Number of Scholarly Articles per Country

Data on Web of Science listed articles and reviews 2004-2013

Margit Palzenberger
palzenberger@mpdl.mpg.de
Max Planck Digital Library (MPDL)

Abstract

Data on the total number of Web of Science articles per country and year for the publication years 2004 to 2013 have been compiled. This includes both the numbers with articles assigned to multiple countries based on all authors as well as articles unambiguously assigned based on corresponding authors.

Introduction

Analyses dealing with the global scholarly journal market need estimates of the total number of articles published and their distribution among countries. Open Access Gold transition strategies that are aiming at the replacement of the present subscription system with open access business models based on article processing charges (see Science Europe 2013) need a special refinement with respect to the "corresponding author". Only those articles with the corresponding author affiliated to a given entity (country or institution) are of relevance for estimates of transition costs, as typically the corresponding author is responsible for settling the invoice for article processing charges (APCs).

There is no definite source for total numbers of articles from all scholarly journals worldwide and any estimates crucially depend on the source chosen, the definition of "scholarly article" and assumptions for projections. The data presented here rely on a subset chosen by Thomson Reuters (TR) for their product "Web of Science". This currently includes more than 12,000 journals from a wide variety of scholarly domains selected by criteria outlined by Jim Testa (2012).

Methods

"Web of Science Tagged data" (© Thomson Reuters, TR) are regularly transferred to and processed for an in-house data warehouse (Oracle 11g). For this analysis the files for SCI, SSCI and AHCI delivered by January 2015 were used. The additionally available product "Conference Proceedings" was not included. Raw data cannot be made available for the public for commercial licensing reasons. However, organizations subscribing to the TR data source should be able to verify the data presented.

Tagged data include information on the authors' affiliations within to data blocks: C1 (repeated) for every author and RP (maximum 1 entry) for the "reprint author" which usually is synonym to the corresponding author. The tags NU within these blocks refer to the countries of the affiliations of the authors.

Only items with the document types article (dt = '@') and review ('r') where selected for further processing. For these total counts and counts per country where compiled for the publication years 2004 to 2013. 2014 was not yet included as Web of Science data typically still have substantial gaps up to May of the subsequent year. Data for 2014 will be published as updated version of this data set during summer 2015.

Counts per country where compiled in two versions:
- total counts allowing for multiple association of an article to several countries
- counts for the corresponding author with only one country per article

Aggregated data are attached as tab delimited ascii files together with further source and variable description in markdown format. Descriptive statistics and graphs (png format) were produced with the aid of the open source statistical package R.

Data Descriptives

global dataset

The first table shows key counts for the raw data used in this data compilation. TR continuously is changing its journal coverage and has broadened it above average in 2008. Thus the increase in article numbers in the 9-years period from 2004 to 2013 (60% in total) does not reflect global increase in an unbiased way.

The number of countries per article increases throughout the 10 years investigated. Thus the overall share of corresponding author articles to all articles declines from 80% to 75%. This might be driven by an increase in cross-country collaboration during these years.

subset	a2004	a2005	a2006	a2007	a2008	a2009	a2010	a2011	a2012	a2013
01 articles total [thousands]	928	975	1.028	1.080	1.157	1.208	1.249	1.326	1.391	1.456
02 articles with affiliation entries	899	944	999	1.050	1.129	1.183	1.226	1.305	1.369	1.437
03 sum of articles per country	1.119	1.180	1.255	1.327	1.433	1.513	1.588	1.706	1.808	1.915
04 corresponding author articles	896	941	995	1.046	1.124	1.178	1.221	1.301	1.365	1.433
01 increase to previous year [%]		5	5	5	7	4	3	6	5	5
ratio 04 to 03 [%]	80	80	79	79	78	78	77	76	76	75

tab. 1

data per country

If data are split for countries (n = 211 + 1 for unclear entries) the following characteristics are found:

For the "median country" the total number of articles has increased during the 9-year period by 120% and thus well above the global ratio of 60%. This can be ascribed to a higher publishing activity as well as a broadening of the TR selection process, both with respect to smaller countries.

The share of articles with corresponding authors varies widely among the countries. For a "median country" 41% of the articles published in 2013 had a corresponding author from that country. This value has also declined since 2004 (56%).

measure	N	Q1	med	mean	Q3
increase of total number of articles [%]					
increase 2013 compared to 2004	196	68	120	355	200
increase 2013 compared to 2009	198	18	39	61	67
share of corresp. to total articles [%]					
2004	196	31	56	51	72
2009	198	30	52	50	69
2013	207	25	41	42	61

tab.2 (N = number of countries, Q1 and Q3 quartiles, median, mean)

The difference between the overall share (75%) and the "median country" (41%) can be understood by the skewed nature of country wise article numbers and a significant correlation of the share of articles with corresponding author to the total number of articles.

Such a pattern is to be expected for different group sizes as smaller groups have a higher probability of inter-group collaboration than larger groups. This, however, might be further modified by the language and/or politically based cohesiveness of a country's scholarly community.

The following table and figure outline this effect. The table shows the top 16 countries making up 80% of all articles in 2013. Their corresponding author article shares lie well above the median value of 41%. The figure shows this pattern for all countries with at least 10 articles. A third degree (cubic) polynomial well models that overall pattern.

| code	country	articles and reviews published in 2013			
		total	corresp.	share [%]	cum corresp [%]
us	United States	388.062	309.004	80	22
cn	China	221.584	199.004	90	35
gb	Great Britain	111.332	73.292	66	41
de	Germany	103.666	70.243	68	45
jp	Japan	79.600	65.830	83	50
in	India	53.527	46.885	88	53
fr	France	72.248	46.839	65	57
it	Italy	63.352	46.301	73	60
kr	South Korea	52.156	44.460	85	63
ca	Canada	64.261	44.378	69	66
es	Spain	56.482	41.045	73	69
au	Australia	54.398	37.581	69	72
br	Brazil	39.799	33.386	84	74
tw	Taiwan	28.142	24.356	87	76
nl	Netherlands	38.218	24.043	63	77
ir	Iran	26.287	23.642	90	79

tab. 3
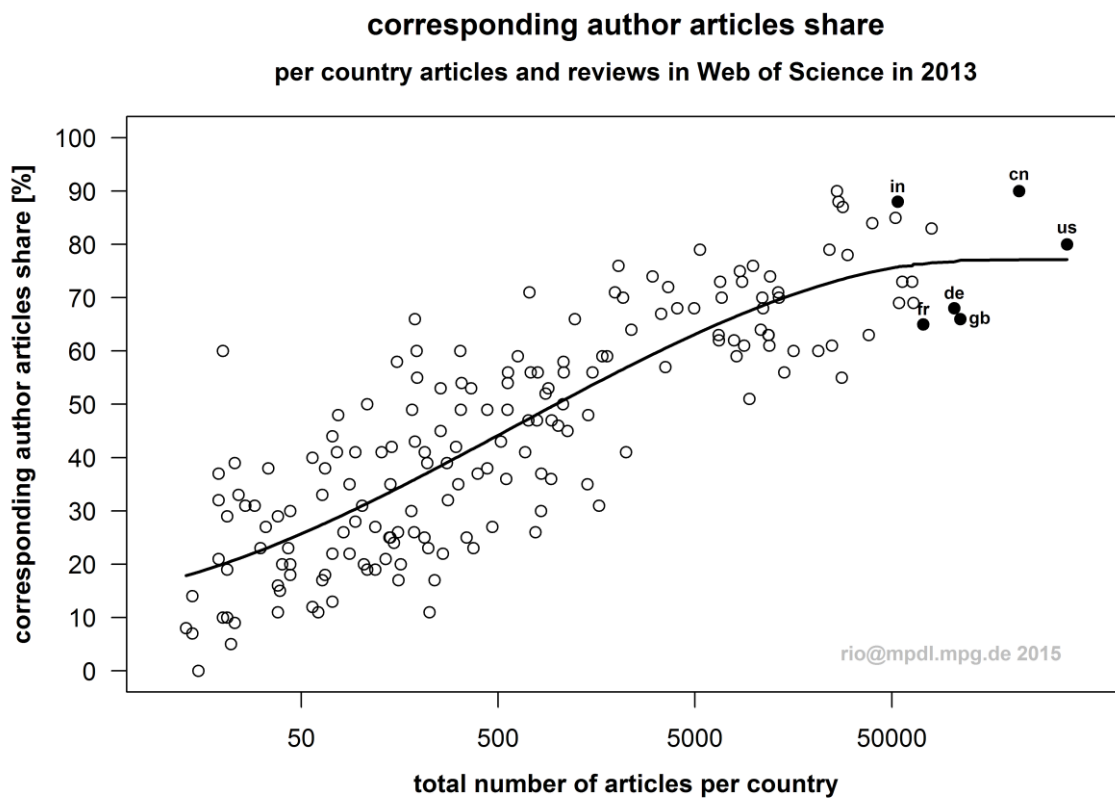
fig. 1

Data Provided

- data definitions (mpdl_rio_wosartbyco.md)
- global data (mpdl_rio_wosartbyco_global.tsv, 11 variables, 1 header + 4 data lines)
- country data (mpdl_rio_wosartbyco_country.tsv, 32 variables, 1 header + 212 data lines)
- figure (mpdl_rio_wosartbyco_share.png, 4800x3600, 132 KB)

References

Science Europe (2013). Science Europe Position Statement: Principles for the Transition to Open Access to Research Publications. Online at http://www.scienceeurope.org/uploads/PublicDocumentsAndSpeeches/SE_OA_Pos_Statement.pdf. [2015-04-24]

Jim Testa (2012): The Thomson Reuters journal selection process. Online at http://wokinfo.com/essays/journal-selection-process/ [2015-04-24]

Acknowledgements