Research Article

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. More recent research reports figures that are far more modest than these earlier estimates and predictions[7]. Mabe explains in the introduction to the 2001 article 'Growth dynamics of scholarly and scientific journals' that improvements to Ulrich's[1] system of classification allow for more realistic estimates[5], and Mabe and Amin follow 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 the ISI database and the application of

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

Bradford's law to estimate the quantity of non-ISI indexed journals to produce a reasonably close figure of 16,000 titles[3]. Therefore the first estimate, using more discriminating techniques for active refereed scholarly journal titles, can be traced to this study and can be stated as being between 15,000 and 16,000 titles in 2001. The global number of active journal titles and global annual article output follow linear growth rates averaging 3.4 per cent and 3 per cent respectively over three centuries[4], and the cumulative number of articles follows an exponential curve.

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. 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 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]. 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 the18th 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.

26,406*0.36= 9506 – the estimated number of ISI titles. 9506*111.7 = 1,061,820

26,406 – 9506 = 16,900 – the estimated number of non-ISI titles. 16,900*26.2= 442,780

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

The total number articles estimated for 2009 assuming the accuracy of the 3 per cent growth rate in the Excel table is 1,477,382. The margin of error of these two estimates is less than 2 per cent.

Discussion - Towards Equity in the Twenty-First Century

The estimate of the global total of scholarly articles that exist is clearly a ball-park figure, rather than a precise number. 50 million articles at face value sounds like serious information overload. However, it is the ability to filter, select and retrieve articles that is of interest to the researcher, and the ability to manage, preserve and make available this research that is of interest to the librarian, information scientist and archivist.

The figure leads us to imagine what a self-organizing global information system of scholarship will come to look like during the 21st century, in the era of global digital communication. Searching for, navigating to, accessing and using research articles are the basis of the communicative part of science, that which rests 'on the shoulders of giants'. While the volume and complexity of research increases, with thousands of databases housing current and archived material, the complexity of access options increases as well. A researcher at University of Ottawa, for instance, gets the benefit of roughly $7 million dollars of subscription-based research per year, a dollar figure that doubled in a decade[10] Provided they can access the internet with sufficient bandwidth, a researcher in the global South may benefit from the increasing percentage of articles that are either published as being open access at the point of publication (the gold route), or available through author deposit to an open-access archive (the green route). Alternatively, they may access thousands of subscription titles through UN programs or publisher concessions (see Table 2). It is no wonder that Willinsky identifies a central search engine, indexing global knowledge resources, as the 'Holy Grail' of the science of indexation. Google Scholar plays this role to a degree, but Willinsky notes that it does not provide the sensitivity required for precise searching such as a subject index [11]. More precise tools, such as those provided by libraries to search databases do provide this, but ultimately cannot provide a global search. Whether the researcher who finds a relevant article title actually can access the article depends on a number of factors: these include whether the article is online or can be accessed in a library or through interlibrary loan; whether they can pay the pay per view cost; whether they have subscription privileges to the journal; or whether the article is available *gratis* through green or gold open access.

50 million peer-reviewed journal articles is an impressive heritage, and a powerful resource for humanity. Yet in order to manage such a resource in a way that is equitable, useful and sustainable, we would do well to take an inventory of where we stand today in terms of access, indexation and preservation of the global body of research. Further research is needed to answer the following questions:

- What are the existing barriers in terms of price for access, search/ navigation/ retrieval/ indexing, availability of technology and isolation from networks for individuals and libraries?
- What percentage of the cumulative global volume of articles is online? What percentage is offline? What percentage of their abstracts is online?
- Of the articles that are online, what proportion is freely available?

- Which offline articles would be found on open library shelves as opposed to being stored in archives? How are offline articles dispersed geographically?
- What are the patterns of digitization of articles originally published in print? How many are open access, in the public domain, or accessible through subscription?
- What are the alternative routes to accessing an article that is not freely available online – e.g. InterLibrary Loan, e-print, re-print? What is the extent of library networks that allow for this?
- What proportion of articles would be considered secure in terms of preservation, and what proportion would be considered vulnerable? What risks are there for the existing system of scholarship? Do we know of articles that have existed but have been irretrievably lost?

It is hoped that describing the quantitative whole of this 'mountain' of literature will provoke further research to refine estimates and to disaggregate the article population, in order to identify routes for search, filtering, navigation, access and retrieval – to mine the mountain as precisely as possible, to serve the curiosity and the pragmatic concerns of humanity. This estimate of the cumulative total of articles in existence comes during a revolution in communications and publishing. The figure of 50 million articles may prompt us to take stock. We can look backwards to understand how the mountain was constructed; consider what it means today; and consider how it can best be directed to expand the benefits of research and the opportunities to engage in it in the future.

Conclusion

This estimate was made possible by the system which it studies, and is itself part of that system. The majority of literature here was easily retrieved online. Although not all the literature cited in this article was freely available, most of it was; the relevant articles could usually be located with the help of colleagues reached through e-mailing networks, as well as through internet search. No expense was needed for the author to obtain this information thanks to a well-resourced library in Canada's national capital. The same cannot be said for every researcher in the world with potential interest in this subject. Moreover, there is no single comprehensive way to search the literature. Therefore, it would appear that the 50 million articles in existence are a resource that is not yet equitably available and does not meet its maximum potential. From a global point of view the information system for peer-reviewed research has barriers that reduce research impact, quality and equity of participation in both use and production of peer-reviewed research. These price and technology barriers are felt especially where access to knowledge is most needed, in parts of the world where the burden of social, economic and ecological

problems are felt the most. According to King et al[12], as of 2004, 31 countries accounted for 98 per cent of the top 1 per cent most-cited literature worldwide. 162 countries accounted for less than 2 per cent, a gross under-representation of the global South. There are also important concerns over preservation; accounting for and managing the global body of literature is a concern that must be shared globally. The good news is that barriers are falling, with Steven Harnad pointing to an inevitable future where the world will have access to:

"The entire full-text refereed corpus online
• On every researcher's desktop, everywhere
• 24 hours a day
• All papers citation-interlinked
• Fully searchable, navigable, retrievable
• For free, for all, forever" [13]

This article calls for action in commemorating this milestone of publishing; the questions provoked are more significant than the result itself, and hopefully their significance is not lost on those who write, publish and read articles like these.

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.sigmaxi.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.

[10] University of Ottawa, 2009. *Collections – uOttawa Library.* Retrieved February 17th 2009 from http://www.biblio.uottawa.ca/content-page.php?c=abt-collections&g=en&s=biblio

[11] Willinksy, J. *The Access Principle*. Cambridge, MA, MIT Press, 2006.

[12] King, D. 2004. The Scientific Impact of Nations: What different countries get for their research spending, *Nature,* 430, 311-316.

[13] Harnad, S., 2003. Open Access to Peer-Reviewed Research through Author/Institution Self-Archiving: Maximizing Research Impact by Maximizing Online Access. *Journal of Postgraduate Medicine*, Vol 49(4), 337-342.

Figures

Figure 1 Number of journals launched per year (Source: *Ulrich's International Periodicals Directory* in Ware[4].

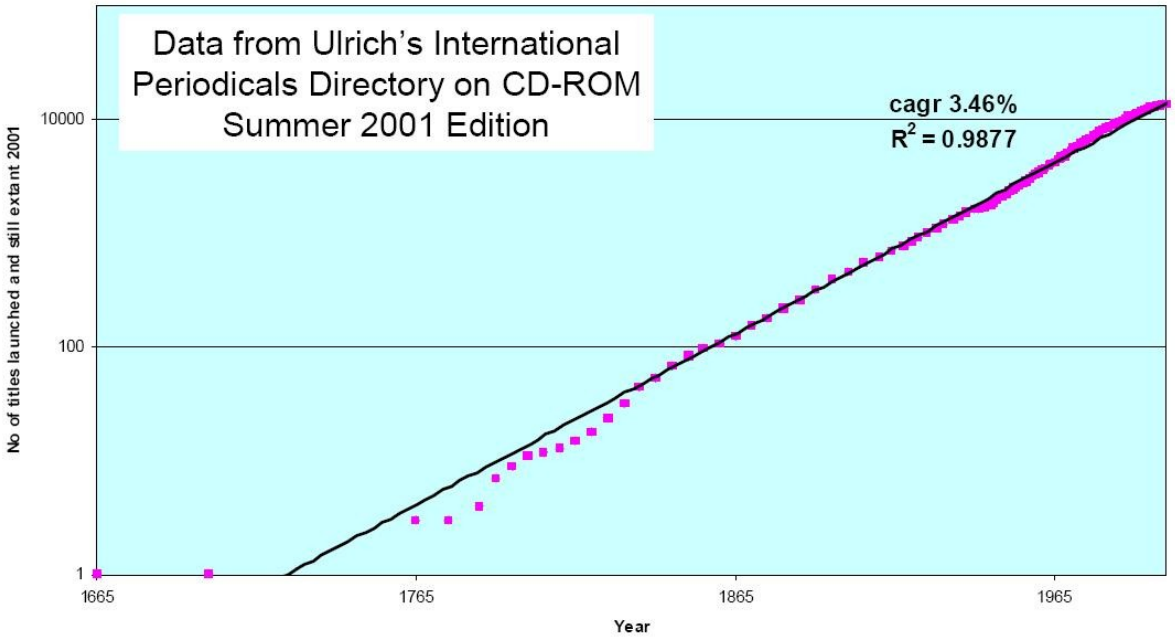

Figure 2 Estimated annual global research article output at 3% annual growth

* 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

Year	Articles Published
1726	344
1727	354
1728	365
1729	376
1730	387
1731	399
1732	411

1733

423

1734

436

1735

449

1736

462

1737

476

1738

490

1739

505

1740

520

1741

536

1742

	552
1743	
	569
1744	
	586
1745	
	603
1746	
	621
1747	
	640
1748	
	659
1749	
	679
1750	
	699
1751	
	720

1752	**742**
1753	**764**
1754	**787**
1755	**811**
1756	**835**
1757	**860**
1758	**886**
1759	**912**
1760	**940**

1761

968

1762

997

1763

1,027

1764

1,058

1765

1,089

1766

1,122

1767

1,156

1768

1,190

1769

1,226

1770

1,263

1771

1,301

1772

1,340

1773

1,380

1774

1,421

1775

1,464

1776

1,508

1777

1,553

1778

1,600

1779

1,648

1780

1,697

1781

1,748

1782

1,801

1783

1,855

1784

1,910

1785

1,968

1786

2,027

1787

2,088

1788

2,150

1789

2,215

1790

2,281

1791

2,350

1792

2,420

1793

2,493

1794

2,567

1795

2,644

1796

2,724

1797

2,805

1798

	2,890
1799	
	2,976
1800	
	3,066
1801	
	3,158
1802	
	3,252
1803	
	3,350
1804	
	3,450
1805	
	3,554
1806	
	3,660
1807	
	3,770

1808	**3,883**
1809	**4,000**
1810	**4,120**
1811	**4,243**
1812	**4,371**
1813	**4,502**
1814	**4,637**
1815	**4,776**
1816	**4,919**

1817

5,067

1818

5,219

1819

5,376

1820

5,537

1821

5,703

1822

5,874

1823

6,050

1824

6,232

1825

6,419

1826

	6,611
1827	
	6,810
1828	
	7,014
1829	
	7,224
1830	
	7,441
1831	
	7,664
1832	
	7,894
1833	
	8,131
1834	
	8,375
1835	
	8,626

1836

8,885

1837

9,151

1838

9,426

1839

9,709

1840

10,000

1841

10,300

1842

10,609

1843

10,927

1844

11,255

1845

11,593

1846

11,941

1847

12,299

1848

12,668

1849

13,048

1850

13,439

1851

13,842

1852

14,258

1853

14,685

1854

	15,126
1855	
	15,580
1856	
	16,047
1857	
	16,529
1858	
	17,024
1859	
	17,535
1860	
	18,061
1861	
	18,603
1862	
	19,161
1863	
	19,736

1864	**20,328**
1865	**20,938**
1866	**21,566**
1867	**22,213**
1868	**22,879**
1869	**23,566**
1870	**24,273**
1871	**25,001**
1872	**25,751**

1873

26,523

1874

27,319

1875

28,139

1876

28,983

1877

29,852

1878

30,748

1879

31,670

1880

32,621

1881

33,599

1882

	34,607
1883	
	35,645
1884	
	36,715
1885	
	37,816
1886	
	38,951
1887	
	40,119
1888	
	41,323
1889	
	42,562
1890	
	43,839
1891	
	45,154

1892	**46,509**
1893	**47,904**
1894	**49,341**
1895	**50,822**
1896	**52,346**
1897	**53,917**
1898	**55,534**
1899	**57,200**
1900	**58,916**

1901

60,684

1902

62,504

1903

64,379

1904

66,311

1905

68,300

1906

70,349

1907

72,460

1908

74,633

1909

76,872

1910

	79,179
1911	
	81,554
1912	
	84,001
1913	
	86,521
1914	
	89,116
1915	
	91,790
1916	
	94,543
1917	
	97,380
1918	
	100,301
1919	
	103,310

1920

106,409

1921

109,602

1922

112,890

1923

116,276

1924

119,765

1925

123,358

1926

127,058

1927

130,870

1928

134,796

1929

138,840

1930

143,005

1931

147,296

1932

151,714

1933

156,266

1934

160,954

1935

165,782

1936

170,756

1937

175,879

1938

	181,155
1939	
	186,590
1940	
	192,187
1941	
	197,953
1942	
	203,891
1943	
	210,008
1944	
	216,308
1945	
	222,798
1946	
	229,482
1947	
	236,366

1948	**243,457**
1949	**250,761**
1950	**258,284**
1951	**266,032**
1952	**274,013**
1953	**282,233**
1954	**290,700**
1955	**299,421**
1956	**308,404**

1957

317,656

1958

327,186

1959

337,001

1960

347,112

1961

357,525

1962

368,251

1963

379,298

1964

390,677

1965

402,397

1966

	414,469
1967	
	426,903
1968	
	439,710
1969	
	452,902
1970	
	466,489
1971	
	480,484
1972	
	494,898
1973	
	509,745
1974	
	525,037
1975	
	540,788

1976	**557,012**
1977	**573,722**
1978	**590,934**
1979	**608,662**
1980	**626,922**
1981	**645,730**
1982	**665,102**
1983	**685,055**
1984	**705,606**

*****

1985

726,774

1986

748,578

1987

771,035

1988

794,166

1989

817,991

1990

842,531

1991

867,807

1992

893,841

1993

920,656

1994

948,276

1995

976,724

1996

1,006,026

1997

1,036,207

1998

1,067,293

******

1999

1,099,312

2000

1,132,291

2001

1,166,260

2002

1,201,247

2003

1,237,285

2004

1,274,403

2005

1,312,635

***

2006

1,352,015

****

2007

1,392,575

2008

1,434,352

2009

1,477,383

able 2

Organization	Programme	Quantity of material
World Health Organization - WHO	HINARI – Health Internetwork	6200 journals
Food and Agriculture Organization – FAO	AGORA	1278 journals
United Nations Environment Programme - UNEP	OARE	3200 journals
International Network for the	- Programme for the Enhancement of	25000 journals & 11000 e-books (to

Availability of Scientific Publications - INASP	Research Information – PERii. - African Journals Online - AJOL	be selected from by consortia), citation &bibliographic databases, document delivery
eIFL	Country-level consortia negotiations	Per country/institution
Directory of Open Access Journals - DOAJ	Gold OA Journals	3884
Directory of Open Access Repositories -OpenDOAR	Open Access Repositories	1300+ repositories