Open access levels and patterns in scholarly articles from India

*Rajesh Piryani, Jyoti Dua and Vivek Kumar Singh**

Open access (OA) has emerged as an important movement worldwide during the last decade. There are several calls now that not only persuade researchers to publish in OA journals, to archive their pre- or post-print versions of papers in repositories, but also institutions and funding agencies to promote OA of research publications. This article examines OA levels and patterns in research output by computationally analysing research publication data obtained from the Web of Science for India during the last five years (2014–2018). Results obtained show that about 24% of research output from India, during the last five years, is available in OA compared to world average of about 30%. More articles are available in gold OA compared to green and bronze OA. Furthermore, OA levels vary in different disciplines, with medical science, physics and biology having higher percentage of their articles available as OA as compared to those like arts and humanities, social science and (surprisingly) information science.

Keywords: Open access, paywall, scholarly articles, scientific publishing.

OPEN access (OA) refers to an academic movement where publications are made available and accessible free of cost with minimum restrictions to different kinds of audience. The Budapest Open Access Initiative (BOAI), first defined 'Open Access' as 'making the publications freely availability on public internet, permitting any user to read, download, copy, distribute, print, search or link to the full texts of these articles, crawl them for indexing, pass them as data to software, or use them for any other lawful purpose, without financial legal or technical barriers'[1]. OA brings with it significant benefits to authors, readers, funders, institutions, etc. and more so for developing and under-developed countries, where enough resources are not readily available with academicians to pay the access charges.

OA now has multiple forms. There are journals that are completely OA and as a result all articles published in them become openly accessible to everyone. The Directory of Open Access Journals (DOAJ) is a community-curated online directory started in 2003, that indexes such journals[2]. Then, there are journals which are usually behind a paywall for access, but make some articles accessible free to everyone, either after payment of an article processing charge or on their own after a particular period from the date of publication. Based on the levels and types of OA, the articles can be classified into the following categories[3].

Gold open access: This refers to an article which is published in a journal that is OA, i.e. it is in a journal where all articles are accessible freely to everyone. Usually OA journals are included in DOAJ.

Green open access: This refers to an article published in a journal that is behind a paywall but the article is uploaded on some repository (either disciplinary or institutional), which makes it accessible free of cost. However, in this case the reuse rights are restricted. Further, journals often impose an embargo period (varying from 6 to 48 months), after which the article can be uploaded on such a repository.

Hybrid open access: This refers to an article published in a closed journal which makes it OA in exchange of an article processing charge paid by authors or their institutions through some agreement with the journal.

Bronze open access: This refers to an article that is free to read on a journal page, but without a license.

Black open access: This refers to an article that is shared on illegal pirate sites, such as Sci-Hub or LibGen. However, this type is not well recognized as OA in the literature.

Closed access: This refers to all other articles that are not openly accessible in legal forms. This includes articles that may be shared on some academic social network or some illegal pirate sites.

The OA initiative practically began in 1991, when Paul Ginsparg (Los Alamos National Laboratory, LANL, New Mexico, United States) set up arXiv, an electronic preprint

Rajesh Piryani is in the Department of Computer Science, South Asian University, New Delhi 110 021, India; Jyoti Dua and Vivek Kumar Singh are in the Department of Computer Science, Banaras Hindu University, Varanasi 221 005, India.
*For correspondence. (e-mail: vivek@bhu.ac.in)

service for the physics community[3]. The advent of the World Wide Web provided the right platform and infrastructure to make OA really possible. However, OA was first defined properly in the 2002 Budapest initiative, which itself was a result of a meeting convened by the Open Society Institute (OSI)[4] in 2001. This was followed by a series of other statements and declarations, including the Bethesda Statement[5,6] and Berlin Declaration[7,8], both in 2003. The OA movement kept gaining momentum thereafter, picking up pace during the last 15 years because of multiple reasons.

Many research funding agencies are now making it mandatory for grant recipients to publish their research outcomes in OA versions. Several disciplinary repositories have emerged during this period that allow authors to freely upload their articles, sometimes the pre- or postprint versions. Authors also upload papers on academic social networking sites like ResearchGate and Academia. A good number of articles are now also uploaded on pirate sites like Sci-Hub. Owing to various reasons, journals are also increasingly making many of their published articles openly accessible.

Open access initiatives in India

The OA culture in India can be traced back to early 90s, when Indian physicists started depositing their preprints in arXiv. They were later joined by mathematicians, computer scientists, biologists, etc. The Institute of Mathematical Sciences (IMSc) at Chennai, set up a mirror server for arXiv. The Vidyanidhi Digital Library[9] set up in 2002 at the University of Mysore is probably one of the first electronic thesis and dissertation initiatives in India, a beginning of repository culture. The call for public access to geographical information made during a meeting of the Indian Academy of Sciences in 1999 (ref. 10), later published as a report in *Current Science*[11], can be treated as one of the initial attempts in India. In 2002, the Indian Institute of Science, Bengaluru set up EPrints@IISc electronic repository. By early 2002, several Indian institutions and organizations had started debating the need for OA and many of them took initial steps in that direction. A report of the National Knowledge Commission discusses about OA and open educational resources[12]. In 2009, Council of Scientific and Industrial Research (CSIR), New Delhi[13] sent a Memorandum to all its 37 laboratories in the country to set up institutional OA repositories, though it took some time for them to start implementing the same.

The more recent effort by the Government of India (GoI) towards OA is the DBT and DST Open Access Policy[14], released in 2014. The Policy has been issued jointly by DBT and DST, the two departments of the Ministry of Science and Technology, GoI, and states that 'since all funds disbursed by the DBT and DST are public funds, it is important that the information and knowledge generated through the use of these funds are made publicly available as soon as possible…'. The guiding principle is public access to publicly funded research outcome. The policy also envisaged that OA will allow percolation of cutting-edge research in higher education curricula, which in turn will raise the standards of technical and scientific education in the country. Through the Policy, institutions were encouraged to set up repositories and deposit all their research outcomes in them. A central harvester *sciencecentral.in* was also established and it was expected that all institutional repositories will be eventually linked to it. The Policy was mandatorily to be followed by institutions receiving core funding from DBT and DST. The *csircentral.net* established by CSIR-URDIP[14] is another example of central harvester that is expected to link the different institutional repositories of individual CSIR laboratories. Indian Council of Agricultural Research (ICAR) has also set up a central repository called *krishikosh*. The more recent National Digital Library[15] project of the Ministry of Human Resources and Development, GoI tried to bring the repository culture to higher education institutions in the country. However, full benefits of these initiatives are yet to be realized.

Research questions

This article aims to explore and analyse the OA levels and patterns in research output from India during the last five years, i.e. 2014–2018. More precisely, it aims to answer the research questions:

RQ1: What percentage of articles from India are OA and what is the prevalence of different types of OA?

RQ2: How has the OA level and type in India varied with time?

RQ3: How do OA patterns in India vary across disciplines and which disciplines show significant annual growth in OA levels?

Related work

There are some recent studies to measure the OA levels in research outputs and also to characterize them with respect to time, publisher or discipline.

Archambault *et al.*[16,17] analysed the proportion of OA peer-reviewed papers at the European and world levels for 2004–2011 and 1996–2013 time periods respectively. They have shown that the tipping point for OA (more than 50% of the research articles are freely available) has been attained in several countries, including Brazil, Switzerland, the Netherlands, USA, as well as in biomedical research, biology, mathematics and statistics.

Multiple studies during 2010–2017 estimated and characterized the OA levels and patterns in scientific

Table 1. Research output from India indexed in the Web of Science (WoS) (2014–2018) that is open access (OA)

Publication year	No. of articles indexed in WoS	No. of articles in WoS having a DOI	No. of articles found on Unpaywall[#]	No. of articles that are OA[#]
2014	67,575	58,310	58,308 (99.99%)	13,834 (23.72%)
2015	70,685	61,695	60,039 (97.32%)	14,347 (23.89%)
2016	76,530	67,857	67,370 (99.28%)	17,280 (25.64%)
2017	78,532	70,861	69,989 (98.77%)	17,014 (24.30%)
2018	84,014	76,780	76,405 (99.51%)	17,856 (23.37%)
Total	377,336	335,503	332,111 (98.99%)	80,331 (24.19%)

[#]Percentage is calculated with respect to the number of articles in WoS having a DOI and found in Unpaywall, i.e. column 4.

publishing[18–23]. They mainly focused on analysing OA levels (gold, green, hybrid) of scholarly articles through various models and studies such as hybrid model for OA, longitudinal studies of OA, and anatomy of green OA, etc.

Piwowar *et al*.[24] used three different samples of 100K articles, each drawn from the Web of Science (WoS), CrossRef and Unpaywall for publications from all over the world. Based on analysis of the samples, they estimated that about 28% the scholarly articles are OA. This accounts for about 19 million of 67 million total articles used in the estimate. They found significant annual growth in OA levels, particularly in gold and hybrid, reaching as high as 45% for the year 2015. They also analysed citation levels of OA articles, concluding that they attain about 18% more citations than average.

Bosman and Kramer[25] collected data from WoS using its oaDOI service, and explored OA levels across research fields, languages, countries, institutions, funders and topics. They found high variations in OA levels on all these dimensions. With respect to countries, they found OA levels varying from 10% to 60%. Further, with respect to disciplines, they found higher OA levels (more than 50%) in disciplines like life sciences, biomedicine and physical sciences compared to relatively lower levels (less than 20%) in social sciences and arts and humanities.

There are, however, no previous recent studies that have focused on characterizing OA levels and patterns in research output from India.

Data and methodology

Data for research publications from India during the period 2014–2018 were obtained from the WoS index[26,27]. A total of 377,336 publication records were found, out of which 335,503 had a DOI. The data were downloaded in June 2019. In order to find out what proportion of data downloaded is OA, the Unpaywall REST API[28] service was used. The data from Unpaywall website were retrieved during 27–29 June 2019. The Unpaywall REST API returns data in JSON format, which contain information like OA status, host type, etc. Programs were written in Python for the processing of WoS records as well as retrieving data from Unpaywall. It was found that out of 332,111 records found indexed on Unpaywall, 80,331 (~24%) were OA.

Since this article is also aimed at characterizing OA levels in different disciplines, each publication record was tagged into one of the 14 broad research disciplines. This categorization into 14 broad disciplines was proposed in a previous study[29] and used thereafter by many publications. The WoS Category (WC) field was used for tagging each publication into 14 different categories, viz. agriculture (AGR), arts and humanities (AH), biology (BIO), chemistry (CHEM), engineering (ENG), environment science (ENV), geology (GEO), information sciences (INF), materials science (MAR), mathematics (MAT), medical science (MED), multidisciplinary (MUL), physics (PHY) and social science (SS). The characterization of research output was done using these tags assigned to each publication record. Different programs were written in Python to computationally process the whole data and produce analytical results.

Results

The first analysis was to find out what proportion of research output from India during last five years (2014–2018) is OA. Results show that out of 335,503 scholarly articles with DOI in WoS, a total of 332,111 records were captured by Unpaywall. Among these records, 80,331 (24.19%) were available in OA. Table 1 shows the detailed statistics for all the years from 2014 to 2018. It can be observed that there is continuous growth in the number of articles that are OA, though proportion-wise it is approximately the same level for all the years. It would be interesting to compare this with the world average. As reported in previous studies for the years 2015 and 2016, approximately 30% of articles worldwide are OA. In the case of India, this level is around 23% of the total research output. Thus, India's OA proportion in its total output is somewhat lower than the world average.

The second analysis was performed on categorization of OA articles into different subtypes of OA, i.e. gold, green, bronze and hybrid. Figure 1 shows the proportion

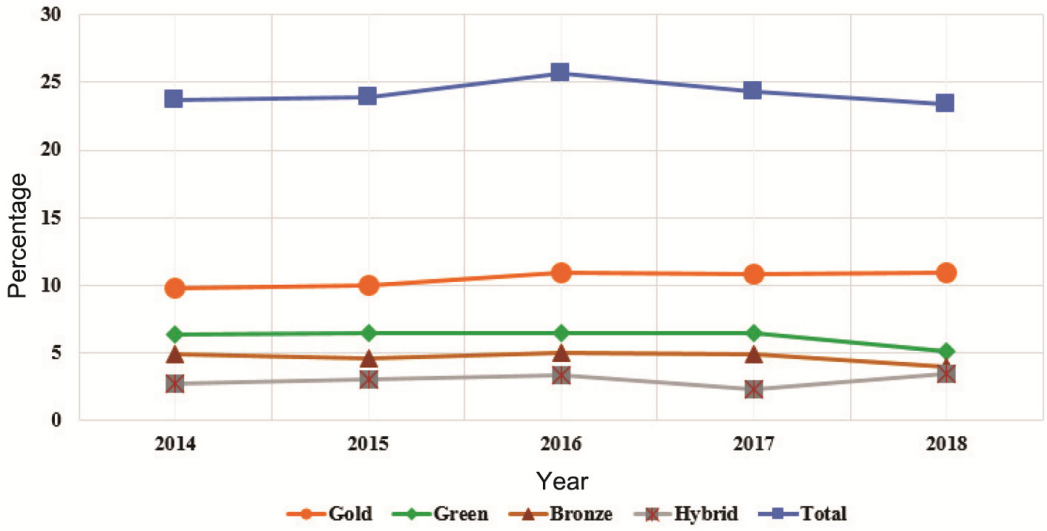

Figure 1. Percentage of articles that are open access, plotted year-wise.

Table 2. Discipline-wise count of articles from India that are Open Access (2014–2018)

			2014		2015		2016		2017		2018	
Discipline	No. of articles in WoS	No. of articles that are OA	No. of articles with DOI	No. of articles that are OA	No. of articles with DOI	No. of articles that are OA	No. of articles with DOI	No. of articles that are OA	No. of articles with DOI	No. of articles that are OA	No. of articles with DOI	No. of articles that are OA
CHE	66,273	8,469	12,617	1,171	12,806	1,115	13,496	1,337	13,630	2,022	13,724	2,824
MED	56,569	27,924	9,835	5,240	10,612	5,469	11,884	6,118	11,667	5,549	12,571	5,548
PHY	38,710	16,843	7,912	2,992	7,127	2,903	7,598	3,411	7,846	3,474	8,227	4,063
BIO	32,510	11,798	6,008	2,369	6,148	2,139	6,394	2,525	6,713	2,514	7,247	2,251
ENG	31,596	4,461	4,526	451	5,252	515	6,389	957	6,980	1,076	8,449	1,462
MAR	27,756	5,935	4,691	815	4,855	915	5,562	1,110	5,897	1,253	6,751	1,842
AGR	17,631	5,561	3,020	839	3,667	1,134	3,516	1,245	3,786	1,264	3,642	1,079
ENV	15,574	4,957	2,563	687	2,901	970	3,164	1,095	3,404	1,138	3,542	1,067
GEO	15,473	4,392	2,698	762	3,044	848	3,038	916	3,179	923	3,514	943
INF	15,022	2,457	1,793	320	2,263	338	2,728	515	3,702	600	4,536	684
MUL	10,158	8,104	1,599	1,257	1,541	1,209	2,351	1,945	2,362	1,803	2,305	1,890
MAT	7,604	3,139	1,273	568	1,405	529	1,487	582	1,649	680	1,790	780
SS	6,492	1,932	1,006	318	1,223	362	1,478	493	1,280	359	1,505	400
AH	498	89	86	17	74	11	98	17	128	29	112	15

of articles from India that are OA, along with the types of OA. It can be observed that gold OA is the most common type of OA, with about 10–12% articles. Green OA is the second common type of OA, with about 6% of articles. This is followed by bronze (around 5%) and hybrid (around 3%). Interestingly, the percentage of green and bronze types has decreased during 2017 and 2018. Gold OA, however, has increased in almost all years, except 2017 and 2018. The slight drop in OA levels in recent two years can be explained due to the fact that many of the recently published articles may still not be available as OA due to several reasons, including embargo periods imposed by journals.

The third analysis was to see if there exist any disciplinary variations in OA levels. Table 2 shows the number of articles from India for each discipline for all the years from 2014 to 2018. It also shows the number of articles for each discipline that are OA. It can be observed that medical science, physics and biology have higher proportion of OA articles compared to other disciplines. Arts and humanities, and social science have relatively lesser proportion of OA articles. Low publication volume in Arts and Humanities and Social Science disciplines in WoS indexed journals and absence of any funding agency driven push for OA may be one reason for lower levels of OA. Interestingly, information science also has a very low percentage of articles that are OA. Figure 2 shows the OA growth curves for all disciplines during 2014–2018. Physics and chemistry show highest growth in percentage of OA papers. Medical science and biology show

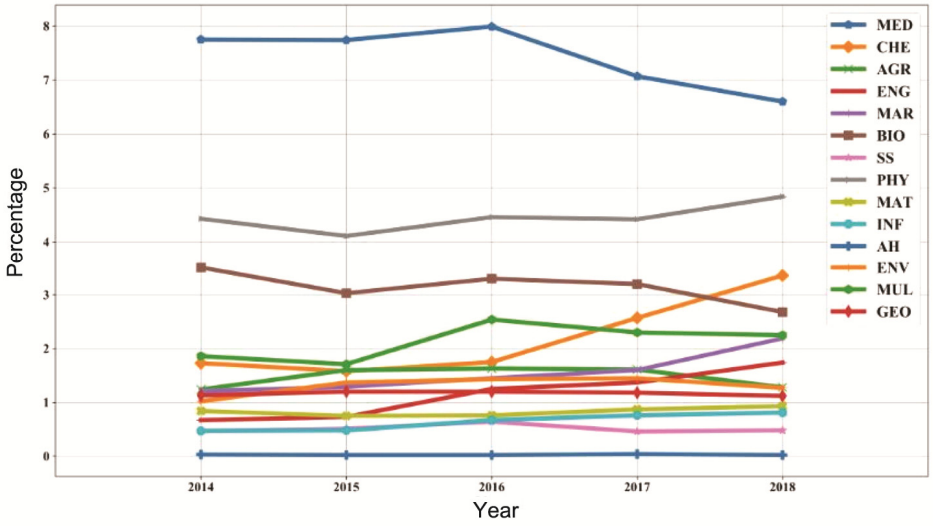

Figure 2. Discipline-wise percentage of open access articles, plotted year-wise.

a small decline in recent periods. Most of other disciplines also show small growth in OA percentage, except arts and humanities, which is the same level, and social science which shows some decline in OA percentage. These disciplinary patterns in OA levels in India are somewhat similar to disciplinary variations across the world as shown in previous studies, indicating higher OA levels in medical science, biology and physics, and low OA levels in arts and humanities, and social science. However, the levels of OA in India for different disciplines are significantly lower than the world average. For example, medical science has about 30% OA articles as the world average compared to only about 8% OA articles in India. Similar differences exist in disciplines like biology and physics.

Conclusion

This study characterizes OA levels in research output from India during the last five years. Interesting findings are obtained from the analysis. First, only about 24% articles from India published during the last five years are OA. This is about 6% lesser than the world average level of OA in WoS data. Secondly, gold OA is the most prevalent type in Indian research output followed by green and bronze OA types. Thirdly, the overall OA level in India has increased little during five years, but some disciplines like physics and chemistry have seen higher growth compared to others. Fourthly, there exist disciplinary variations in the levels of OA in Indian research output, similar to the differentiation seen in the world wide patterns. Therefore, it appears that despite a push from funding agencies and other bodies, the growth in OA levels in research output from India is not significant. More efforts are needed from all quarters of the research ecosystem to promote OA research output in the country. It may be noted that this study does not look at research papers deposited in academic social networks, which if counted as OA can some what change the results. Exploring the proportion of papers from India deposited in academic social networks like ResearchGate and Academia thus remains an interesting exercise for the future.

1. www.budapestopenaccessinitiative.org (accessed on 25 June 2019).
2. Directory of open access. Open access Repositories Operational Status, 2018; https://doaj.org (accessed on 25 June 2019).
3. https://arxiv.org/ (accessed on 25 June 2019).
4. https://en.wikipedia.org/wiki/Budapest_Open_Access_Initiative#cite_note-google.se-3 (accessed on 25 June 2019).
5. https://en.wikipedia.org/wiki/Bethesda_Statement_on_Open_Access_Publishing (accessed on 25 June 2019).
6. http://legacy.earlham.edu/~peters/fos/bethesda.htm (accessed on 25 June 2019).
7. https://en.wikipedia.org/wiki/Berlin_Declaration_on_Open_Access_to_Knowledge_in_the_Sciences_and_Humanities (accessed on 20 August 2019).
8. https://openaccess.mpg.de/Berlin-Declaration (accessed on 25 June 2019).
9. Urs, S. R. and Raghavan, K. S., Vidyanidhi: Indian digital library of electronic theses. *Commun. ACM*, 2001, **44**(5), 88–89.
10. Indian Academy of Sciences; https://www.ias.ac.in/ (accessed on 20 August 2019).
11. Ramachandran, R., Public access to Indian geographical data. *Curr. Sci.*, 2000, **79**(4), 450–467.
12. NKC, Report to the Nation 2006–09, National Knowledge Commission, 2009; http://14.139.60.153/bitstream/123456789/112/1/National%20Knowledge%20Commission-Report%202006-2009%20.pdf (accessed on 20 August 2019).
13. http://www.csir.res.in (accessed on 20 August 2019).
14. DBT and DST Open Access Policy on open access to DBT and DST funded research; http://www.dst.gov.in/sites/default/files/APPROVED%20OPEN%20ACCESS%20POLICY-DBT%26DST-%2812.12.2014%29_1.pdf (accessed on 20 August 2019).
15. National Digital Library of India, https://ndl.iitkgp.ac.in/ (accessed on 20 August 2019).

16. Archambault, E., Amyot, D., Deschamps, P., Nicol, A., Rebout, L. and Roberge, G., Proportion of open access peer-reviewed papers at the European and world levels 2004–2011. European Commission, Montreal, Canada, 2013; http://science-metrix.com/pdf/SM_EC_OA_Availability_2004-2011.pdf (accessed on 25 June 2019).
17. Archambault, É., Amyot, D., Deschamps, P., Nicol, A., Provencher, F., Rebout, L. and Roberge, G., Proportion of open access papers published in peer-reviewed journals at the European and world levels – 1996–2013. European Commission, 2014; http://science-metrix.com/sites/default/files/science-metrix/publications/d_1.8_sm_ec_dg-rtd_proportion_oa_1996-2013_v11.pdf (accessed on 25 June 2019).
18. Björk, B. C., Welling, P., Laakso, M., Majlender, P., Hedlund, T. and Guðnason, G., Open access to the scientific journal literature: situation 2009. *PLoS ONE*, 2010, **5**(6), e11273. https://doi.org/10.1371/journal.pone.0011273
19. Björk, B. C., The hybrid model for open access publication of scholarly articles: a failed experiment? *J. Am. Soc. Inf. Sci. Technol.*, 2012, **63**(8), 1496–1504.
20. Björk, B. C., Laakso, M., Welling, P. and Paetau, P., Anatomy of green open access. *J. Assoc. Inf. Sci. Technol.*, 2014, **65**(2), 237–250.
21. Björk, B., Hybrid open access – a longitudinal study. *J. Informetr.*, 2016, **10**(4), 919; https://doi.org/10.1016/j.joi.2016.08.002
22. Björk, B.-C., The open access movement at a crossroad: are the big publishers and academic social media taking over? *Learn. Publish.*, 2016, **29**(2), 131–134; https://doi.org/10.1002/leap.1021
23. Björk, B. C., Gold, green, and black open access. *Learn. Publish.*, 2017, **30**, 173–175; doi:10.1002/leap.1096
24. Piwowar, H. *et al.*, The state of OA: a large-scale analysis of the prevalence and impact of open access articles. *Peer J.*, 2018, **6**, e4375; https://doi.org/10.7287/peerj.preprints.3119v1 (accessed on 20 August 2019).
25. Bosman, J. and Kramer, B., Open access levels: a quantitative exploration using Web of Science and oaDOI data (No. e3520v1). *Peer J. PrePrints*, 2018.
26. https://apps.webofknowledge.com/ (accessed on 20 June 2019).
27. https://images.webofknowledge.com/images/help/WOS/hs_wos_fieldtags.html (accessed on 20 June 2019).
28. http://unpaywall.org/products/api (accessed on 23 June 2019).
29. Rupika, Uddin, A. and Singh, V. K., Measuring the university–industry–government collaboration in Indian research output. *Curr. Sci.*, 2016, **110**(10), 1904.

Received 23 August 2019; revised accepted 21 September 2019

doi: 10.18520/cs/v117/i9/1435-1440