Journal Article Identifiers

Gillian Page, Pageant Publishing

Introduction

The idea of giving each item published in every journal a unique identifying code has been around for some time. It has mostly been seen in the context of interlibrary loan, but it is clear that, if such codes existed, they could be used for many other purposes.

A good deal of work has been done in the USA by the Serials Industry Systems Advisory Committee (SISAC) on the Serial Issue Identifiers (SIID) and Serial Article Identifiers (SAID), but little information has spread about their proposals outside the USA. At the same time, the sponsors of the Adonis scheme (which provides copies of articles for interlibrary loan purposes) have developed a numbering system for their own purposes. There is also the European system, Biblid, with which some readers may be familiar.

Article identifiers are likely to affect almost everyone professionally involved with serials: librarians, document delivery centres, producers of software for library systems, publishers, copyright licensing centres, indexing and abstracting services, online vendors of serials, and possibly even journal editors, contributors and readers. It is therefore desirable that we should all be informed about the various porposals and have the opportunity to put forward our views.

This article arose from a couple of meetings in early October, one at the British Library Document Supply Centre and the other in association with the Frankfurt Book Fair, at which presentations were made of the SISAC and Adonis proposals. It is written by a publisher and so may show a lamentable lack of understanding of what library users, librarians, document delivery centres, copyright licensing centres and so on do, but perhaps it will provoke them into commenting on the points made here. I hope to hear from you.

The uses of article identifiers

Current interest in codes for journal articles has mainly been stimulated by those interested in developing library systems software for interlibrary loan. However, there are many other possible uses to which codes could be put; the suggestions given here are not necessarily comprehensive.

If the code could be allocated at an early enough stage, it could be used to track the article through the editorial and production processes. It could help editors who like to check the references supplied by authors. It could make the identification and recording of what has been copied and the payment of royalties or photocopying licensing fees to copyright owners much easier. It could be used to call up the article from an online service or (as in Adonis) from a CD-ROM or other storage medium.

The possible uses do not stop there. For instance, if statistics were available showing the number of copies made of particular papers, it could help publishers in promoting journals and editors in selecting papers, for they would know what was in greatest demand. There could be publications that reprinted the most commonly requested papers in a given field. The statistics might be used in assessing the publications of people seeking tenure, promotion, research grants or fellowships. It has been suggested that subject codes could be incorporated, and then the identifier could be used to advise interested parties of papers in advance of publication, or to gather other statistical information.

These further uses are not without hazard. The Copyright Clearance Center has found that some businesses are sensitive to others knowing exactly what they are copying. A paper might be in demand because it is so open to attack; because it sets out the current problems so well without making any

contribution to their solution, or because it offers an easier method of identifying a substance. That a paper is frequently copied or cited is not necessarily a signal that it is a great and original contribution to the subject, just as popular paperbacks do not always reflect the highest achievements of literature. Both scientists and librarians may be uneasy about too much advance information on unpublished papers.

Current proposals

At least three schemes have been proposed for article identifiers. The US-based Serials Industry Systems Advisory Committee (SISAC) has put forward an extension of the Serials Issue Identifier (SIID) to give a Serials Article Identifier (SAID) and this has been made consistent with the European-based Biblid (Bibliographic Identification). Its major defect is unwieldiness. In a fixed pitch format, it commonly consists of 56 digits: enough to defeat a computer, not to mention a human being.

The third system is the Adonis number. This has the advantages of comparative brevity and simplicity, consisting of the ISSN, the last two digits of the year of publication, and a five digit number unique to the article. It runs more or less sequentially throughout a volume, and is not unlike what many publishers are printing as Copyright Clearance Center codes on the first page of each article. The Adonis number is page-independent, which allows it to be allocated at an early stage (provided the year of publication is known). But because it is page-independent, it may be impossible to generate consistent codes independently. A library could not take a volume of a journal which does not have the codes and use a formula to arrive at the code for any particular article. They would therefore need conversion tables to get from the article to the article number and document delivery centres would need a key to get from the article number to the printed pages for journals they do not hold in machine readable form. As currently proposed, the Adonis number does not therefore seem adapted to the transfer of orders from library to document delivery centre.

Some basic considerations

Clearly none of the schemes mentioned above is perfect; indeed no system can take account of all possible interests and all possible uses. But it would be unfortunate if that led to a proliferation of systems, each adapted to only one kind of use. A major problem is that so many different people are involved in the processes of journal publishing and document delivery and each has different capabilities and different requirements.

Some of the basic points that should be considered in devising a system of unique article identifiers are set out below. Some are very basic (numbers 1 to 4 for instance); others are more a matter of taste. The list does not claim to be comprehensive; readers will almost certainly add to it.

1. It must be workable. The more difficult it is to use, the less likely it is to be used. The easier it is to make a mistake, the more errors will occur. For the time being at least article identifiers will require human intervention at many stages. The longer the code, the more likely that mistakes will be made.

2. It must also be suitable for computer manipulation. Even computers have restrictions on the length of code that they can sort easily.

3. It should be possible for anyone who has the relevant journal issue to find the article quickly from the article identifier without resorting to some other source of information. The simplest method might be to give the page number(s).

4. The article number should be readily available to those who order from document delivery centres. A common characterstic of those ordering documents is that they do not already have copies of them. Requests for documents are initiated by library users who learn of things that might interest them from a variety of sources, including *Current Contents* and the *Science Citation Index*, other indexing and abstracting services, word of mouth, and

references in other periodicals. Unless these sources carry the article numbers, how will the individual requesting a document or the library ordering it know what the number is?

Is there any point in a system that requires the ordering library to convert the traditional reference (author, title, journal, date of publication, volume and page numbers) into a unique article number for transmission to the document delivery centre, which the document delivery centre then converts back to a traditional reference with page number in order to find the printed copy of the article that has been ordered? Of course it would be possible to provide translation tools, but it seems unlikely that the total system would be more efficient or more economical.

5. The rules should be universal, clear and unambiguous, so that anyone can apply them to construct identifiers for all articles in any journal, given copies of it. That would make it possible to generate article identifiers for articles published prior to any article identifier scheme or for which the publishers have failed to provide codes.

The identifier should not be affected by ambiguities about what counts as an article and what does not. Two people using a system that identifies the first item as 1, the second as 2, and so on might come up with very different numbers for the last item if one included letters to the editor, filler items, and every short note on books received, while the other omitted them.

6. If the article identifier were printed on the first page of an article, that would allow both the document delivery centre and the recipient of the document to check that what had been ordered was what was being supplied.

7. If the structure of identifying codes is such that they can be easily sorted into order of publication within a journal (or an ISSN), document delivery centres could automatically sort requests into a 'picking order' of issues from the shelves, or articles from a CD-ROM. The lists would start with the first article in the journal and end with the last article in the most recent issue.

8. If the article identifier gave the first and last page of an article, it would be clear how long the article was, which should be helpful in estimating the costs, accounting for copies made and making payments for photocopying licensing schemes.

9. It is not unknown for journals to publish retractions of articles or corrections to what has been published. It would be helpful if the article identifier could tie these to the original articles, but that may not be possible.

10. The article identifier should be allocated at the earliest possible moment. That could help in spreading information about forthcoming articles and in the running of editorial and production routines.

11. If all citations of articles also carried the article identifer number, it would help those who wanted copies of the items referred to. The disadvantage is that the reference would become longer. Many authors are likely to be unwilling to include them. Proof-reading would be difficult, and it is likely that many errors would creep in. As is well known (St. Cyr, Domelsmith & Houk, 1980) authors often make elementary mistakes indicating year of publication, volume and page numbers. If included in references, the identifier could also be helpful in the editorial offices of those journals that check bibliographic references against their own database of previously published material.

12. Wherever an article identifier is used, it should be given in full. If for instance, the ISSN is omitted from the contents list printed in the journal, or from the first page of articles, few are likely to recognise that it is missing or to hunt for it elsewhere.

I think many people concerned with libraries, document delivery systems, or photocopying licensing, would agree that most things in this list are desirable. The problem is that they are inconsistent. For instance, if a document delivery centre is to be able to get from the article identifier to the actual pages required, without going through a conversion table (4), then the identifier must give the page numbers. That is also required if all ambiguities and doubtful decisions are to be avoided (5). But many journals do not allocate page numbers until a late stage in the production process. It would therefore not be possible to allocate the identifier at a very early stage (6).

Some questions

As well as listing some basic considerations, there are some questions that might be asked about the composition and use of article identifiers. My impression is that sometimes the answers have been taken for granted, or the implications have not been fully considered. For instance:

What is the minimum information that is absolutely necessary to identify an article satisfactorily? It is tempting to start the article identifier with the proposed issue identifier. But is this the right approach? It inevitably leads to a code that is very long. Is all the information needed to identify a journal issue necessary to identify an article?

Many journals are published in volumes in which the pages are numbered consecutively and issued in a number of parts/issues/numbers. For these journals, is it necessary to give the issue number as well as the volume and page numbers? Once journals are rebound, the issue number is irrelevant for document delivery purposes; and issue numbers are only given in citations when there is no volume number. In any case, many journals, especially scientific, technical and medical ones, give the pagination on the spine of each issue. Would it not be better to press for all publishers to do this than to clutter up the article identifier with a block of numbers that should be unnecessary?

Does the article identifier need to give the date of publication? Is it not enough to give the volume number or the issue number for journals not published in volumes? If the year is given, should it be the year in which the issue was published or the year in which the volume was completed? They will be different in a fair number of cases.

If a number of items begin on the same page, do we need to distinguish between them, or at least between those that start and end on that page? The basic unit in document delivery is a page; fractions of a page are not relevant. Can we manage without recording which articles (as distinct from which pages) have been photocopied? Does it matter for online searching?

It might be useful to be able to distinguish between different articles on the same page of a newspaper or tabloid publication. But can we envisage these publications printing article identifiers? After all, few, if any, manage an ISSN. Their online files often contain reports and stories that have not been published and so would not be covered by the scheme. Perhaps we can ignore this complication in the interests of getting a scheme that is workable.

Even if the identifier omits the date, the issue number if there is a volume number, and any means of distinguishing between two items that start and end on the same page, the article identification code is likely to be rather long. It might, for instance, consist of the ISSN (8 digits plus a hyphen), the volume or issue number (perhaps 4 digits) and the page number (perhaps 4 digits for the first page only, or nine if both first and last are given separated by a hyphen). A check digit would be a sensible provision, and separators between the fields. That brings us to about twenty digits if only the first page of the article is indicated.

Do we need a code that has fixed length fields, or variable length fields, or the possibility of using either?

Can we expect publishers to allocate and print numbers of say twenty digits on each article and in the contents list of the journal? Can publishers be persuaded that the extra effort will bring some benefit to them and not simply to document supply centres which currently copy their journals without payment? They are more favourably disposed towards photocopying licensing centres. Would incorporating this identifier make their admininistration less costly and benefit publishers that way? Could it reduce administrative costs in libraries so more could be spent on publications?

What will happen when mistakes are made? Will publishers (or anyone else) get the codes right every time? A traditional learned journal might have between 6 and 20 articles in an issue, so the problems would not be so great. But what of those which publish a great many small items? For some journals it could perhaps add another page to each

issue, apart from the high labour costs in organising it. If publishers cannot be persuaded to allocate and print article identifiers, what will the source be? How will the identifiers be promulgated?

Who will be the bold library clerk who returns a list of requests from a distinguished professor or the director of the research department saying that unless a twenty digit code is clearly printed beside each item, the library will not attempt to obtain them? Who is the librarian who will defend this requirement to a Senate or Faculty meeting? Or will some member of the library staff have to convert requests for documents into article identification codes? Where is the author (who already finds producing his list of references one of the most tedious and least rewarding aspects of writing a paper or a book) who will insert codes of even half this length beside each reference?

What are the real costs and benefits of the system? Initially many of the costs could fall upon the publisher, but they will eventually be passed on to the library community in the form of higher subscription prices. Will the savings that can be made in the total costs to libraries of document delivery be greater than the increases in subscription prices?

One might even ask if the possible advantages in having article identifiers are the same as they were when the idea first surfaced? Facsimile transmission, the sending of information from one computer to another and the searching of text are now much cheaper and faster than might have been expected then. Perhaps there are other potential uses for article identifiers that have not so far been considered?

Conclusions

In theory, there are many possible uses for unique article identifiers. In practice, unless they are readily accessible and easier to use than conventional references, or give added information, they will be ignored. It would be a pity if, for want of thought, an impracticable system was adopted, for that will put the cause of article identifiers back for many years. Before any scheme is drawn up, let alone adopted, the information community - authors, editors, printers, publishers, librarians, document delivery centre, readers, photocopying licensing centres, and suppliers of library software - needs to consider how they will be used; what uses are most important; how they will work in practice; what is essential and what can be sacrificed if need be. What is the minimum length that we can allocate to any field in order to accommodate virtually any learned journal?

Clearly, it will not be easy to work out a scheme that will satisfy the basic minimal criteria, and it certainly cannot be done without more research. This might be done by collecting opinion and fact from as large a body of volunteers as can be persuaded to put forward their views and attempting to reconcile the conflicts, or it could involve a more formal study of, for instance, the sources of requests for documents. Information, opinion and ideas from all groups involved and from many countries should be sought before any standards are proposed.

Can funds be found to set such a project in train? It would surely be less expensive to do that than to introduce an identifier that proved in practice to be unsound, inefficient or unacceptable to one group or another.

References

W.A. St. Cyr, L.N. Domelsmith, K.N. Houk Experimental Approaches to the Date of Origin of Koopmans' Theorem. The Journal of Irreproducible Results, 26, (1) 23-24.

Pageant Publishing is at 5 Turners Wood, London NW11 6TD.