Blackwell data is the
main focus of the paper, with Emerald data being used selectively. The datasets used were:

1. **Blackwell main**: Two month’s worth of logs for Blackwell Synergy (February–March 2003), in which
data was related to a subscriber database maintained by Blackwell. 10.5 million transaction were
monitored.

2. **Blackwell pilot**: In addition one day’s (17th September 2003) Blackwell Synergy logs were analysed—a
little over half a million user transactions in all, which constituted a test-bed for analyses conducted
above.

3. **Emerald main**: One month (June 2002) of Emerald’s digital library server logs was analysed. Nearly 2.5
million transactions were monitored.

Of course, the fact that the datasets covered different time-periods means that any comparisons between
the two publisher platforms have to be treated with caution, but it was not the intention to do this. The size
of the datasets were enormous, nevertheless we are only commenting on a month or two of data and our
results should be looked at in this light.

In all cases the raw logs were obtained and subject to standard deep log techniques, parsed and then
processed by SPSS. For full details of the methods see Nicholas, Huntington, Lievesley, and Wasti (2000).

The working definitions for the usage metric employed in the paper, items viewed/requests made was a
'complete' item returned by the server to the client in response to a user action. Typically this might include
an abstract, an article or a table of contents. A complete item might be all the pages, charts, etc. from an
article, and this is recorded as a single item and hence are quite different from traditional server log files that
record pictures and text documents separately. The Blackwell logs also recorded views to the home page
and a returned search screen.

In the cases of Synergy we embellished and supplemented the usage data with data about the user’s
occupation/academic status, we call this deep log analysis. User background data held on a registered
user database was related, via an identification number, to the usage logs generated in February 2002 by
registered users. The user database contained records of over 500,000 registered users. The database
was not a complete record of subscribers entering the site. This was because there was a number of
ways that subscribers entering the site were recognised, for example they maybe identified as coming via
a trusted proxy server user, a society member, a location such as a university, or were users at a given IP
address and so on. The number of subscribers entering the site via their user name and password was rel-
atively small—about 10%. The log file of these subscribers was extracted and supplemented with information
extracted from the form that users fill in this gives information on the users occupation and place of
work.

5. Results

5.1. Year of publication of journal viewed (Blackwell main study)

Fig. 1 gives the distribution of all items (ToCs, abstracts and full-text articles) used during January/Febr-
uary 2003 by year of publication. The dip in 2003 should be largely discounted because only two months
(January and February) of 2003’s journal issues were included in the analysis. Clearly use appears to fall off
sharply with publication age, and after 1997 there is very little use indeed.
Table 2 charts the decline rather more precisely by tabulating the percentage share for each year. As can
be seen well over half of all use (53.7%) was accounted for by items published within the last 15 months.
2002 accounted for the highest proportion of use—a third of all use in fact. 2003, although being only rep-
resented by two months worth of data still accounted for the second highest proportion of use (21%). No
year in the nineties accounted for more than 8% of use. The download half-life, being the median of the age
distribution,8 was 2002.

5.2. Allowing for a declining archive

Of course, the apparent downturn in usage with age could be a function of the database’s declining
archive. Thus Synergy has good coverage of journal issues back to about 1997, but before that coverage
is patchy, with a few older items stretching back to 1981. To put this in more specific terms, while 658 jour-
nals were covered by Synergy, about 7% had archives of just one year, a quarter (25.5%) had archives aged
of three years and half (52.2%) had archives of up to five years. Only 1 journal was available online for the
period 1981–2003 covered by Fig. 1. Fig. 1 then provides a misleading pattern of decline as usage patterns
will in part reflect the availability of material online. Hence comparatively greater downloads will be asso-
ciated with years where there is a greater availability of journals online, just because there are more journals
and articles available to search.

A relatively simple way to overcome or allow for this is to look at usage of a subset of journals that were
available over the same time frame. It was found that 206 journals were available throughout the six year
period 1998–2003. Table 3 shows the results, and breaks down use by item type.

---

8 We have adapted the concept of citation half life to downloads, which in the former’s case is the time it takes for a journal (or, in
this case, a set of journals representing a field) to attract 50% of all the citations it is ever likely to attract.
Interestingly, the net result is not that much different, although usage for 2003 is rather higher and 2001 and 2002 somewhat lower, so the results show a slightly greater concentration in the very new. The download half-life (median) is again 2002, the arithmetic mean 2001.3 while Huber’s robust estimator is 2001.6.

5.3. Year of publication of item (content) viewed (Blackwell main)

Analysis by content or type of item viewed is rather unusual but interesting, in that it tell us whether the age of the various types of content (table of contents, abstract and full-text articles in PDF or HTML) differs. There is in fact a variation by type of item viewed. PDF full text use is particularly weighted towards recent usage though a greater use of table of contents was made in earlier years.

Fig. 2 provides another way of looking at the type of content used by date of publication as it gives the percentage share of use by type of content (ToCs, abstracts, HTML and PDF articles) for each year during the period, 1996–2003. There are significant differences between years. As previously noted PDF views account for a much higher proportion of recent views—29% of 2003 views but only 22% of 1998 ones. The importance of table of contents for identifying or evaluating older material is also confirmed; with 37% of views of 1998 publications being of this type, while they only accounted for 32% of views to the most recent years; there was also marginally less use of HTML articles in the more recent years. The differences were small, but possibly significant. The fact that the increase in the use of table of contents with journal age was almost matched by a decline in the use of full-text articles, which fell slowly over the six year period from

Table 3
Percentage usage by date of publication for subset of journals with the same length of archive

<table>
<thead>
<tr>
<th>Year</th>
<th>1998</th>
<th>1999</th>
<th>2000</th>
<th>2001</th>
<th>2002</th>
<th>2003</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>7.3</td>
<td>9.7</td>
<td>12.1</td>
<td>15.8</td>
<td>31.4</td>
<td>23.8</td>
</tr>
<tr>
<td>ToC</td>
<td>9.0</td>
<td>10.0</td>
<td>11.4</td>
<td>15.8</td>
<td>28.7</td>
<td>25.1</td>
</tr>
<tr>
<td>Abstract</td>
<td>6.8</td>
<td>9.3</td>
<td>11.9</td>
<td>15.0</td>
<td>32.6</td>
<td>24.4</td>
</tr>
<tr>
<td>HTML</td>
<td>8.2</td>
<td>11.2</td>
<td>14.0</td>
<td>17.3</td>
<td>30.3</td>
<td>19.0</td>
</tr>
<tr>
<td>PDF</td>
<td>5.5</td>
<td>8.9</td>
<td>12.1</td>
<td>16.0</td>
<td>33.5</td>
<td>24.1</td>
</tr>
</tbody>
</table>

Year | Median | Mean | Robust
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>2002</td>
<td>2001.3</td>
<td>2001.6</td>
<td></td>
</tr>
<tr>
<td>2002</td>
<td>2001.2</td>
<td>2001.4</td>
<td></td>
</tr>
<tr>
<td>2001</td>
<td>2001.1</td>
<td>2001.2</td>
<td></td>
</tr>
</tbody>
</table>

Fig. 2. The distribution of use between ToC abstracts and articles by year of publication, 1998–2003.
40% (2003) to 37% (1996), indicates that users adopt a differential searching strategy of examining the full text of the newer issues and examining the table of contents of older issues.

5.4. Table of content use and article downloads

To explore this finding further a ratio was created by dividing the number of article views (PDF and HTML) by the number of table of content views. This provides an estimate of how frequently a table of content view converts to a download. Clearly there are other ways that users might access a full-text article, for example via the search facility or via an external site that might front the dataset. However, even allowing for this, the ratio is believed to be informative of table of content use. Values of less than one indicate a greater incidence of table of content conversions. Fig. 3 gives the average values of the ratio over the period 1989–2003. What is apparent is the steady decline in the ratio, apart from 1997 and 1998, over time. That is, comparing the relative use of article and table of content use over time, older journals were far more likely to have their table of contents scanned before they were downloaded. Plainly users were weighing things up first before they chose an older article.

5.5. Discipline differences

It was decided to take the table of contents analysis further by exploring the differences between journal subject coverage as determined by Blackwell subject categories, and the average ratio of each was calculated. The highest value for the survey period was recorded for medical journals (1.42) and argues that the use of table of contents in medical journals is more likely to result in an article download; the values for science (1.07) and professional journals (1.03) were lower and similar but the value for academic (social science and humanities) journals (0.27) indicates a poor conversion from a table of contents view to an article download. The conversion ratio for academic journals over time (1997–2003) never rose above 0.5. Academic journals were introduced relatively late on to Synergy (from 1997 onwards) and many of these were only available in a PDF format. Their introduction had a significant impact on the overall distribution and table of contents to article conversion ratio.

\[ F = 300.9, \text{ DF } = 3.2864, p = 0.000. \]
5.6. Abstract use and article downloads

We have demonstrated how an interesting relationship exists between table of content views and the propensity to download articles over time and thought it worthwhile to conduct a similar investigation concerning abstracts. Fig. 7 looks at the abstract to article conversion ratio over time. The values for all years were greater than 1 arguing that in all years there was a greater use of articles compared to abstracts. However, there is a pattern in the figure from 1994 onwards, discounting 2003, with the conversion rate from abstract view to download increasing year on year. This suggests that abstracts have a navigational or relevance-check value for older material. The absence of a pattern before 1994 needs further investigation but may result from the introduction of journals to the database during this period. The low figure recorded for 2003 may well reflect users keeping up to date by checking abstracts first before going on to view articles.

5.7. Download format preference

It was decided to examine the differences in the use of download format (HTML and PDF) over time to see what could be discovered. We only included in the analysis those journals available in either of the download formats. Fig. 5 shows the distribution in use between PDF and HTML over journal issue year. As can be seen there is a marked tendency for more recently published journal articles to be viewed in a PDF format. For example 72% of articles were viewed in a PDF format in 2003 compared to 53% of articles so viewed with a publication date of 1997. This might reflect a different user distribution. Thus previous research (Nicholas, Huntington, Watkinson, Jamali, & Hamid, in press) has found that students were more likely to view articles in an HTML format while Professors and teachers prefer viewing articles in a PDF format. Given this Fig. 5 would argue a marginally greater use of current articles by professors and teachers, while students might well make up more use of older published titles.

5.8. Year of publication of item viewed by subject (Blackwell main)

Further variation in usage over time is apparent when we compare journals by subject using the broad Blackwell subject categories (Table 4).
As might have been expected academic (social science and humanities) journals attracted the highest proportion of use to back numbers. Nearly a quarter (23.9%) of use is accounted for by publications issued before 2000, while scientific journals clearly account for the highest proportion of very recent usage—57.6% of usage is accounted for by journals published in the most recent 15 months. These differences and variations are further discussed below.

A way of allowing for the database’s declining archive is to weight the figures taking into account both use and the number of journals available in that period. This was done by dividing the ratio of year use (articles) to total use by the ratio of journal availability to total number of journals. We have done this for the four key Blackwell subjects. The results are given in Fig. 6, yearly figures above one show higher than expected use while those below one record lower than expected use. The results suggest that issues of professional and science journals older than five years were more likely to be used than medical and academic (social science and academic journals) journals. However, it has to be said that there were very few academic journals available before 1997 and this may impact on this finding.

Fig. 7 gives the breakdown share of item use for journal subject by year of publication. The relative use of academic (social science and humanities) publications increases with age. Thus they accounted for 7% of all use in 2000 but only 4% in 2003. This was also true of professional journals, they accounted for 14% of use in 2000 but only 11% of use in 2003. It certainly appears that academics and professionals make marginally greater use of back issues. The proportional use of medical journals was much the same over the period, while the share of science journals has increased from about 29% to 36%, suggesting a greater interest in more recent publications here, which is something we would probably expect.

Table 4
Percentage usage by date of publication for subset of journals with the same length of archive: subject differences

<table>
<thead>
<tr>
<th></th>
<th>1998</th>
<th>1999</th>
<th>2000</th>
<th>2001</th>
<th>2002</th>
<th>2003</th>
<th>Median</th>
<th>Mean</th>
<th>Robust</th>
</tr>
</thead>
<tbody>
<tr>
<td>Medical</td>
<td>6.0</td>
<td>9.2</td>
<td>12.3</td>
<td>16.6</td>
<td>32.0</td>
<td>23.9</td>
<td>2002</td>
<td>2001.3</td>
<td>2001.5</td>
</tr>
<tr>
<td>Academic</td>
<td>11.7</td>
<td>12.2</td>
<td>13.2</td>
<td>16.7</td>
<td>31.6</td>
<td>14.6</td>
<td>2001</td>
<td>2000.9</td>
<td>2001.1</td>
</tr>
<tr>
<td>Professional</td>
<td>8.8</td>
<td>11.3</td>
<td>13.0</td>
<td>18.8</td>
<td>30.2</td>
<td>17.9</td>
<td>2001</td>
<td>2001.0</td>
<td>2001.2</td>
</tr>
</tbody>
</table>

As might have been expected academic (social science and humanities) journals attracted the highest proportion of use to back numbers. Nearly a quarter (23.9%) of use is accounted for by publications issued before 2000, while scientific journals clearly account for the highest proportion of very recent usage—57.6% of usage is accounted for by journals published in the most recent 15 months. These differences and variations are further discussed below.

A way of allowing for the database’s declining archive is to weight the figures taking into account both use and the number of journals available in that period. This was done by dividing the ratio of year use (articles) to total use by the ratio of journal availability to total number of journals. We have done this for the four key Blackwell subjects. The results are given in Fig. 6, yearly figures above one show higher than expected use while those below one record lower than expected use. The results suggest that issues of professional and science journals older than five years were more likely to be used than medical and academic (social science and academic journals) journals. However, it has to be said that there were very few academic journals available before 1997 and this may impact on this finding.

Fig. 7 gives the breakdown share of item use for journal subject by year of publication. The relative use of academic (social science and humanities) publications increases with age. Thus they accounted for 7% of all use in 2000 but only 4% in 2003. This was also true of professional journals, they accounted for 14% of use in 2000 but only 11% of use in 2003. It certainly appears that academics and professionals make marginally greater use of back issues. The proportional use of medical journals was much the same over the period, while the share of science journals has increased from about 29% to 36%, suggesting a greater interest in more recent publications here, which is something we would probably expect.
5.9. Year of publication of item viewed by individual journal title

5.9.1. Emerald main study

Not only do we have differences between broad subject fields but we also have quite clear differences between individual journals belonging to the same discipline. To demonstrate this we turn to the EmeraldInsight. Tables 5 and 6, respectively, show age analyses for selected Library and Information Science (LIS) and Management/business journals—just those titles with relatively long archives (nine years) were included. There are enormous differences between subjects and amongst titles within a subject. This of course points to the dangers of generalising decay data. Table 5 shows the decay for the five LIS journals with the longest archives. Half of the article downloads from Interlending and Document Supply occurred within...
the first year, and this compares to a figure of 29% for Library Review. OCLC Systems and Services' value waned most rapidly, with 90% of downloads being to the most recent five years of the journal, whereas the figure was less than 75% for Library Review, perhaps reflecting the practical nature of the former.

Table 6 shows the decay rates for 11 management/business journals. In general use drops off less quickly than it does for library titles. Once again there are some big differences between journals. Thus, taking downloads for the last five years the International Journal of Operations Production Management Hospitality decays much more slowly than the average (56% of downloads only were made to issues up to five years old) and Journal of Services Marketing more rapidly (comparable figure, 81%). Put another way, nearly a quarter of all downloads to Journal of Operations Production Management Hospitality concerned volumes six to nine years old. Management Decision was a journal of the week during data collecting time—meaning downloads were free during this period, and this would account for high representation of hits from the first year.

We were especially interested in the decay rates of the top performing journals according to article use and Figs. 8–10 show the distribution of articles use by year for the top three performing management journals. The three journals are *Industrial & Commercial Training* (online from 1994), *European Journal of Marketing* (1993) and *Integrated Manufacturing Systems* (1994). Even the briefest glance at the histograms shows that we have three very different age profiles. *Industrial & Commercial Training* probably has the 'classic' continuous, decaying profile, although the decline is gentle rather than rapid. In the case of the *European Journal of Marketing*, while there is a declining trend it is bumpy and patchy; indeed, sometimes older volumes attract more downloads than younger ones (1996 and 1997 are cases in point). The *Integrated Management Systems* distribution is something of a hybrid with a short and sharp decline in use followed by quite a long stabilisation (the years between 1999 and 1995 show hardly any difference in use).
5.9.2. **Blackwell main study**

Figs. 11–14 show the age distribution of article downloads (in percentages) for a sample journal selected from each of the Blackwell Synergy subject groupings listed in Fig. 4. As in the case with Emerald, there are clearly three very different distributions on view. The Journal of Neurochemistry bears more resemblance to the age profile to Emerald’s Integrated Manufacturing Systems than any of the Blackwell titles. Two thirds of all use occurred in the most recent three years, and then use dropped regularly but slowly. Cephalalgia’s profile was marked by the very heavy downloads to one year (2002), which accounts for nearly one third of all use. Journal of Advanced nursing has a very flat distribution, with year usage totals never exceeding 19% and use holding up well back as far as 1998. The performance of the Scandinavian Journal of Psychology was notable as 50% of its use occurred in just one year, 2002.
5.10. Year of publication of type of item viewed by status of user (Blackwell pilot)

The following three Figs. 15–17 examine the age of the item viewed by occupation of the user who viewed it, first for ToCs, then abstracts and finally full text articles.

In regard to table of contents views (Fig. 15), professionals and practitioners tended to consult the more recent ones and researchers the older ones—something we would probably expect given the need for researchers to conduct literature reviews. Professionals and practitioners accounted for nearly a third of ToC views for 2003 (31%). For journals older than 2001 researchers were the dominant group, about 40% of views to journals published before 2001 were accounted for by these subscribers.

A slightly different pattern emerges with views to abstracts (Fig. 16). In terms of the most recent publications (2003) researchers made up the modal group, 29% of researchers accessed a 2003 abstract. And they were also the most significant group for viewing abstracts of journals published before 2000. Undergraduates’ use of abstracts appeared to increase with the age of the journal. Perhaps there is a lag between
publication and the introduction of the accompanying ideas into the classroom. Postgraduates followed a similar pattern to undergraduates.

In terms of articles viewed (Fig. 17) it is the practitioners/professionals who accounted for the highest proportion of full-text views for the most recent year and the researchers for the highest proportion of views to pre-2000 material.

5.11. Year of publication of journal article by type of user according to type of subscription held (Emerald main)

For Emerald it was possible to categorise users according to their access rights. ‘Big Deal’ users had most rights, having access to the full-text of over a hundred journals. Non-Deal users generally had access to typically 6–12 specific titles. For non-subscribers (limited) full-text access rights could be obtained through
three routes: (1) the ‘journal of the week’ promotion, when two titles from Emerald’s list were featured; (2) by becoming a trialist for a month which gave access to five full-text articles from Emerald’s list; (3) by paying for it by credit card.

To illustrate the form of analysis Fig. 18 provides the date distribution of abstract, article and table of content views for the European Journal of Marketing, broken down by type of subscriber. Interestingly, the breakdown in use of back issues between subscriber types is proportionally the same and there is little evidence that Big Deal subscribers, those with the greatest freedom to exploit the archive, used back issues more than non-Deal users or non-subscribers.

6. Conclusions

Tenopir and King (2000) after a comprehensive view of the literature came to the view that nearly two-thirds of all use occurred in the article’s first year of publication. This was largely based on hard-copy
studies. Our own figures—about 55% for 15 months based on a subset of Blackwell journals which had the same length of archive, showed there was, as expected, somewhat less concentration on the new. We acknowledge that we were working with relatively short time frames—rarely more than 10 years. Of course, our figure is not strictly comparable to the Tenopir and King figure. Tenopir and King were looking at the historic decline in use of the same article while here we are looking at the use in the current period of all articles published in current and previous periods. However, both studies would arguably estimate the same historic event under the assumption that aggregation over a large set of articles allows us to refer to the result as being representative of the average or the “standard” article albeit here a Blackwell standard article. Assuming this then, the fact we are looking at the usage of a group of articles of various ages at the same point in time is much the same as looking at the same article from one year to the next over a given number of years. In fact this assumption must be the case if Tenopir and King claim their rule generally holds for all articles. Our approach potentially provides a less biased and more accurate picture as it is based on many articles rather than a few.

Fig. 16. Year of publication of item viewed (Abstracts) by type of subscriber—pilot study.

Fig. 17. Year of publication of item viewed (articles) by user status.
While we did find that use dropped off less than might have been expected from our review of the literature we found little evidence to support Odlyzko (2000) observation that use actually increased with age. This might be because the Odlyzko data is some four years old now and matters have moved on.

While these findings are of interest our prime aim was to demonstrate what deep log analysis could tell us about the use by date of publication. Copious log data meant we could investigate far more deeply and robustly than conventional self-report or document delivery studies. Thus we examined date of item used in regard to:

- **Types of content/item viewed**—abstract, table of contents, full text article. There were some differences here. Chiefly and tellingly, there was some evidence to suggest that contents pages had greater longevity and users frequently consulted these before downloading the full-text of older articles.
- **Subject of journal.** Surprisingly perhaps, subject differences were not as great as one might have supposed from a reading of the literature, although, as might have been expected the back issues of social science and humanities journals were made more use of. It appears digital library searching is reducing the differences between subjects—a kind of digital publication date default could be emerging.
- **Individual journals.** The differences here were much more significant. In general the journals we featured in our analysis had very different usage age profiles, and again this seemed to have little to do with subject, and more to do with function or purpose. Also enhanced digital visibility, through offers, such as Emerald’s ‘Journal of the Week, might be more influential in explaining differences in use. Much more research is needed here.
- **Types of user, categorised by occupation and type of subscriber.** This is a form of analysis generally unique to deep log analysis. In regard to table of contents, professionals and practitioners consulted the more recent ones and researchers the older ones; in the case of abstracts what stood out was that undergraduates’ use of abstracts increased with the age of the journal and that researchers made proportionally the greatest use of abstracts of the most recent journals; and in terms of articles viewed it was the practitioners/professionals who accounted for the highest proportion of full-text views for the most recent year and researchers the highest proportion of older articles (pre-1999). Interestingly, the breakdown in use of back issues between subscriber types was proportionally the same and there was little evidence that Big Deal subscribers, those with the greatest opportunity to exploit the archive, used back issues more than non-Deal users or non-subscribers.

![Fig. 18. Abstract, article and table of content views for the European Journal of Marketing broken down by date of publication and type of subscriber as a percentage bar chart.](image-url)
As people, because of reasons of time pressures, easy access and digital visibility increasingly confine their searches to the digital environment, ironically, it could be that their information horizons, especially in regard to date of material used will be expanded and shrunk at the same time. They might get access to more journals but it is possible that, unless publishers recognise the value of the archive to the user and how it is used, they could see the age of the journal archive reduced somewhat. If this proves to be the case and it becomes a case of library catalogues, bibliographic references and a consequent wait for older material and on-demand, full-text, visual and interactive for newer material, then there is a real risk of our best older texts disappearing from view. Undoubtedly this will hit the humanities and social sciences the hardest, with another digital divide possibly opening up. There is a need for retrospective conversion in these disciplines, more for searching than reading, because most people prefer to read in printed form. We need to be able to search old full-text documents if we are to fully exploit them. Without online representation, few people will find out about old, hard-copy documents. A fixed form of representation and stable form of preservation does not guarantee real access to printed documents. The printed world and online world are much more connected than we think. Printed material will become obsolete much faster without online representation. The printed world is also much more dependent on online visibility and retrieval than before and this dependency will grow and grow in the future.

Even a decade ago, Line (1993) could see this coming. He envisaged that the emergence of virtual libraries and online journals, and that the improvement of browsing and searching facilities might affect obsolescence because in an electronic environment all volumes of the journals had the same opportunity to be retrieved and used—the level-playing field we referred to earlier, and the archives of electronic journals usually do not go back more than say 10 or 15 years ago and this may lead to neglecting the old materials which are not available in electronic format. His words could prove to be particularly prophetic.

While it does appear that easy access to large digital libraries like Blackwell Synergy and EmeraldInsight does lead to a greater exploitation of journal archives the visibility/accessibility of older material can be enhanced in other ways, and price comes most obviously to mind. For pay as you go users, a small but growing group of consumers who are potentially enormous in number—for instance they actually constitute Emerald’s biggest group of users, it does not make sense to price all articles at the same price—very little else in the consumer world is priced like this, and price is clearly a barrier to use and an important retrieval characteristic. Clearly, there is considerable scope to price older, less used items, below their more used cousins. If the consumer or library only has a pound to spare then they might well be happy content with something that shows a little age, but the richer consumer perhaps not. Currently publishers have grown fat on Big Deals, which sew users up beforehand so there is little incentive to consider incentives (probably the opposite) but with a possible move to a pay-by-usage model then they might consider measures to make older material more attractive of accessible.

References


