

Measuring the Open Access Friendliness of Indian Institutions through Data Carpentry

Abhijit Roy and Parthasarathi Mukhopadhyay

The present study has investigated the open access friendliness of top Indian academic institutions that have been ranked in the National Institutional Ranking Framework (NIRF) 2020. Open access friendliness has been measured based on two areas: i) the OA publication area and ii) the use of a license in OA publications. The present study has analyzed, over the last ten years, a total of 6,04,586 primary publication data (2010–2019). The data carpentry tool, named OpenRefine, has fetched the open access status of the institutional publications. The indicator assessed the selected institutions' open access friendliness. The final OAF ranking shows that 36.63% (n=37) of institutions have an OAF score above 50 and OAF scores ranges from 36.51 at the lowest to 74.99 at the highest. The overall OA share for top 100 institutes in India in the given publication time frame is 23.73%.

Introduction

The adaptations of the open access policy changed scholarly communications worldwide, and over the last decade, peer-review journals migrated from a printed to digital format (Björk, 2012; Piwowar et al., 2019). More than half of European and global scientific peer-reviewed papers are available on the Internet and have a citation advantage (Archambault et al., 2013; Archambault et al., 2014). India has also adopted the open access policy and gotten an advantage in scholarly impact. In 2004, the Indian National Academy signed the Berlin Declaration. After that, other institutions adopted the same and at present, 23-24% of publications have been distributed through different open access routes by Indian academic and research institutions (Nazim, 2021; Singh et al., 2020). However, it should be noted that appropriate indicators must be designed to build awareness and materialize the open science (OS) culture. Some researchers have attempted to measure open access indicators based on open access colours, open access repositories, and citations at the global and regional levels (Alperin et al., 2014; Gomez et al., 2009; Maddi, 2019, 2020; Robinson-Garcia et al., 2019). A team of researchers (Robinson-Garcia et al., 2019, 2020) have designed open access indicators at the university level and analyzed the open access uptake worldwide (including in India) based on Unpaywall and the Web of Science. They have found out that worldwide universities' openly share publications at 43%, universities in the UK have a total OA share of 74%, and India has the nineteenth-highest OA share worldwide. A scientometric study conducted by Mukhopadhyay (2022), has analyzed the open access friendliness of the Indian Institutes of Technology (IITs) based on a 100-point OAFI scale and found that newly established IITs have adopted more open access friendliness than older IITs. Roy and Mukhopadhyay conducted another study on central universities of India in 2022. They have found that, out of the top central universities, Vishya Bharati University has the most open access friendliness.

Keeping in view the scenarios as discussed, the present study has measured the open access friendliness of Indian top institutions that are ranked in the National Institutional Ranking Framework (NIRF) 2020 overall top 100 categories. The open access friendliness indicator as developed by (Mukhopadhyay, 2022) includes four major areas: i) OA share, ii) licensing share in OA publications, iii) OA citation impact, and iv) OA altmetric impact. This study examined the open access friendliness of the top Indian institutions based on only two areas: (i) open access publication sharing at the institution level and (ii) the share of OA publications with a legal licensing. These two areas are subdivided into five factors with a total 100-point weightage scale.

2. Overview of the Methodology

This section includes two groups – a) steps and processes related to data fetching, data organization and data extraction; and b) feeding data into a selected OAFI scale modified to suite the needs of the present study.

2.1 Data Carpentry Steps

The steps related data source identification to data xtraction are as follows -

2.1.1 Selection of Institutions

The aim of the present study is to develop an open-access-friendliness ranking framework for the top Indian higher education institutions (HEIs). The sample of the research study includes all of the top 100 institutions that have been ranked in the NIRF 2020. A total of 101 institutions were included in the overall ranking of NIRF 2020 (a combined list of seven subject domains) (National Institutional Ranking Framework, 2020).

2.1.2 Development of primary dataset

Based on a suitable query that included the affiliation ID [AF-ID (eight-digit ID number)] of the institutions, a total of 6,04,586 records of primary publication data for these top 100 institutions were collected from the largest bibliographic and citation database, Scopus. The present study has collected only institutional affiliation publications (2010–2019) in the CSV file. In the last ten years, primary publication data (all types of documents) has been merged into a single CSV file by using a suitable script and then imported to a data carpentry tool, OpenRefine. Out of these primary publications, 13.51% (n = 81,687) of documents have been published without DOI. In Table 1, we have listed the 17 Indian institutions whose primary publications are above 10,000. The publications without DOI have been excluded from farther analysis as the next steps are all based on DOI as unique input element.

Table 1: Primary dataset of top 17 Indian institutions with 10K+ publications (2011-2019)

SL	Name of the institutions	Total publications (2010-2019)	Publications With DOI	Rank in NIRF 2020
1	Indian Institute of Science	25,761	23,856	2
2	Indian Institute of Technology Kharagpur	22,365	20,889	5
3	Indian Institute of Technology Bombay	21,179	19,378	4
4	Indian Institute of Technology Madras	20,891	18,989	1
5	Indian Institute of Technology Delhi	19,930	18,185	3
6	University of Delhi	17,952	16,143	18
7	Indian Institute of Technology Roorkee	16,000	14,713	9
8	Jadavpur University	16,226	14,651	12
9	Vellore Institute of Technology	19,431	14,334	28
10	Anna University	18,059	14,015	20
11	Indian Institute of Technology Kanpur	13,199	12,296	6
12	Banaras Hindu University	13,514	11,543	10
13	Indian Institute of Technology Guwahati	11,975	11,384	7
14	Panjab University	10,145	9,324	44
15	University of Calcutta	10,303	9,240	11
16	Aligarh Muslim University	10,575	9,224	31
17	Manipal Academy of Higher Education	10,316	8,503	14

2.1.3 Development of secondary dataset

The DOI element in the primary publications data set has a very important role for developing the secondary dataset. The OpenRefine data carpentry tool has collected the open access status of publications from Unpaywall through a data wrangling process. Table 2 shows that a total of 5,11,916 primary publications with open access status have been responded to by Unpaywall, against 5,22,899 publications with DOI. Only 2.10% (n=10,983) open access status of primary publications were not given by Unpaywall.

Table 2: Unpaywall responded to publications status

REST/API call structure for OA status	No. of queries sent	Responses received
“https://api.unpaywall.org/v2/” + value + ?email=<your-mail-id-here>”	5,22,899	5,11,916
value is DOI	Publications with DOI	97.90% of publications with DOI

2.1.4 Data Extraction

Unpaywall provides us with publication status in JSON (JavaScript Object Notation) format. In this final step, the data carpentry tool has extracted the essential data using some GREL syntax. A set of examples is given in Table 3.

Table 3: GREL -an example to extract data from the Unpaywall dataset

Response from Unpaywall in JSON	GREL for data extraction	Extracted data
"is_paratext": false, "published_date": "2015-01-01", "year": 2015, "journal_name": "Journal of Epidemiology and Global Health", "journal_issns": "2210-6014", "journal_issn_l": "2210-6006", "journal_is_oa": true, "journal_is_in_doaj": true, "publisher": "Atlantis Press", "is_oa": true, "oa_status": "gold",	value.parseJson().is_oa	True
	value.parseJson().journal_name	Journal
	value.parseJson().journal_is_oa	OA journal
	value.parseJson().oa_status	True
	value.parseJson().is_oa	True

2.2 Indicators for Open Access Friendliness (OAF)

The open access friendliness indicators are based on two areas (open access publications and use of license in OA publications), distributed on a 100-point scale. These categories were proposed by Mukhopadhyay in 2022. This study has divided these two main areas into a total of five factors with their respective weightage (Table 4).

Table 4: Areas, groups and factors for calculating Open Access Friendliness Indicators

Areas (weightage)	Groups within the areas (with distributed weightage)		
OA publications Areas (weightage: 60%)	OA share (Group weightage:30%)	Gold & Green share (Group weightage: 15%)	Repository share (Group weightage:15%)
OA licensing Areas (weightage: 40%)	OA license share (Group weightage:20%)	Gold and Green license share (Group weightage: 20%)	

The scope of these five OA factors under the two major areas are as follows-

2.2.1 OA Share

OA share means the total number of OA publications in the last ten 2011-2019 years. Similarly, total publications means the number of publications of a given institution that have received status (close or open) from Unpaywall. The value is obtained through the formula- (Number of total OA ÷ Number of total publications) *30.

2.2.2 Gold & Green Share

Gold and green share means the total numbers of gold and green OA publications during the period 2011-2019. The two routes (Gold & Green) are considered in this study because these two routes are more legalistic than the other two routes (Bronze & Hybrid). The value is obtained through the formula- (Number of Gold & Green OA share ÷ Number of total OA publications) *15.

2.2.3 Repository Share

The natural destination of a green OA publication is an open access repository (OAR). An OAR is considered the best OA location for a green OA publication. This group value quantifies the availability of green OA publications in repositories, as the best OA locations, across the globe including archiving in their institutional repositories. The indicator is calculated by the formula- (Number of green OA in institutional repositories ÷ Total numbers of green OA publications) *15.

2.2.4 OA License Share

OA license share means the use of the appropriate OA licensing (mainly Creative commons) in OA publications. It is another important issue for measuring the open access friendliness of OA publications. The users' rights depend on the attached license of an OA publication. This factor is calculated by the formula- (Total number of OA published with a license ÷ Total numbers of OA publications) *20.

2.2.5 Gold & Green License Share

Similarly, Gold and Green license share means the total number of Gold and Green OA published with a license in the last ten years (2010-2019). The value is obtained through the formula- (Total number of Gold & Green OA with a license ÷ Total number of Gold & Green OA publications) * 20.

3. Results

An analysis of JSON merged datasets (5,22,899 publications with DOI out of which 5,11,916 publications responded to by Unpaywall) shows that 3,90,447 publications (76.27%) are closed access and only 1,21,469 have been published in open access (23.73%). These 1,21,469 OA publications have been distributed into four routes: -Gold OA (58,777 i.e., 48.39% of total OA publications); Green OA (32,690 i.e., 26.91% - of total OA publications); Bronze OA (17,599 i.e., 14.49% of total OA publications); and Hybrid OA (12,403 i.e., 10.21% of total OA publications). The last ten years of OA publication growth is presented in Figure 1. In the licencing scenario, a total of 48.98% (n=5,9,497) of OA was published with legal licencing in the last ten years.

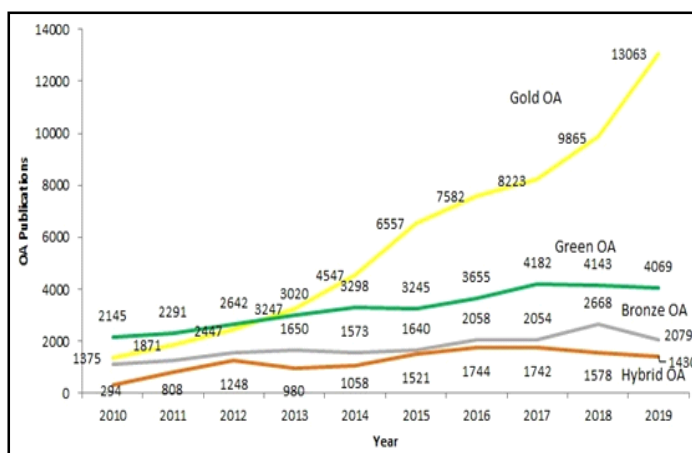


Figure 1: Growth of the OA types in Indian top institutions in the years (2010-2019)

3.1 The OAF Score of overall institutions

The OAF score is the sum of Factor I (OA share), Factor II (Gold & Green share), Factor III (Repository share), Factor IV (OA license share) and Factor V (Gold & Green license share). These show that the newly established university JSS Academy of Higher Education and Research (2008), has achieved the highest OAF score 74.99 (out of 100) in overall categories. On the other hand, the oldest established university - university of Calcutta (1957) got 46.52 (out of 100) and achieved the 57 OAF rank in overall categories. Out of 101 institutions, Thapar Institute of Engineering & Technology and Indian Institute of Technology Roorkee both got 40.50 scores (out of 100) and achieved 93 OAF rank overall categories. The Koneru Lakshmaiah Education Foundation University has secured the last position out of 101 institutions. The OAF has scored 36.51 marks (out of 100). In overall categories out of 101 institutions, only 11.88% (n=12) institutions have secured OAF scores above 60 (out of 100) in Table 5.

Table 5: Ranked list of 12 top institutions by OAFI

Name of Institutions and NIRF ranking 2020	Factor I (30)	Factor II (15)	Factor III (15)	Factor IV (20)	Factor V (20)	OAF (100)	OAF Rank
JSS Academy of Higher Education and Research (54)	18.82	12.55	10.73	16.12	16.78	74.99	1
Sawai Man Singh Medical College (91)	18.46	13.02	12.27	13.37	14.77	71.88	2
Sri Ramachandra Institute of Higher Education And Research (51)	17.92	11.82	11.11	14.44	16.10	71.39	3
Datta Meghe Institute of Medical Sciences (97)	17.52	12.76	11.79	13.30	14.86	70.22	4

Name of Institutions and NIRF ranking 2020	Factor I (30)	Factor II (15)	Factor III (15)	Factor IV (20)	Factor V (20)	OAF (100)	OAF Rank
Mysore University (47)	11.67	12.65	12.62	13.94	14.81	65.69	5
Manipal Academy of Higher Education (14)	14.33	12.06	10.78	13.37	14.91	65.45	6
University of Jammu (90)	12.84	9.36	12.85	14.83	14.77	64.64	7
Indian Institute of Science Education & Research Pune (25)	15.56	11.29	13.86	11.85	11.65	64.19	8
King George's Medical University (50)	17.29	11.90	11.98	10.50	12.20	63.86	9
Dr. D. Y. Patil Vidyapeeth (75)	17.59	10.02	10.02	10.88	13.79	62.31	10
University of Madras (41)	9.03	12.34	10.74	13.74	15.26	61.11	11
Visva Bharati (69)	10.84	9.91	13.07	13.71	13.42	60.94	12

In overall categories out of 101 institutions, Bharath Institute of Higher Education & Research has secured the top position 19.56 (out of 30) in Factor I (OA share), whereas the Indian Institute of Technology Mandi has achieved the first position 13.19 (out of 15) in Factor II (Gold & Green share), Indian Institute of Science Education & Research Pune has ranked the first position 13.86 (out of 15) in Factor III (Repository share) and the JSS Academy of Higher Education and Research has ranked the top position in two Factors (Factors IV: OA license share and Factor V: Gold & Green license share). In the final rank list of 101 institutions are under the study, Sawai Man Singh Medical College got a score of 43.74 (out of 60) in Area 1 (OA publications Areas) and the JSS Academy of Higher Education and Research got 32.90 (out of 40) and secured the first position in Area II (OA licensing Areas). If we deeply observe Table 6, we find that out of those 101 institutions 63.37% (n=64) rank below fifty.

Table 6: OAFI scores of top 100 institutions

Overall OAFI score	Number of the institutions	Percentage
74.99 - 70.00	4	3.96%
69.99 - 60.00	8	7.92%
59.99 - 50.00	25	24.75%
49.99 - 40.00	59	58.42%
39.99 - 36.00	5	4.95%

3.2 The OAF score of Engineering and technological institutions

Of those 101 institutions, only 37 Engineering and technological institutions (IITs, NITs etc.) have got rank in overall categories. Out of 37 Eng. and Tech. institutions the newly established IIT Bhubaneswar has secured the first position, the OAF score is 57.52 (out of 100). On the other side, the National Institute of Technology Silchar has got 38.80 (out of 100) and got the last position. Of these 37 Eng. and Tech. institutions, Thapar Institute of Engineering & Technology and Indian Institute of Technology Roorkee both have got 40.52 scores and achieved 32nd positions in these categories. In the final ranking in this group, the Indian Institute of Technology Gandhinagar has got 34.45 (out of 60) and achieved the first position in Area I (OA publications Areas), while the Indian Institute of Technology Bhubaneswar has got 28.01 (out of 100) and achieved first positions in Area II (OA licensing Areas). If we deeply observe Table 7, we find out that only 8.11% (n=3) Eng. and Tech. institutions' OAF scores above fifty whereas 91.89% (n=34) Eng. and Tech. institutions OAF score below fifty.

Table 7: OAFI score of 37 Engineering and Technological Institutions

OAFI score	Number of Eng. and Tech. institutions	Percentage
57.52 - 50.00	3	8.11%
49.99 - 40.00	31	83.78%
39.99 - 36.00	3	8.11%

3.3 The OAF score of universities

In the NIRF 2020 overall categories, out of the top hundred institutions 52 universities secured rankings in the range of 2-100. Out of these 52 universities, the newly established university JSS Academy of Higher Education and Research got 74.99 (out of 100) in the OAF score card and secured the first position. On the other hand, Koneru Lakshmaiah Education Foundation University got the last position. The OAF score was 36.51 (out of 100). In this category, the JSS Academy of Higher Education and Research got the first position in both areas – Area I (OA publications Area) and Area II (OA licensing Area). Table 8 displays that 46.15% (n=24) of universities are having OAF scores above fifty.

Table 8: OAFI score of 52 Universities

OAFI scores	Number of universities	Percentage
74.99 - 70.00	1	1.92%
69.99 - 60.00	5	9.62%
59.99 - 50.00	18	34.62%
49.99 - 40.00	26	50.00%
39.99 - 36.00	2	3.85%

3.4 The OAF score of health institutions

Out of 101 institutions, seven institutions are attached with the health institutions. Of these seven institutions, Sawai Man Singh Medical College has achieved the first position with OAF score 71.88 (out of 100). Table 9 displays the OAF scores of the health institutions. Interestingly, 42.86% (n=3) of institutions have secured scores above seventy, 28.57% of institutions have scores above sixty, and 14% of institutions have OAF scores above fifty. Only one of the health institutions OAF scores was below fifty.

Table 9: OAFI score of 7 Health institutions

OAFI scores	Number of health institutions	Percentage
71.88 - 70.00	3	42.86%
69.99 - 60.00	2	28.57%
59.99 - 50.00	1	14.29%
49.99 - 40.00	1	14.29%
39.99 - 36.00	0	0.00%

3.5 The OAF score of other institutions

In this group, a total of five institutions are included in the OAF ranking. The Indian Institute of Science Education & Research, Pune has achieved the first position in this group with an OAF score of 64.19 (out of 100). If we observe the OAF scores in Table 10, we find out that 80% (n=4) of other institutions got OAF ranked above fifty.

Table 10: OAFI scores of 5 other institutions

OAFI scores	Number of Institutions	Percentage
64.19-60.00	1	20.00%
59.99-50.00	3	60.00%
49.99-40.00	1	20.00%
39.99-36.00	0	0.00%

4 Conclusions

The present study has investigated the open access friendliness of the Indian top institutions based on a 100-point weightage scale in two areas (OA publication area and use of license in OA publications). These two areas are subdivided into a total of five factors to design the 100-point OAF scale.

Out of the 101 institutions, the OAFI scores show that the newly established institutions have scored better in open access friendliness. If we observe deeply the OAF ranking, we find that in case of Eng. & Tech.

Institutions, 8.11% (n=3, out of 37 institutions), for universities 46.15% (n=24, out of 52), for health institutions 87.71% (n=6, out of 7) and in case of other institutions 80.00% (n=4, out of 5) have ranked above 50 in 100-point scale of the OAF ranking. Of these top Indian institutions, only 3.96% (n=4) have achieved an OAF score of above 70. Two of these three belong to health institutional group, while only one belongs to university group. In the Eng. & Tech. group, 91.89% (n = 34, out of 37) institutions have scored below fifty in the OAF ranking. The possible reason for this surprising finding may be the fact that newly founded institutes right from beginning of their academic journey adopted the culture of OA as the advocacy of OA began in India right from 2006 (Harnard, 2008). The older institutes carry lots of legacy (like the University of Calcutta) and were late adopters of OA philosophy. This fact of late adaptation of OA principles are quite evident in their publication's patterns since 2010.

References

1. Alperin, J. P., Babini, D. S. de, Fischman, G., & Consejo Latinoamericano de Ciencias Sociales. (2014). Open access indicators and scholarly communications in Latin America. Available at <http://biblioteca.clacso.edu.ar/clacso/se/20140917054406/OpenAccess.pdf> (Accessed on 28/07/2022)
2. Archambault, É., Amyot, D., Deschamps, P., Nicol, A., Provencher, F., Rebut, L., & Roberge, G. (2014). Proportion of Open Access Papers Published in Peer-Reviewed Journals at the European and World Levels—1996–2013. Available at <https://digitalcommons.unl.edu/scholcom/8> (Accessed on 28/07/2022)
3. Archambault, E., Amyot, D., Deschamps, P., Nicol, A., Rebut, L., & Roberge, G. (2013). Proportion of Open Access Peer-Reviewed Papers at the European and World Levels—2004-2011. *Eur. Comm.* Available at http://www.science-metrix.com/pdf/SM_EC_OA_Availability_2004-2011.pdf (Accessed on 28/07/2022)
4. Björk, B.-C. (2012). The hybrid model for open access publication of scholarly articles: A failed experiment? *Journal of the American Society for Information Science and Technology*, 63(8), 1496–1504. doi: 10.1002/asi.22709
5. Gomez, N., Bustos-Gonzalez, A., Santillan-Aldana, J., & Arias, O. (2009). Open Access Indicators and Information Society: The Latin American Case. *OCLC Systems & Services: International Digital Library Perspectives*, 25(2), 82–92. doi: 10.1108/10650750910961884
6. Harnad, S. (2008). India, Open Access, the Law of Karma and the Golden Rule. *DESIDOC Bulletin of Information Technology*, 28(1), 35-40. URL: <http://publications.drdo.gov.in/ojs/index.php/djlit/article/view/250/102>. (Accessed 9 September 2008).
7. Maddi, A. (2019). Construction of a Normalized Open Access Indicator (NOAI). Available at <https://hal.archives-ouvertes.fr/hal-02328158/file/DT-CEPN-2019-08.pdf> (Accessed on 28/07/2022)
8. Maddi, A. (2020). Measuring open access publications: A novel normalized open access indicator. *Scientometrics*, 124(1), 379–398. doi: 10.1007/s11192-020-03470-0

9. Mukhopadhyay, P. (2022). How green is my Valley? Measuring open access friendliness of Indian Institutes of Technology (IITs) through data carpentry. In *Panorama of open access: Progress, practices & prospects* (pp. 67–89). Ess Ess. doi: 10.5281/zenodo.6511080
10. National Institutional Ranking Framework. (2020). Department of Higher Education Ministry of Education Government of India. Available https://www.nirfindia.org/nirfpdfcdn/2020/pdf/Report/IR2020_Report.pdf (Accessed on 28/07/2022)
11. Nazim, M. (2021). Analysing Open Access Uptake by Academic and Research Institutions in India. *DESIDOC Journal of Library & Information Technology*, 41(02), 108–115.
doi: 10.14429/djlit.41.02.16324
12. Piwowar, H., Priem, J., & Orr, R. (2019). The Future of OA: A large-scale analysis projecting Open Access publication and readership. *BioRxiv*. doi: 10.1101/795310
13. Robinson-Garcia, N., Costas, R., & Leeuwen, T. N. van. (2020). Open Access uptake by universities worldwide. *PeerJ*, 8, e9410. doi: 10.7717/peerj.9410
14. Robinson-Garcia, N., Costas, R., & van Leeuwen, T. N. (2019). Indicators of Open Access for universities. *ArXiv:1906.03840 [Cs]*. Available at <https://arxiv.org/ftp/arxiv/papers/1906/1906.03840.pdf> (Accessed on 28/07/2022)
15. Roy, A., & Mukhopadhyay, P. (2022). Measuring Open Access Friendliness of Indian Central Universities through Data Carpentry. *SRELS Journal of Information Management*, 59(3), 131–139. doi: 10.17821/srels/2022/v59i3/170100
16. Singh, V. K., Piryani, R., & Srichandan, S. S. (2020). The case of significant variations in gold–green and black open access: Evidence from Indian research output. *Scientometrics*, 124(1), 515–531. doi: 10.1007/s11192-020-03472-y

Keywords: Open Access Friendliness; Open Access Indicators; Open Access; OpenRefine; Unpaywall; National Institutional Ranking Framework (NIRF)

About Authors

Mr. Abhijit Roy

M.Phil Scholar

Department of Library and Information Science, University of Kalyani, Kalyani, Nadia, West Bengal

Email: royabhijit75@gmail.com

Prof. Parthasarathi Mukhopadhyay

Professor

Department of Library and Information Science, University of Kalyani, Kalyani, Nadia, West Bengal

Email: psmukhopadhyay@gmail.com