

Research Article

This is a post-print version of an article published in:

Jinha, A. E. (2010). Article 50 million: An estimate of the number of scholarly articles in existence. *Learned Publishing*, 23(3), 258-263. doi: 10.1087/20100308

Article 50 Million: An Estimate of the Number of Scholarly Articles in Existence

Arif E. Jinha

Faculty of Post-Doctoral and Graduate Studies

University of Ottawa



Abstract

How many scholarly research articles are there in existence? Journal articles first appeared in 1665, and the cumulative total is estimated here to pass 50 million in 2009. This sum was arrived at based on published figures for global annual output for 2006, and analyses of annual output and growth rates published in the last decade.

Introduction

From the first model of the modern journal, *Le Journal des Sçavans*, published in France in 1665, followed by *Philosophical Transactions* published by the Royal Society in London later that year[1], the number of active scholarly journal titles has increased steadily. In 2006 there were roughly 23750 titles[2]. There are direct correlations between the numbers of researchers, journals and articles[3]. Björk *et al.*[2] argue that changes in the dynamics of literature-based research, provoked by the communications revolution, have made the article itself relevant today as the basic molecular unit of research communication.

The correlations are revealed by studies in the past decade on global research output that have reported the growth rate and annual figures for researchers, journals and articles [3][4][5][6]. Researchers retire, but more new researchers emerge. Journals fold, but a higher number are introduced for the first time. Changes over time in the number of active researchers and journals describe the dynamics of both publishing and research, and the increase in absolute size of active production[5]. However, the article has a static nature that makes it unique as a metric. Articles, once created and published, are rarely destroyed. They can always be re-activated and through citation each article occupies a position in the architecture that researchers can continue to build upon. The article is born essentially through the efforts of journals and their publishers, but articles survive the death of journal titles. Though disciplines develop distinct fields of inquiry, there are ultimately no fixed boundaries in scholarship – this is a single system of documented written knowledge.

Therefore a metric that describes the quantitative whole of this system – the global total of all modern scholarly journal articles in existence at the present moment or at any point in time – can be useful as a starting point for research into the structure of the system itself. Further, getting better estimates of the global volume of research can enable information scientists to achieve a great deal. The estimates allow them to map the geography of knowledge production, identify routes to retrieval of articles, extract

content, while ensuring its preservation and its availability for use. This paper presents an estimate for the total number of all peer-reviewed articles published worldwide since 1665. Included is a replication of earlier studies showing the current numbers of active journal titles, reported here for the year 2009.

Literature Review

Inquiry into the scope of production of scholarly articles through peer-reviewed journals and the universe of journal titles and articles has never been precise. However, several works exist that attempt to quantify global output of scholarship dating to the post-War 'Big Science' period as well as more recent works from the 1990s until present.

In 1963, Derek de Solla Price plotted the growth of journal titles from 1665 to 2000 giving rise to predictions of an astronomical 1 million journal titles by year 2000. Price also identified key relationships between research investment, the numbers of researchers, and the numbers of journal titles, abstracts and articles. These relationships have been carried forward in more recent research. Estimates of the numbers of journal titles worldwide were made by King et al in 1977 at 57,400 and in 1995 by Meadows and Singleton at 70,000 to 80,000[7]. More recent research reports figures that are far more modest than these earlier estimates and predictions[7]. Mabe and Amin explain in the introduction to the 2001 article 'Growth dynamics of scholarly and scientific journals' that improvements to Ulrich's¹ system of classification allow for more realistic estimates[5], and Mabe follows up in the 2003 article 'The growth and number of journals' with the argument for a novel approach based upon this[3]. Earlier estimates are considered high because researchers were unable to differentiate peer-reviewed journals from other periodicals and could not differentiate active journal titles from those that had closed. Significantly, the growth rate cannot be taken as exponential or cumulative as Price had assumed, and this can explain why we do not have anything near 1 million journal titles today.[3][7].

Mabe[3] used search terms in Ulrich's classification system to filter in scholarly, refereed and active journal publications as well as the 'AND NOT' function for several terms that disqualify a database resource from being included in the definition of a scholarly/scientific journal. From this, Mabe estimated the global number of journal titles to be 14,694 in 2001. Mabe also followed up on Price's identification of the relations between the numbers of researchers, journal titles and articles to test how reasonable the estimate happened to be. This was done through identifying the number of titles indexed

1 Ulrich's Periodicals Directory (Ulrichsweb.com) has been a global source of periodicals information since 1932.

in the ISI database and applying Bradford's law to estimate the quantity of non-ISI indexed journals . The estimation by the second method produced an estimate of 16,000 titles[3]. Therefore the first estimate that defined the parameters more accurately and included only active, refereed and scholarly journal titles can be traced to this study with the result of 15,000 and 16,000 titles in 2001.

Tenopir and King[6] estimated a global annual output of 1 million articles at the turn of the twentieth century, an estimate based on empirical data on the number of active researchers and the average research output per research author. Björk et al produced an estimate of 23,750 journal titles for 2006 using the method introduced by Mabe[3][5] except that Björk et al did not include the 'AND NOT' filter in their study (this does not appear to be necessary in any case). The authors were then able to make the first estimate of global annual output of articles. In order to do this, they distinguished ISI-indexed titles that as a rule produce more articles than non-ISI titles and then determine the average number of articles per title for each category (by an indirect method for ISI described in the study, and by statistical sample of non-ISI titles.) Calculating the sum of titles multiplied by the average number of articles per title for each category gives an estimate of 1,346,000 (rounded) articles for 2006. Incidentally, though ISI titles represent 36 per cent of the total number of active journals, ISI articles represent 70 per cent of the total number of articles[2].

At the time of revising this article, a lively discussion occurred on the American Scientist Open Access forum regarding the wide variance in estimates, demonstrating the ongoing difficulty in getting precise numbers[8]. Morris [9] discusses the limitations of relying on Ulrich's database, stating: "the directory's publishers are entirely reliant on the information supplied by the publishers of the journals listed therein. New journals are often not listed immediately. There can therefore be no hard-and-fast guarantees as to the completeness, currency, or accuracy of that information" [9, pg. 299]. Concern was expressed on the AMSCI forum that a greater number of smaller journals, particularly those published in languages other than English and those published in developing countries would be more likely to be excluded leading to both a skewed view of the universe of academic publishing and an underestimation of its size[8]. This issue was discussed further by Tenopir and King [7] in the recent book 'the Future of Academic Publishing. However, aside from embarking on a manual method of counting titles, Ulrich's remains the most comprehensive database for determining worldwide totals and the most sensitive to filtering for key distinctions such as active, refereed and scholarly titles[7,9]. Moreover, results from Ulrich's have been consistent with what we understand about the relationships between the numbers of

researchers, titles and articles as well as the growth rates [2][3]. Included in this study is a replication of the basic method of searching Ulrich's to determine the number of active journal titles in 2009. From this figure, we can produce an estimate for global annual article output assuming no great change in the average number of articles per title for ISI and non-ISI titles since 2006.

Methods

The estimate is based on the measurement of global output of scholarly output in 2006 reported by Björk et al[2] , and rests on the assumption that[4], Mabe[3], Ware[4], Mabe and Amin 5] and Tenopir and King[6] are correct in reporting a steady increase in the number of researchers, journals and articles over three centuries. While the average rate of increase in the number of journals is reported by these authors to be 3.26 per cent, Ware (2008) reports a growth in article output of roughly 3 per cent per year. This produces a doubling time of just under 24 years.

We chose 1726 as the initial year for our calculations, because it corresponds to the beginning of the line of steady growth of journals shown in Figure 1 (reproduced from *Ulrich's Periodicals Directory* 2001 in Ware[4]). The literature reports steady growth for over two to three centuries[3][4]. The earliest period of publishing, from 1665 to the middle of the 18th century shows less predictable growth. Ulrich's does not produce results for the number of journals as far back as 1726 so this method cannot be applied to determine the start figure. It appears reasonable to start 1726 with a number greater than zero but negligible to the global quantity today. Additionally, the number of articles can be set to a figure that produces closely matched results to estimates for global annual output in the past decade when the 3 per cent growth curve is applied, a form of backward mapping.

When the number of articles for 1726 is set to 344, the curve corresponds closely to Tenopir and King's[6] estimates of annual output at the turn of the millennium as well as the estimate by Björk *et al.* [2] for the number of articles in 2006. This was done using an Excel spreadsheet. Table 1 shows the estimated annual output for each year from 1726, and the cumulative total. Mabe [3] reports that journal growth experienced its largest year-to-year increase during the Big Science period from 1946 to 1976, with lower-than-average rates before WWII and after 1976. The author used the same multipliers for the changes in the journal growth rate to adjust the article growth rate and calculated a sum taking into account this variability. Since the results for both calculations were almost identical, the more straightforward calculation based on the average growth rate is reported here.

Results

This method yielded an estimate of nearly 50 million articles by the end of 2008, with the figure expected to pass 50 million in 2009. In actual fact, the year that the sum of all scholarly articles passes 50 million cannot be determined precisely, but we can report that this result is the first estimate to follow from the current evidence for the grand total of all scholarly articles that exist at the time of writing in 2009 and publishing in 2010. In good humour, the authors can make the claim that this article could itself be the 50 millionth to be published in history!

- Estimated total at 31 December 2008 = 49,234,626
- Estimated total at 31 December 2009 = 50,712,009

Replicating Björk *et al.*[2] and using the search terms 'Academic/Scholarly', 'Refereed' and 'Active' in Ulrich's, the total number of active journal titles for the year 2009 is 26,406.

Assuming little change in the proportion of ISI titles (36 per cent) and using the averages given by Björk *et al.* for articles per title in ISI and non-ISI journals, the total number of published articles estimated for 2009 by relation of journal titles to articles is calculated by estimating the number of ISI titles and subtracting that from the total number of titles for 2009, then multiplying each category by its respective average number of articles. The average number of articles per title reported by Björk *et al.* for ISI titles is 111.7; for non-ISI, the average number of articles is 26.2.

To determine the number of ISI titles, we multiply the total number of titles by 35%.

- $26,406 * 0.36 = 9506$ – the estimated number of ISI titles.

To determine the number of ISI articles, we multiply the number of titles by the number of articles per title for ISI journals.

- $9506 * 111.7 = 1,061,820$ – the estimated number of ISI titles.

To determine the number of non-ISI titles, we subtract the number of ISI titles from the total number of titles.

- $26,406 - 9506 = 16,900$ – the estimated number of non-ISI titles.

To determine the number of non-ISI articles, we multiply the number of non-ISI titles by the number of articles per title for non-ISI journals.

- $16,900 * 26.2 = 442,780$

To determine the annual global output of articles for 2009, we sum the number of ISI and non-ISI articles.

$$1,061,820 + 442,780 = 1,504,600$$

Summing these, the global output of articles in 2009 is 1,504,600.2

The total number articles estimated for 2009 using the 3% growth curve and produced in the Excel table is 1,477,382. The difference between the estimate by the method employed by Björk and that produced by the growth curve is less than 2%.

Discussion and Conclusion

The estimate of the global total of scholarly articles that exist is clearly a ball-park figure, rather than a precise number. However, the study of the size, growth and composition of a global body of scholarship has moved forward in this decade. We can better determine global annual output of scholarships through our understanding of a) the relationship between numbers of researchers, journals and articles; b) the year to year growth rates for the number of active titles and the number of published articles; c) the relationships between ISI and non-ISI journal titles; and d) the improvements to Ulrich's classification system. However, further investigation is needed to test the robustness of each of the relationships and indeed the comprehensiveness of Ulrich's database.

50 million peer-reviewed journal articles is an impressive heritage, and a powerful resource for humanity. In order to manage such a resource in a way that is equitable, useful and sustainable, we would do well to take ongoing interest in where we stand in terms of the access, digitization, search and indexation, and preservation of this global library of knowledge.

Arif Jinha

217 Moffatt Street

Carleton Place

Ontario K7C 3K9

Canada

Phone: 613-889-7506

Email: arif@stratongina.net

Appendix 1 – Calculated data for annual and cumulative article totals (Excel)

<http://dx.doi.org/xxxxx>

Acknowledgements

The author would like to acknowledge the assistance of Azim Jinha with the calculations, and the advice and editorial help of Robin Beecroft (Searchligher) and Moustapha Diack.

References

- [1] Brown, H. 1972. History and the Learned Journal. *Journal of the History of Ideas* 33: 365-78.
<http://www.jstor.org/stable/2709041>
- [2] Björk, B., Roos, A. and Lauri, M. 2008. Global annual volume of peer reviewed scholarly articles and the share available via different Open Access options. *Proceedings of the ELPUB2008 Conference on Electronic Publishing, Toronto, Canada, June 2008*. <http://oacs.shh.fi/publications/elpub-2008.pdf>
- [3] Mabe, M. 2003. The growth and number of journals. *Serials* 16: 191-197
- [4] Ware, M. *Scientific publishing in transition: an overview of current developments*. Bristol, Mark Ware Consulting, 2006.
- [5] Mabe, M. and Amin, M. 2001. Growth dynamics of scholarly and scientific journals. *Scientometrics* 51: 147–62. DOI:10.1023/A:1010520913124.
- [6] Tenopir, C. W. and King, D. W. *Towards Electronic Journals*. Washington DC, Special Libraries Association, 2000.
- [7] Tenopir, C.W. and King, D.W. 2009. The growth of journals publishing. In *The Future of the Academic Journal*, Cope, B., and Phillips A. Eds. Chandos Publishing/Woodhead Publishing Ltd. ISBN 1 84334 416 5.
- [8] American Scientist Open Access Forum. 2009 Archives. See discussions with subject line 'Number of Scholarly Journals in the World.' <http://listserver.sigmaksi.org/sc/wa.exe?A1=ind09&L=american-scientist-open-access-forum&F=l>

[9] Morris, S., 2007. *Mapping the journal publishing landscape, how much do we know?* Learned Publishing 20(4): 299-310.

Figures

Figure 1 Number of journals launched per year (Source: *Ulrich's International Periodicals Directory*, reproduced with permission from Mabe[3].)

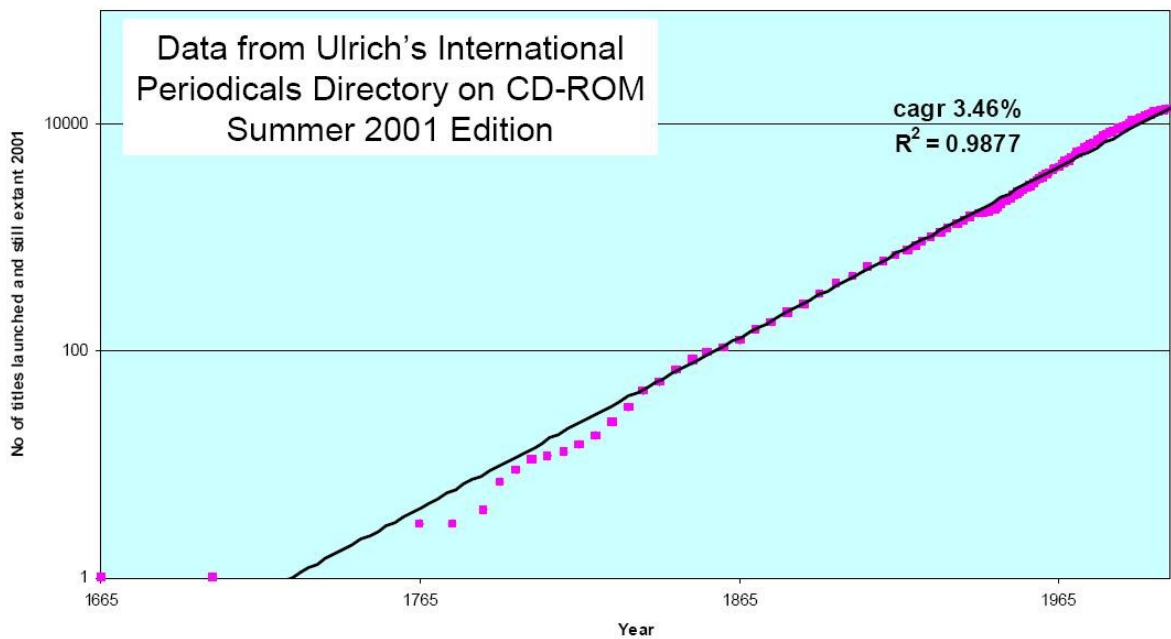


Figure 2 Estimated annual global research article output at 3% annual growth

- * Year 1985 – doubling time of just under 24 years.
- ** 1999 – Corresponds to estimates by Tenopir and King[6], for research output in the late 1990s, of 1 million articles per year.

- *** 2006 – 1.35 million articles corresponds very closely to Björk et. al’s estimate of xxxx (2006).
- **** 2007 – 1.4 million articles corresponds closely to Ware’s estimate of xxx.

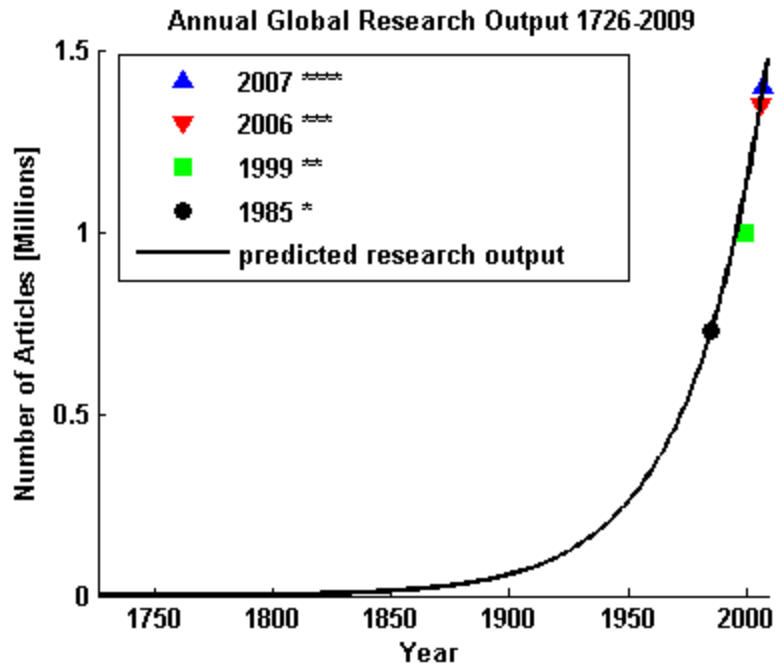


Table 1 Published scholarly journal articles per year 1726-2009

(Can be created from Excel file)