

Digital Dilemmas: electronic challenges for the scientific journal publisher¹

Michael A Mabe

*Elsevier Science Ltd, Langford Lane, Kidlington, Oxford OX5 1GB
m.mabe@elsevier.co.uk*

Excited by what has now become possible, participants in the creation of the new digital library often overlook fundamental (often implicit) aspects of the paper paradigm that still need to be maintained in the digital universe. In the realm of scholarly and scientific communication, the issues of first publication, version control, referencing, stable and immutable archives, as well as strategic issues blurring the distinctions between the information players all have to be reconsidered. These matters are reviewed with an indication of what has and has yet to be achieved.

We are in the midst of the third great cultural revolution of humankind, the move to a digital environment for all information. Living through a revolution is never pleasant; things can change in unpredictable and catastrophic ways. For those that work in the information industries, never has the traditional Chinese curse 'May you live in interesting times!' been more appropriate.

Much has been written about the advantages of digital libraries. Less common has been an examination of what such a change would mean for particular segments of the information world. This paper is an attempt to review the broad issues that surround the creation of digital libraries of learned or scholarly journals.

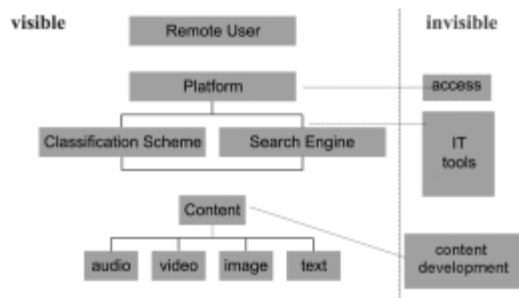
A schematic structure for a generic digital library is illustrated in Figure 1. This structure shows the user-visible elements as well as the user-invisible modules that are required. The digital library vision is for a single, continuously available, interoperable platform. Work on the generic expectations of its potential users (readers) suggest they want (*inter alia*):

- a one-stop-shop platform that is interoperable, accessible remotely for 24 hours a day and 7 days per week;
- information that is interconnected, for example by reference linking;

¹ Revised version of a paper originally presented at the DELOS Workshop, Pisa, November 1999.

- a system that allows easy searching and browsing, possessing downloadable and printable files;
 - multimedia applications;
 - an application that is free at the point of use.
- Trying to achieve these stringent demands in the context of the other players in the information chain is the challenge facing information professionals. The dilemmas that arise in trying to achieve this for the learned journal marketplace form the basis of this paper.

Figure 1. Schematic generalised digital library

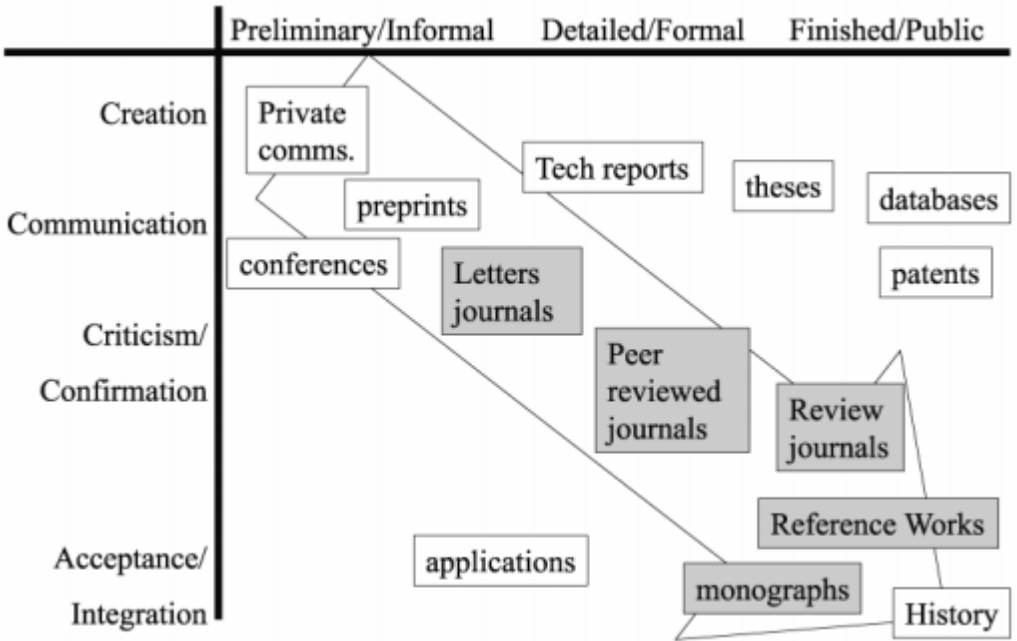


STM journals: a special case

What are they and who uses them?

Scientific, technical and medical (STM) journals are quite distinct from other types of serial publications, such as magazines and newspapers. They have often been described as the 'minutes of science' [1], a metaphor which reflects their principal formal role. They

Figure 2. Science communication and the rhetorical flow



are not 'scientific newspapers'², the news function in science has usually occurred through informal communication within each scholarly community via conferences, meetings, telephone conversations and (especially in physics) via the circulation of draft versions of papers before publication, called pre-prints. STM journals are principally formal records of research that has been methodologically checked by the research community, through the process of peer review. They are also the main instruments in the rhetorical process of science by which the observations of individual scientists are given weight, turning their observations into objective and universal results, which become the building blocks of accepted theories. The rhetorical position of

the journal, in the context of other communication vehicles, is shown in Figure 2. STM journals are mainly used by researchers in academia and industry. They use them as authors and as readers and, for most research journals, the set of authors and the set of readers strongly overlaps.

What do they do? [2, 3, 4]

The primary function of the research journal is registration. That is, it registers authors' research claims by providing an independent mechanism that connects an observation to an author and to a date of acceptance. In so doing they establish priority of the published author over other competing unpublished authors. Authors are also seeking to place their claims in the best possible journal: (a) to give the highest level of validation to the research, and (b) thereby maximally rewarding the authors' research (and the author). Thirdly, they communicate the results to the readership of the journal. The title of the journal acts as a branding mechanism, identifying the content, approach and likely quality of its papers. In doing so, it selects a tightly defined market of readers from a larger potential pool

² All research level serial publications are loosely called 'journals' even though some of them are clearly intended to be specialist magazines for disciplines or professions (e.g. *The Lancet* or *British Medical Journal*) as well as publishing primary research. Such specialist magazines or professional journals have very large readerships, most of which are neither potential authors nor researchers. This makes for a very different type of information vehicle, one driven by reader needs rather than author needs, which affects both its editorial policy and business model.

and disseminates results directly to them. Lastly, as a formal record of validated research it preserves (archives) the investigation and its results for the scientific record. It is immediately apparent that the journal is principally set up to serve the needs of authors and that these may not represent the same desires of those researchers when they are in reader mode. It is from this contradiction, heightened by the move to digital delivery, that a series of dilemmas flow for the scientific journal publisher.

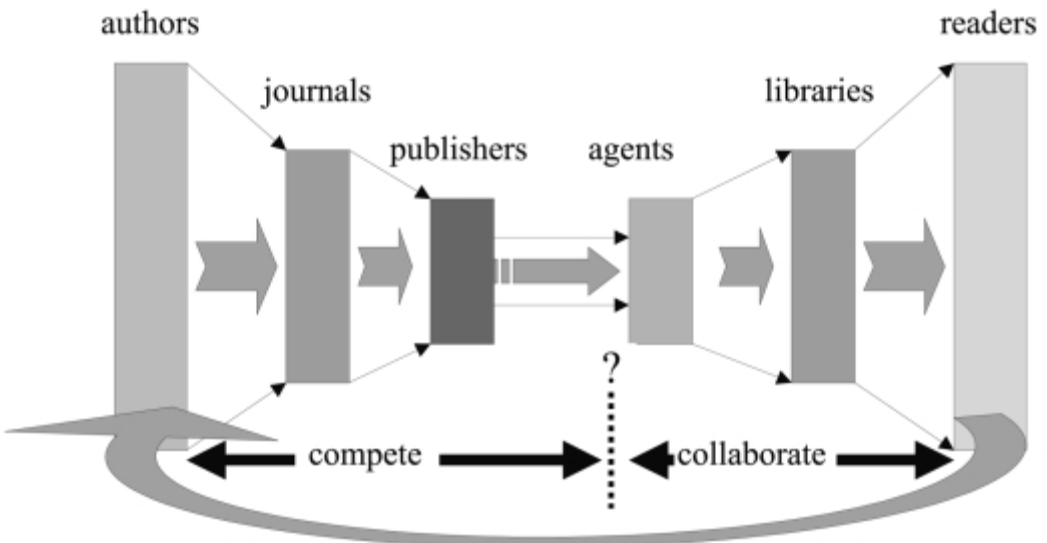
The strategic dilemma

The essence of the strategic dilemma is represented in Figure 3. Authors compete with each other to get their research accepted by the right (*one* particular) journal. Publishers compete to provide the right journal for top authors. Most journals are sold to libraries via subscription agents. These agents compete with each other to supply the widest range of journals, and partly collaborate, partly compete with the publishers. Libraries, on the other hand, collaborate with each other to satisfy readers in their desire to be able to access *all* important journals. For most research journals, the authors and the readers are nearly identical, but their behaviour

in author mode and reader mode is radically different. In the former, they focus on a few key journals, in the latter on a wide range of relevant titles. This poses a number of major questions for the players in the system: Where does competition stop and collaboration begin? Whom does one satisfy first? How can this conflict be overcome?

For the publishers, satisfaction of authors precedes but does not negate satisfying readers, not least because without any authors there cannot be a publication. Each publisher will strive to create journals that attract as many of the best authors as possible. At the same time, to satisfy readers' requirements, each publisher will try to provide a range of titles covering the areas of interest of each reader community. The introduction of a digital universe has heightened the dilemmas. The growth in electronic platforms provided by publishers for their own and sometimes other publishers' journals has begun to blur the distinctions between the competitive role of publishers and the collaborative role of the agents. The publishers seem to be expanding their role to encompass that of the traditional agent, while the agent is starting to take on some of the functions of the publisher. The

Figure 3. *The strategic dilemma*



increasing importance of the overall collection and the platform upon which it is made available has made some observers (wrongly) believe that the individual journal no longer has any significance.

For the librarian the broad needs of the reader are paramount. They want to acquire the necessary titles, but they do not want to have to deal with a multitude of publishers. In the paper world they have dealt with a small number of subscription agents, but in the electronic world they (and their readers) would prefer one source and one platform to learn how to use. As customers, they want one source, but they do not want one supplier! This has led some to collaborate over resource sharing and attempting to set up their own platforms. It has also led to the development of the agent's role. The agents increasingly see themselves as the honest broker, creating multi-publisher journal aggregations for the libraries.

The content dilemmas

What readers want versus what authors provide

A similar set of contradictions arises in considering the electronic expectations of readers in the light of what these same individuals are prepared to supply as authors. For paper publication, most authors rarely supply more than minimal flat text and diagrams, often not following the notes for authors and frequently of sub-standard reproduction quality. This problem is amplified when material is submitted electronically, with conversion between different operating systems and programmes causing chaos with special characters and symbols. Circumventing these difficulties has always been the major production role of the publisher in creating quality finished content from that which was submitted. In addition, the task of enhancing the text with links, both internal to the article and external, providing computer-aided indexing and classifications, are also left to the publisher.

When we move from simple electronic versions of flat text and diagrams to multimedia presentation, the problems become more acute. In the first place, it is still rare for authors to submit any multimedia [5]. Those that do

usually have some multimedia creation as part of their normal research workflow, such as surgical researchers videotaping operations or theoreticians or modellers using computer simulations or animations. The amount of effort even these authors are prepared to expend on multimedia efforts is limited, however. For most authors, it is the 'fact' of being published that is important, not the 'act'. Rewards are gained from publishing good articles in the best journals, not for preparation of sophisticated reader aids.

If readers say they want any of the above electronic enhancements, but authors do not supply them, what should be done? Research into the wants of electronic readers are bedevilled with the problems of users saying they want something, but not actually using it in practice [5]. Even given that they really do want to see multimedia, is it feasible for anyone other than the author to make up the difference? Can the cost of doing this be recovered by charging for extra value? The experience of the first CD-ROM book producers suggests not³ [6]. The general public was unwilling to pay much more for extra functionality over the basic content price of the paper product. Where this leaves the STM journal is unclear, but the old mantra 'content is king' still seems a wise one.

The archive and the back catalogue

The move to digital unpacks a number of previously inseparable items. The traditional journal issue delivered the 'browser', the content and the archive mechanism in a single entity. Electronic libraries are faced with providing their own hardware and browser, and purchasing access to content. Purchasing hardware is an additional, cyclical cost since technology and software developments necessitate roughly triennial upgrading. Questions over who performs the archival function and how it

³ The UK publisher Dorling Kindersley clearly thought that the added-value of the CD-ROM versions of its mass market reference works and guidebooks could be reflected in a premium price over the basic content. The failure of CD-ROM sales to match expectations together with other problems led last year to DK's acquisition by Pearson. (See reference[6].)

should be accomplished remain unanswered. It has not been a core competence of publishers, but it may become so in the future. Unlike a paper archive, the electronic one will almost certainly have to be regularly migrated from old to newly developed storage technologies with increasing regularity. The experience of the recording industry shows how over barely a century the archive has had to move from wax cylinders, to 78s, to mono LP, to stereo, to CD, to DAT, MiniDisc, DVD and now MP3. We cannot expect the future of journal archiving in the digital era to be very different. In addition, every migration will have consequences for the rich interlinking envisaged by most providers. The links will need refreshing and checking even when the technology does not change.

Demands are already arising concerning the digitization of the backlist. How feasible is this, given that there are an estimated 29 million articles in existence since the commencement of journal publishing in 1665, and that this total is doubling every 20 years? Digitization of the backlist and maintaining the archive are tasks with significant cost implications. Whoever has the archival task, how will they charge for it? Who will pay for it?

The meta-content dilemmas

In addition to these content-related issues, there are a number of information management aspects unique to the scholarly journal that are affected by the move to the digital domain. What constitutes first publication? Which version of a digital publication is the definitive version for the archive? How can readers' desires for interconnection be delivered? How will citations be managed in a non-paper environment?

First publication and definitive copy

Priority protection is a key deliverable for authors using the journal system. The drive to improve speed so that authors can win priority races has led many journals to publish their electronic versions up to one or two months in advance of print. Such speed is often at the expense of accuracy. Errors that occur in the first electronic version are often corrected in the print version. Equally, additional (non-print-

table) elements may be present in the electronic version but not in the print edition. For the purposes of the 'minutes of science', it is extremely important to establish which really is the final, archival edition. At present, there is no agreement on this, but current practice tends to favour paper. It is unclear how long this situation can (or should) continue.

The waters are muddied further in those disciplines where preprints (early draft versions of articles informally circulated for comment within the research community) have been common. The circulation of electronic versions of preprints, which later find their way into the formal journal literature, and their collection into web-accessible databases has begun to blur the traditional definition of first publication: first appearance in public. Is the date of publication the date of submission of the preprint to the preprint database? Or is it the date when it was received as a 'manuscript' at the journal editorial office? To what extent are the preprint and the final paper the same work? The scholarly community itself has no universal answer to these problems. There is, however, some agreement in some quarters.

First public appearance (whether in paper or online) of a peer reviewed paper is becoming accepted as the point at which first formal publication has occurred. For many, only the first appearance of an accepted, peer-reviewed paper counts. This has recently become the policy of Academic Press and John Wiley and Sons among others. However, there is no consensus about whether appearance as a preprint (electronic or paper) can be used to claim priority. Some journal editors view the preprint culture as pernicious and dangerous (especially in the clinical sciences). Others are concerned that the clear boundaries between formal and informal communication are being blurred. After all, in those disciplines where preprint circulation is common, the principal role of the preprint is to advertise the author's forthcoming paper and to circulate it for informal comment. Informal responses were encouraged, not least to improve the quality of the draft paper before its eventual formal submission to a journal. In this respect, the preprint was functioning in a similar way to a conference presentation. Both form part of

the grey literature, and it is legitimate to question why one would want to archive such materials when their informal function was past. The fear of many academics is that the spread of an archived preprint culture will create yet another layer to read when there is much too much already.

Pointers

The desire of readers to have the ability to move seamlessly from article to article via their citation links generates further complications. Universal reference linking requires unique identifiers for each article, in the same way that the ISBN uniquely identifies each published book. A first pass at this was the creation of the Publisher Item Identifier (PII) and related reference numbering, such as SICI. The PII, which is widely used by a number of large journal publishers, is formed from a number of elements: a prefix S (for 'serial', allowing also 'B' for 'book'); the International Standard Serial Number (ISSN) of the journal; a unique item number for the article in question; a computer generated check digit. The main flaw with this system is its adherence to the notion of paper as the primary and sole catalogue-able source. How is the same article in CD-ROM, online, etc. manifestations to be differentiated? The answer lies in the growing acceptance of a single protocol for numbering every potential instance of a work, not just in traditional publishing, but software, audio and video too: the DOI or digital object identifier [7].

The DOI avoids the problems of other numbering systems by disentangling the work itself from its potentially multiple manifestations. Thus, a Beethoven symphony can be given a unique number, representing the original creation of the composer, which is differentiated by further numbers indicating whether it is appearing as a musical score, a recording, a video etc. Not only is the system media-independent but it also provides a ready means to track rights-holders.

References and citation

Even if the DOI or some other system is used, the question remains as to whether authors will use these 'random numbers' in place of

the traditional and (relatively) transparent citation system of author/date/journal/volume/issue/page. As with the current experience of authors and multimedia, it seems unlikely that authors will abandon traditional citations. They will require someone else (probably the publisher) to convert the alphanumeric citation string to the DOI (or whatever) that allows pointing and reference in cyberspace. This is an additional task not present in the paper world: it presupposes that there will be a universal numbering system; and that its owners (such as the DOI Foundation) have set up a global database for the purpose. It will clearly cause problems for citation counters, such as ISI, who are currently faced with several streams of citations for each medium manifestation of the article. The data capture and processing of multiple streams of citation data for every single article appear daunting, to say the least.

STM community response

The response of the community of STM publishers has been mixed. Both the publishers and the subscription agents, responding to the strategic issues of the digital transition, have moved to create aggregations of journals and related services: the frequently touted but rarely delivered one-stop-shop. The essence of each of these platforms is a set of journals covering many subject areas with browsing, searching and printing facilities. Some of the publishers and subscription agents have developed platforms with multiple publisher material, cross-linking of articles and entry points to wider, secondary service abstracting and indexing layers.

Elsevier's ScienceDirect™ is a good example of this approach [8]. This platform contains a significant and growing proportion of the world's scholarly, scientific, technical and medical literature. Many other publishers have mounted their journals on the platform, and other parties are in negotiation to join. ScienceDirect™ provides its users with linkage between references and other articles in its database of full-text articles (presented in both html and pdf formats) and extensive abstracting and indexing layers, found in no other platform. These in their turn allow

searching and navigation to the wider universe of non-ScienceDirect™ material. Publishers Academic, Wiley and Springer have created platforms for their journals, but these lack the rich layer of secondary services found on ScienceDirect™.

While the archiving issue remains unresolved from the technical perspective, there is a growing realisation that academic authors and readers will need reassurance about the permanence of the systems in which their articles reside. Elsevier recently announced [9] that it would commit to maintaining an archive of all its published materials for as long as it remained in existence. In the event of it being unable to do so, it would hand over its archive to competent authorities to ensure its continued existence.

At the 1999 Frankfurt Bookfair a number of the major STM players announced [10] a reciprocal linking policy that would allow readers to navigate links between their respective electronic holdings. Called the CrossRef project, this initiative now has most of the key journal publishers involved and looks to be moving towards satisfying one of the great dissatisfactions of users of electronic journals. It will not, of course, cure the problem of the lack of backfile material in the collective database, which barely goes back to 1996. Although this issue will largely cure itself as time moves on (assuming that 1996 material remains available), for many disciplines the size of backfile desired is comparatively large. In mathematics, for example, references decades or centuries old are not uncommon. In chemistry, several decades of material would be required. According to recent studies, on average, 15% of readings are to material over six years old, 6% to material over ten years old [11].

In tandem with the CrossRef project support for DOI as a unique identifier system is growing, and this at least shows signs of facilitating some of the functionality readers say they want.

Outstanding issues

Although tremendous advances have been made in the last few years (and things have been changing so fast that we often forget that the World Wide Web has only just celebrated its

tenth birthday), many issues remain unresolved. The extent to which the current technologies have been embraced varies widely. Adoption rates differ both by discipline and by geographic region. Some observers are concerned that, although most of the scholarly community worldwide has access in some form or other to email and (less commonly) the Internet, the distance between the leading edge software and hardware and the most backward implementations is growing.

It is still unclear how digital archiving should be implemented and, more crucially, who should do it. As we saw above, archiving of richly interlinked material is not a static business, it requires constant intervention. We also saw that the desires of readers for extension of the digital backlist are not without considerable problems as to the extent to which this needs to be done. In both cases, a feasible business model is lacking which would enable the archival function to be performed and render the digitization of the backfile economic.

Perhaps the most thorny of the unresolved issues lies with what counts as first publication. Although this falls outside of the control of digital librarians and publishers into the domain of the academic world, the choice of one convention over another could have quite alarming effects for the other players in the information chain.

This paper began with a Chinese curse, so it is appropriate it should end in a similar vein. Looking over what has been achieved and what remains to be done we can at least take comfort in Chairman Mao Zedong's aphorism: 'Every journey begins with the first step'. We have only just taken that first step.

References

1. Velterop, J. Keeping the minutes of science. In: Collier, M. and Arnold, K., eds. *Electronic library and visual information 2*. Proceedings of the 2nd International ELVIRA Conference, De Montfort University, 2-4 May 1995. London: ASLIB, 1995.
2. Ziman, J.M. *Public knowledge: the social dimension of science*. Cambridge: CUP, 1968.

3. Ravetz, J.R. *Scientific knowledge and its social problems*. Harmondsworth: Penguin, 1973.
4. Meadows, A.J., ed. *Development of science publishing in europe*. Amsterdam: Elsevier, 1980.
5. Pullinger, D. and Baldwin, C. *SuperJournal Project. Final report*. (forthcoming 2001) Also available at: <http://www.superjournal.ac.uk/sj/>
6. *The Independent*, The Literator column, 23 December 2000.
7. Paskin, N. E-citations: actionable identifiers and scholarly referencing. *Learned Publishing*, 13, 159-166, 2000. Also available at: <http://www.doi.org>
8. Science Direct. <http://www.ScienceDirect.com/>
9. *Elsevier Science Information*, 15, 1999, 1-2. Also available at: <http://www.elsevier.nl/locate/esi>
10. *Information World Review*, January 2000, 1. Also available at: <http://www.crossref.org/>
11. Tenopir, C. and King, D.W. *Towards Electronic Journals*, SLA, 2000, 189.