The growth and number of journals

The number and the growth characteristics of scholarly journals have been matters of some debate for a considerable time. Many assumptions about the future of the scholarly communication system have been predicated on very high estimates of the number of journal titles and their seemingly inexorable growth. This article argues in favour of a novel approach to estimating journal numbers and provides results that are consistent with other lines of enquiry. An analysis of the results also allows a model of journal growth to be developed that matches other statistical observations.

MICHAEL MABE
Director of Academic Relations, Elsevier, and Visiting Professor, Department of Information Science, City University, London

Introduction

How many journals are there? How quickly are they growing? And why? There are few more disputed questions in scholarly publishing. Trying to put a value to the number of titles, no matter how approximate, would be of interest to librarians and publishers, as well as information scientists. It is certainly true that we know that the number of active titles has grown but determining the size of that growth has led to no consensus. There have been a number of studies that have tried to answer these questions. Meadows [1, 2] has examined this problem on a number of occasions and has produced estimates of the number of journals ranging from 10,000 in 1951 to 71,000 in 1987. Other research has estimated that the number of scholarly periodical titles being published at the end of the twentieth century would exceed one million[3].

The huge size of these estimates and their variability has inevitably spawned jeremiads about the flood of literature and its effect on scholarship. There have been a few voices that have pointed out that the state of being overwhelmed by the literature may be more perceptual than actual [4]. However, these reasoned responses have been infrequent and usually ignored. The purpose of this article is to provide the broad publishing and serials readership with data from a recent study on this topic, which may help to illuminate some of the debates.

The study reported here [5] was conducted using *Ulrich's Periodicals Directory* on cd-rom, Summer 2001 Edition. There are many other potential sources for such an analysis, but only *Ulrich's* attempts to cover all serial publications and to classify them by a number of criteria. *Ulrich's* also has the undeniable advantage of being available in a readily researchable cd-rom format as well as online.

A question of definition

The main reason there have been so many varying estimates of the number of learned periodicals in the world is almost entirely down to the simple matter of definition. What exactly constitutes a journal? More importantly, what constitutes a learned journal? The most important characteristics of a learned journal are the nature of its content, whether the title is actively publishing at the present point in time, and whether the content has been through a peer review system to ensure its quality. Unless these key distinctions are taken into account when estimating learned journal numbers from directories such as *Ulrich's*, erroneously high values will be obtained.

The approach adopted in our study was to

define the scholarly, scientific or learned journal as having the following characteristics:

It is a serial publication.

It is a type of serial classified as 'academic/scholarly' in the Ulrich's database.

Its most important characteristic is that it is 'refereed', as defined by the Ulrich's classification scheme.

It is still publishing at the present day and is therefore defined as 'active'.

Excluding other irrelevant categories of publication that exist in Ulrich's classification options allows us to refine the definition set further.

This leads to a search query with the following characteristics:

'active' AND 'academic/scholarly' AND 'refereed'
ANDNOT
'A&I Services', 'Audiocassette', 'Bibliography', 'Braille', 'Broadsheet', 'Catalog', 'Consumer Publication', 'Corporate Report', 'Directory', 'Government Publication', 'Internal Publication', 'Looseleaf', 'Magazine', 'Newsletter', 'Newspaper-distributed', 'Newspaper, 'Record', 'Standard', 'Tabloid', 'Talking Book', 'Trade Publication', 'Video Cassette', 'Yearbook'.

When this filtration exercise is performed on the *Ulrich's International Periodicals Directory* on cd-rom, Summer 2001 Edition, the number of active, refereed academic/scholarly serials comes to 14,694 for 2001. This figure is noticeably at odds with estimates given by other workers mentioned above but almost certainly represents a much more realistic number.

One way of testing the reasonableness of this number is to conduct a thought experiment with ISI data. The ISI citation database covers about 8,000 journal titles. Because of the operation of Bradford's Law[6], a bibliometric version of the Pareto Law (often called the Matthew Principle: 'to him that hath shall be given'), the ISI journal set represents about 95% of all journal citations found in the ISI database. An extrapolated version of the Bradford Law plot for the ISI journals is shown in Figure 1. Due to the logarithmic nature of the relationship, moving from 8,000 source journals and 95% citation coverage to nearly 100% coverage almost certainly means the doubling (or more) of the number of journals covered. This would represent about 16,000 titles, a figure closely similar to the calculated value for 2001.

Another approach is to use estimates of the number of researchers and authors to calculate how many articles are being published each year and to use that to estimate journal numbers.

Figure 1. An extrapolated version of the Bradford Law plot for journals covered by ISI

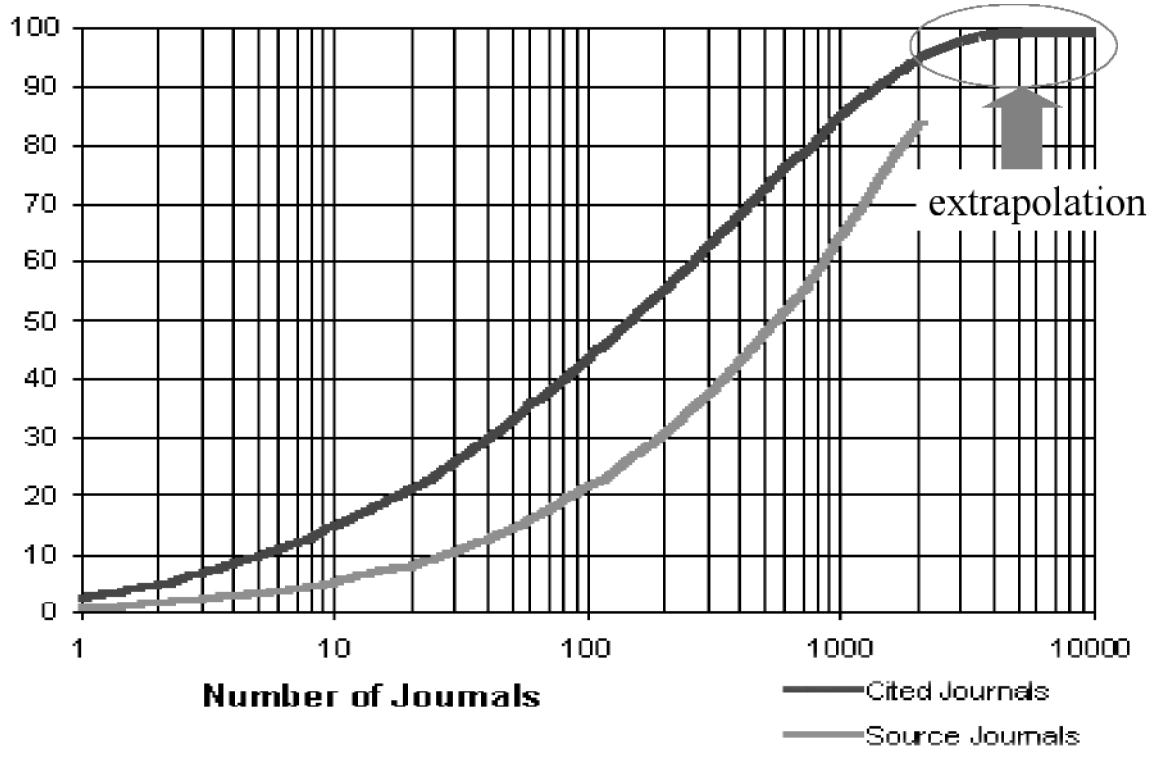
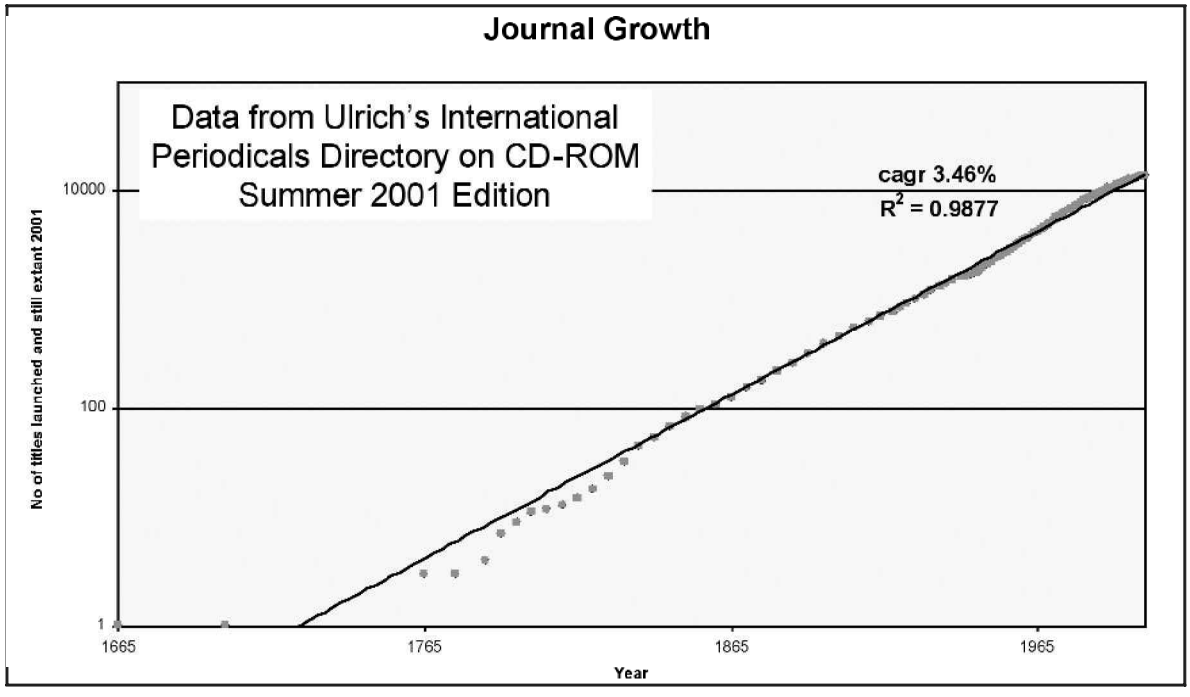


Figure 2. Active refereed scholarly/academic journal title growth since 1665 (log scale)



Surveys of the number of researchers in the world conducted by UNESCO estimate that there are about five to six million of whom about one million are unique repeat authors every year [7]. According to work by Tenopir and King [8] the average productivity per author is about one paper per unique author per year. Consequently there are about a million papers being published each year. For the ISI journal sets the average journal publishes 100 articles per year. Therefore, assuming that the non-core journals not covered by ISI were to be the same average size we can calculate that there must be at least 10,000 journals publishing currently. This is almost certainly an underestimate, as non-core formats are likely to publish fewer than 100 papers annually on average. If the non-core set were assumed to publish only 50 papers per year, the journal total would become 13,333.

Although these approaches are far from perfect, nevertheless they do provide some assurance that the Ulrich's data figure of 14,694 active, peer reviewed scholarly and academic journals is closer to the truth than the hundreds of thousands sometimes quoted by other researchers.

Growth of active journals since 1665

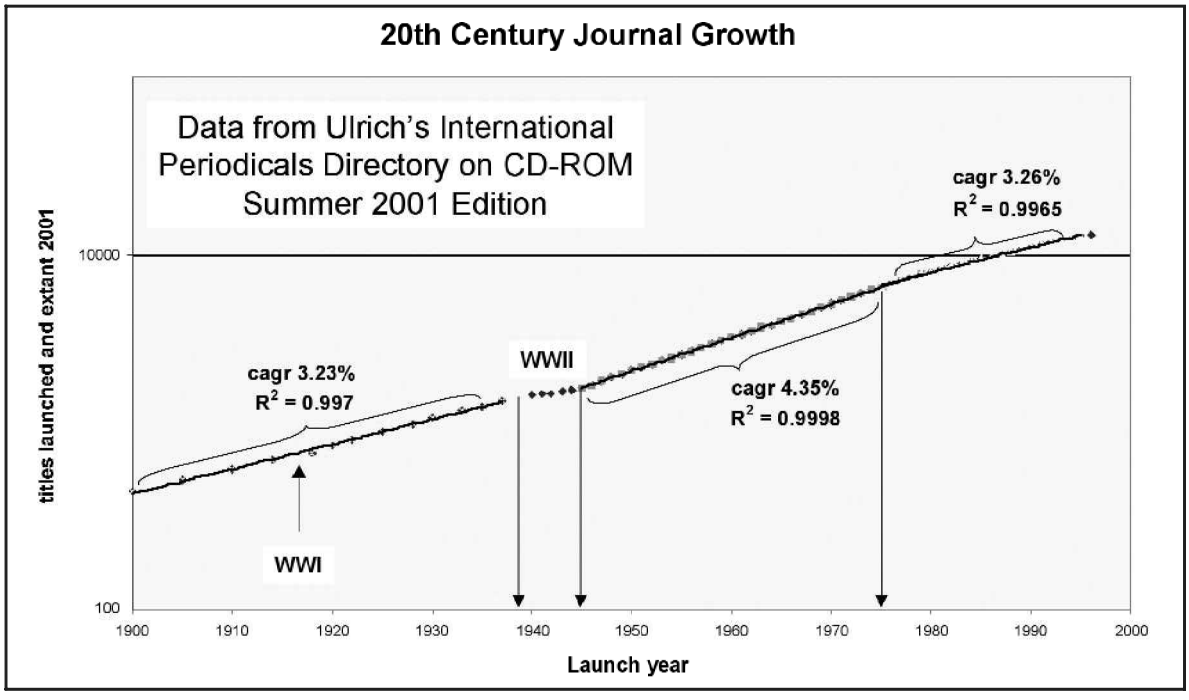
It is possible to repeat the calculation of journal numbers for the year 2001 for any previous year in the Ulrich's database. If such a calculation is done for every year since 1665 it is possible to build up the collective growth curve of the number of scholarly and scientific journals launched and still active for each of the last 338 years. A graph of this data is shown in Figure 2, where a logarithmic scale allows the calculation of the average growth rate. Remarkably, for most of the last three centuries, the growth rate of active peer reviewed scholarly and scientific journals has been almost constant at 3.46% per annum. This means that the number of active journals has been doubling every 20 years.

Figure 3 (overleaf) looks at just the twentieth century part of this growth curve in more detail. From 1900* to 1940 the number of active journal titles grew at an annual rate of 3.23%, a doubling time of 22 years. Amazingly, this growth was hardly affected by major world events such as the First World War. From 1945 to 1976 there was another period of consistent and constant growth with an annual figure of 4.35%, representing a doubling time of 16 years. From 1977 to the present day there is a third period of consistent and constant growth, this time at 3.26%.

The first noteworthy observation is that the growth characteristics between 1900 (or even 1860)

** In fact it is possible to trace this linear growth back as far as 1860 without substantially altering the growth rate*

Figure 3. Active refereed scholarly/academic journal title growth in the twentieth century



and 1940, and 1977 to the present day are almost identical, despite the fact that we are dealing with completely different epochs in world history. Less surprising is the post Second World War boom with a high growth rate of 4.35%.

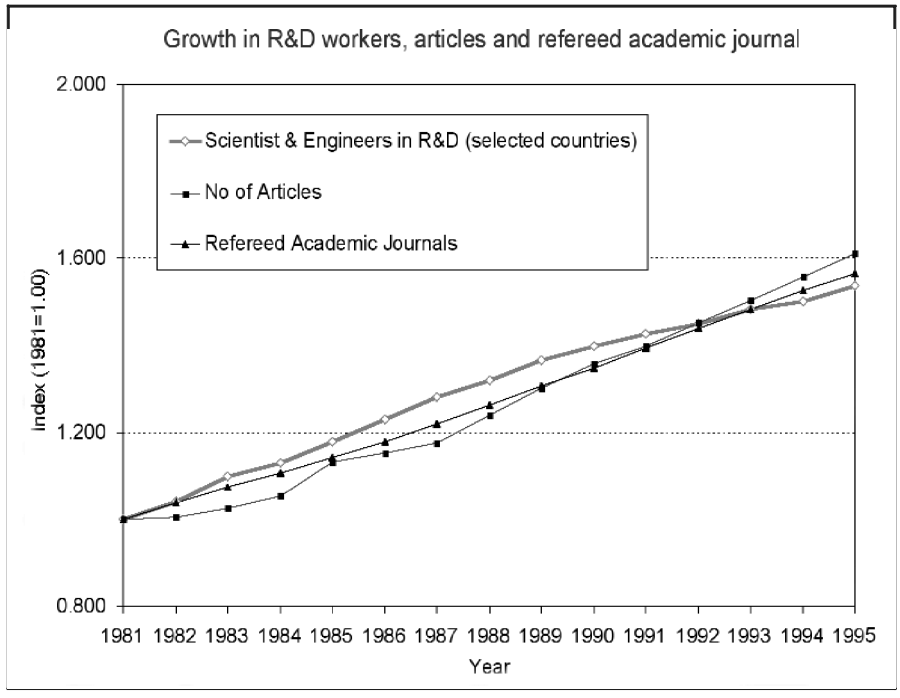
How can this data be interpreted? One approach is to look to the publishing environments during each of these episodes and link them to wider socio-economic developments. The period from 1900 to 1940 could be characterised as small scale or 'innocent' science. During this period there is little funding made available for research from governments and we can assume that the growth is almost entirely due to the collective growth behaviour of all the academic disciplines together. From a publishing perspective this period corresponds to one in which almost all learned publishing is in the hands of the scholarly societies.

The second episode, from 1944 to 1976, corresponds to de Solla Price's 'Big Science'. Science and technology had apparently won the war, so they could now win the peace. Globally, governments invested extremely highly in the advancement of science and technology. This is the era of nuclear weapon technology, the space race and NASA moon landings. In such conditions we might reasonably expect the journal system to grow at its maximum rate, and indeed this is what we find. In terms of the publishing environment we are dealing with a world where commercial publishers, large and small, have appeared on the scene. The system is moving from one largely driven by learned societies to a mixed market of commercial and society players.

The final period, from 1977 to the present day, may be thought of as the period of disillusionment. It is a time when the failure of the science and technology investments of the preceding era to deliver the (over) ambitious expectations of a more innocent age result in disappointment, disillusionment and scepticism. The oil crisis of the 1970s, the increasing public awareness of potential ecological disaster and the turning away from nuclear technology in the 1980s, have led to a relative slow down of government support for research. In the publishing world the journal system continues to be in a mixed market ownership. Yet, despite all of this, journal growth still continued at a constant rate for the whole period, although it has returned to its pre 1940 values.

The return to the same growth rate in the last episode as the first is highly suggestive of a systematic feature. This type of growth behaviour is called 'self-organising' and can be seen in ecosystems in equilibrium. Since such systems tend to grow at a standard rate, when greater levels of nutrient are added this standard growth rate increases until the excess nutrient has been

Figure 4. Growth rates of R&D workers versus journals and articles



consumed. Once this has happened, the system returns to its earlier pre-excess growth rate. The growth of journals during the twentieth century follows this paradigm exactly.

Causation

How can this growth be explained? Figure 4 offers the best explanation for most of the phenomena described. It shows the relative growth rates of the number of research workers, scholarly articles and refereed academic and scholarly journals using the value of each series in 1981 as an index point of 1.00. The connection between growth and the number of journal titles and growth in the number of researchers is unmistakable. This certainly makes sense from a publishing perspective where each journal can be viewed as the sociological outcome of a new grouping of researchers. While it is possible that the journals themselves are merely responding to the growth in research funds, the correlation between R & D funding and the number of refereed scholarly and scientific journals is weaker or more indirect than that between the number of journals and the number of researchers (Fig. 5).

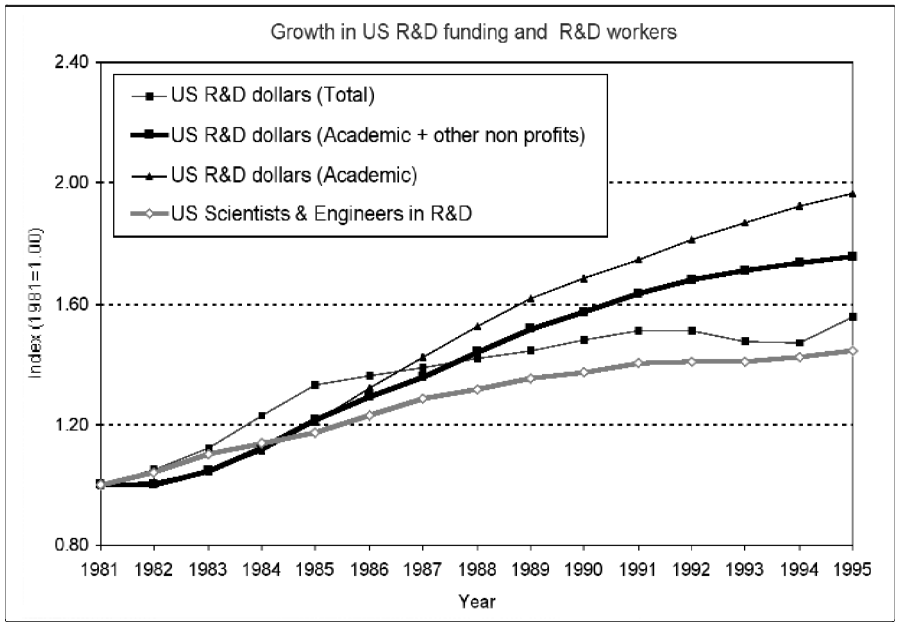


Figure 5. Growth in R&D funding and R&D workers

Overall the figures indicate that an increase of about 100-refereed papers per year globally results in the launch of a new journal. Given that the ratio of unique authors to papers ranges from 1.1 to 0.8 for the period under review, and allowing for rejection of papers, this suggests a growth in the potential annual author community of about 100 to 150 for each new journal, corresponding closely to the size of a typical specialised workshop or conference. Recent work suggests that the reader community as a whole maybe between five and ten times larger than the core author community [7,8]. In this case the potential reader community serviced by the new journal could be anywhere from 500 to 1,000 individuals.

New journal titles can thus be seen as the outcome of a number of competing factors. Connecting journal growth rates with observed sociological phenomena is difficult but not impossible. The growth in new titles is the outcome of attractive forces (such as going with the crowd or peer pressure or the advantages of social interaction) making subject specific social groupings grow in number, versus splitting forces (unwieldy sizes, formation of cliques) that would tend to break apart groups that have grown too large. Given the tendency of each research grouping to have a journal, we could model the characteristics of journal growth in terms of the average number of these groupings at any one time.

Does information technology affect any of this? Harnad and others [9,10] have argued that the new technologies will free scholars from the 'papyrocentric' stranglehold of the journal. This seems mistaken as it confuses the value and purpose of a journal with the medium in which it is published. While new technologies will certainly affect usage behaviour it is much more debatable as to whether it will affect the basic functions of a journal. So far the evidence suggests that the basic dynamic of journals reflecting groups of scientists and the development of science is unlikely to be affected by new technology in anything other than a purely mechanical way.

What of the future?

There is some evidence that after 1997 the rate of growth of new journal titles is starting to fail to keep up with the extrapolated 3.26% rate. This has to be interpreted with some care, since we are very dependent on the efficiency and consistency of classification of periodical characteristics in the Ulrich's database. It seems to take between three to five years for the database to catch up with and classify the moving edge of journal publishing. Consequently figures for journal growth in the period 1997 to 2002 would be expected to show a dip. Previous incarnations of this study showed declines in 1994–1996 which as time moved on later disappeared. The key question will be whether this dip for 1997–2002 persists in four or five years' time. Given what we have learned above about the fundamental mechanisms for journal growth it seems likely that provided the growth in the number of researchers in the world continues so will the growth in the number of journals, and therefore we should not see a decline.

Having said this, though, there is clearly an upper limit to the number of scholars that any society can support. The US National Science Foundation [11] expresses the number of researchers supported in the US in terms of researchers per ten thousand of the general population. Clearly when research numbers per ten thousand reach the limit of economic sustainability then there can be no further growth. As far as we can tell we remain a long way from this maximum.

Many critics will argue that none of this research takes into account the degree to which the number of journals actually launched is affected by the buying power of customers. While it is certainly the case that journal prices have risen rapidly over the last 30 years or so, overall there has been little effect on the growth of active, refereed journal numbers. It is of course quite possible that in the coming decades this aspect of the system will predominate over the almost entirely author-driven one that has been the case for the last 337 years, but at the present this appears not to be the case.

Conclusions

Journal growth rates have been remarkably consistent over time, with average rates of 3.46% from 1800 to the present day. For the twentieth century, the phenomena of growth appear to show a system that is self-organising and in equilibrium, with 3.25% growth predominantly from 1900 (or

even 1860) to 1940 and from 1976 to the present, and an intervening higher growth period from 1945 to 1975. This consistent growth is a factor yet to be taken on board in the world of serials publishing and librarianship, and has important consequences for all members of the publishing chain. While a general causation seems to have been identified (researcher growth rates), detailed examination of how the 3.25% rate arises in practice will keep information scientists occupied for some time to come.

References

1. Meadows, A.J., in: Woodward & Pilling *The International Serials Industry*, 1993, Aldershot, Gower, pp24-27.
2. Meadows, A. J., *Communicating Research*, 1998, London and San Diego, Academic Press, pp15-16.
3. De Solla Price, D., *Little Science, Big Science*, 1963, Columbia University Press.
4. Garfield, E., In truth, the flood of scientific literature is only a myth. *The Scientist*, Vol.5 (17). 12 2, Sept 1991.
5. Mabe, M., and Amin, M., Growth of scholarly and scientific journals *Scientometrics* Vol.51(1), pp147-162, 2001.
6. Garfield, E., Bradford's law and related statistical patterns *Current Contents*, Vol.19. 5-12 12 May 1980; Bradford, S. C., *Documentation*, 1950, Washington DC, Public Affairs Office.
7. Mabe, M., and Amin, M., Dr Jekyll and Dr Hyde: author-reader asymmetries in scholarly publishing *ASLIB Proc.*, Vol.54(3), pp149-157
8. Tenopir, C. W., and King, D. W., *Towards Electronic Journals*, 2000, Washington DC, SLA.
9. Harnad, S., Free at last: the future of peer-reviewed journals *D-Lib Magazine* Vol.5 (12), 1999. http://www.dlib.org/dlib/december99/12harnad.html
10. Odlyzko, A. M., The rapid evolution of scholarly communication, 2000: http://www.research.att.com/~amo/doc/rapid.evolution.pdf
11. National Science Foundation: Science and Engineering Indicators: http://www.nsf.org

Michael Mabe
Director of Academic Relations, Elsevier
Langford Lane, Kidlington
Oxford OX5 1GB
Visiting Professor. Department of Information Science, City University
London
E-mail: m.mabe@elsevier.co.uk