Annals of Library and Information Studies
Vol. 69, September 2022, pp. 225-237
DOI: 10.56042/alis.v69i3.63837

Measuring the open access friendliness of the state universities in India through data carpentry

Abhijit Roy[a] and Parthasarathi Mukhopadhyay[b]

[a]M. Phil Scholar, Department of Library and Information Science, University of Kalyani, Kalyani, Nadia, West Bengal-741235, India, Email: royabhijit75@gmail.com

[b]Professor, Department of Library and Information Science, University of Kalyani, Kalyani, Nadia, West Bengal-741235, India, Emai: psmukhopadhyay@gmail.com

Received: 16 June 2022; revised: 23 August 2022; accepted:02 September 2022

The present research aims to investigate the open access (OA) friendliness of the twenty-six (26) state universities that have been ranked in the National Institutional Ranking Framework's (2021) 'Overall Category' (Rank Band 1: Top 100). The open-access ranking framework, as designed, is based on four major OA areas: i) proportion of OA publications by a given state university; ii) use of a license in sharing OA publications; iii) citations received by OA publications; and iv) popularity of OA publications in socio-academic web-space as indicated by altmetric attention scores. Ten years (2012–2021) primary publication data for the 26 state universities from Scopus was collected. We analyzed a total of 1,57,971 publications; 1,60,723 citations received by these publications; and altmetric attention scores received by 26,996 publications from these 26 state universities listed in the NIRF 2021 overall category. It applies an open-source data wrangling tool (OpenRefine) and a number of ODbL-based (Open Database License) data sources to gather the necessary data elements required for this large-scale study. The study finds that almost 98% of the state universities enlisted in the top 100 overall category passed the 50 marks on a 100-point scale, and the newly established state universities' open access friendliness is greater than the older ones. The final open access friendliness ranking places King Georges Medical University at the top, with leading scores in areas I (share of OA publications) and III (OA citation share), while Panjab University has the highest value in Area II (use of license in OA publications) and Delhi Technological University has the highest value in Area IV (OA altmetric attention score).

Keywords: Open access, National Institutional Ranking Framework, State University, Unpaywall, OpenRefine

Introduction

The open access movement and the advancement of information and communication technology (ICT), particularly the Internet, have accelerated global open access to scholarly communications over the last 20 years[1,2]. India has also adopted the open access policy in 2000 and has got advantages from scientific impact and international collaboration[3,4]. India has achieved the 11th position in the world for OA publications, with more than 808K articles and a 31% share of the total publications in OA routes[5]. But to date, no research framework has been presented to measure the open access friendliness of the state universities. A few researchers have endeavoured to measure open access indicators at the global, regional, and country-level based on OA publications; OA resources covered by repositories; and journals listed in DOAJ from 2009 to 2020[6–11]. Some researchers explored OA publication types, OA versions, availability of repository copies, names of the repository institutions, publishers' detailed information, and future impact of the OA publication status on a large scale, which has been traced from Unpaywall[12,13]. The Unpaywall is a free bibliographic open access database developed by OurResearch (https://ourresearch.org/).

Robinson-Garcia and others, who formulated the OA indicators of 963 universities worldwide (including India) found different OA types at the institutional level, which show the highest OA contributions in different disciplines. The publication data for this study has been collected from the Web of Science and Unpaywall[14,15]. Another scientometric study conducted by Mukhopadhyay[3] has examined the open access friendliness of the Indian Institutes of Technology (IITs) and found that the newly established IITs are more open access friendly in comparison to the old IITs.

The current study aims to assess the open access friendliness of state universities using

Mukhopadhyay's open access friendliness ranking framework[3]. OA friendliness has been measured based on the four primary areas related to OA: 1) the proportion of publications that are open access; 2) the use of the Creative Commons license in the creation of open access publications; 3) citations to open access publications; and 4) the altmetric attention score for open access publications. Each of the primary areas is also subdivided further to pin-point the focus, like "Share of OA publications", which is further subdivided into three factors: the share of OA documents in total publications; the percentage of Green and Gold OA in total OA publications; and Green OA availability in institutional repositories as the best OA locations. The sample of the research study includes the 26 state universities; those are included in the National Institutional Ranking Framework (2021) of overall categories. The names and ranks of the universities are displayed in Table 1.

Review of literature

This study focuses on assessing the open access friendliness of the top-ranked state universities in India. Several studies have been conducted globally by researchers to measure open access status but with limited sets of data. For the first time worldwide, 930 universities' OA publications have been examined by Robinson-Garcia et al based on a combination of data from WoS, Unpaywall, and the Leiden ranking[14,15]. They analyzed a total of 46,21,721 publications during the period 2014–2017 and found that worldwide universities' OA share is 43%. The highest share of OA (74%) has been achieved by British universities, and India has achieved the 19th position in OA share. A large-scale study has been conducted by Martín-Martín et al. to examine the OA status of publications indexed by Google Scholar in all countries[16]. They have collected a total of 22,69,022 publications (with DOI) from WoS core collections for the period from 2009 to 2014. According to the study, 35.8% of OA at the global level was published on publishers' sites and repositories. They have also found that Indian institutions hold a total of 23.1% (15.7% by OA publisher and 7.4% in repositories) of the documents available in OA routes.

Bhat [17] studied 17,516 research articles belonging to Indian Institute of Science, All India Institute of Medical Sciences, Bhaba Atomic Research Centre,

Table 1 — Ranks of the state universities in NIRF (2017-2021, arranged by 2021 ranks)

Sl. no.	Name of the institutions (year of establishment)	Rank NIRF lists (2017-2021)				
		NIRF 2017	NIRF 2018	NIRF 2019	NIRF 2020	NIRF 2021
1	Calcutta University,1857	27	21	12	11	11
2	Jadavpur University,1905	12	13	13	12	14
3	Savitribai Phule Pune University, 1949	18	16	17	19	20
4	Bharathiar University, 1982	45	20	21	21	22
5	Anna University, 1978	13	10	14	20	25
6	Mysore University, 1916	57	NL	80	47	34
7	Panjab University, 1882	54	33	34	44	38
8	Kerala University, 1937	47	47	35	42	43
9	University of Madras, 1857	64	29	33	41	47
10	Andhra University, 1926	69	36	29	36	48
11	Mahatma Gandhi University, 1983	NL	52	49	49	52
12	Delhi Technological University,1941	NL	100	71	62	54
13	Alagappa University,1985	NL	43	47	64	57
14	King George`s Medical University, 1905	NL	23	42	50	60
15	Osmania University,1918	38	45	43	53	62
16	Gujarat University, 1949	NL	NL	NL	60	62
17	Cochin University of Science and Technology, 1971	NL	99	94	89	65
18	Gauhati University, 1948	43	61	65	72	67
19	University of Kashmir, 1948	NL	71	79	78	76
20	Madurai Kamaraj University,1966	NL	81	69	84	83
21	Guru Nanak Dev University,1969	NL	86	81	88	85
22	Bharathidasan University,1982	NL	94	86	77	90
23	Sri Venkateswara University, 1954	68	74	72	68	92
24	Calicut University, 1968	93	NL	90	76	95
25	Mumbai University,1857	NL	NL	NL	95	96
26	Guru Gobind Singh Indraprastha University,1998	NL	NL	95	NL	99

(NL – Not Listed)

Indian Institute of Technology, Delhi and Indian Institute of Technology, Kharagpur based on Scopus data for the period 2003-2007 and found a low percentage of articles published in open-access journals. Out of these five institutions, the All India Institute of Medical Sciences (AIIMS), New Delhi, is the most OA-friendly institute with an OA publication share of 19.37%. Piryani & Dua[18] have examined open access levels and patterns of scholarly publications in India during the period 2014-2018. The authors have collected a total of 3,77,336 publications from WoS. They have fetched the OA publication status from Unpaywall by using REST/API-based content negotiation and found that 24% of research papers are distributed in OA routes (Gold OA, Green OA, Hybrid OA, and Bronze OA).

Nazim[19] has analyzed the open access uptake of Indian academic and research institutions based on the CWTS Leiden Ranking database, 2020. A total of 36 universities have been ranked in this ranking system in 2020. Out of these, the Indian Institute of Science (IISc) is the most productive institution in terms of OA. This study also shows that 23% of those institutions' publications are distributed through OA routes, with Green OA taking the first place.

An open access friendliness indicator (OAFI) has recently been proposed by Mukhopadhyay, where he examined the open access friendliness of the Indian Institutes of Technology (IITs)[3]. The ranking framework has four primary areas (OA publications share, OA license share, OA citations share, and OA altmetric share) with a 100-point weightage scale. This framework has tested large publications (1,59,107), citation (21,69,395), and altmetric (24,308) data of 16 IITs (listed in the top 100 overall category of NIRF 2020) and found that newly established IITs' OAFI scores are higher than the older ones.

Roy & Mukhopadhyay[20] have examined the open access friendliness of central universities in India based on two primary areas–OA publication share and OA license share. This research has surveyed ten central universities (listed in NIRF 2021 overall category) with a total of 85,916 publications and reported that 28.43% of publications have been distributed via different OA routes (Gold OA, Green OA, Hybrid OA, and Bronze OA). It is also found that, out of 10 universities, Visva Bharati University has achieved the first position with an OAFI score of 64.04 (out of 100). The current review of the literature makes it abundantly clear that numerous studies have been conducted in relation to open access publications on a global and regional scale, but only two studies have looked at the open access friendliness of Indian institutions, and no one has yet explored the open access friendliness of the country's top ranked state universities.

Objectives of the study

- To apply data carpentry methods to collect OA status and other associated data (citations, altmetric score, best OA location, OA licensing etc.) for the publications of the state universities in India listed in the top 100 overall category of the NIRF 2021 for the period from 2012 to 2021;
- To find the growth of OA contributions within the respective categories (Gold OA, Green OA, Hybrid OA, and Bronze OA) in the last ten years (2012–2021) of the 26 state universities; and
- To implement a proposed OAFI framework [3] for the ranking of the state universities in India according to their OA friendliness.

Methodology

The present study explores the open access friendliness of state universities that have obtained a rank in the National Institutional Ranking Framework (2021) in the overall category. The NIRF (2021) was published by the Ministry of Education (Government of India) on September 9th, 2021. The overall category of the NIRF is a combined list of seven subject domains. Those are Engineering, Management, Pharmacy, Law, Medical, Architecture, and Dental. The overall category has been divided into three bands: rank band I (1-100), rank band II (101-150), and rank band III (151-200).

A total of 442 state universities exist in India in the states and union tertiaries as approved by UGC, India (https://www.ugc.ac.in/stateuniversity.aspx), but only 26 state universities have been included in the top 100 NIRF 2021 categories, with ranks ranging from 11 to 99. These state universities (arranged by respective ranks) are: Calcutta University (11); Jadavpur University (14); Savitribai Phule Pune University (20); Bharathiar University (22); Anna University (25); Mysore University (34); Panjab University (38); Kerala University (44); University of Madras (47); Andhra University (48); Mahatma Gandhi University (52); Delhi Technological University (55); Alagappa University (57); King George`s Medical University (60); Osmania University (62); Gujarat University (63); Cochin University of Science and Technology (65); Gauhati University (67); University of Kashmir

(76); Madurai Kamaraj University (83); Guru Nanak Dev University (85); Bharathidasan University (91); Sri Venkateswara University (92); Calicut University (95); Mumbai University (96); and Guru Gobind Singh Indraprastha University (99).

Development of primary datasets

The primary publications of the 26 state universities (last ten years, 2012–2021) have been collected from Scopus by using the '**Affiliation Search'** facility. The search results have been downloaded year-wise (like 2011) as 'CSV' files for each of the selected state universities on January 31, 2022. Each of these CSV files has been amalgamated into a single CSV file by using a suitable script. The summary of the data is given in Table 2. The publications with DOI play an important role in developing the secondary dataset for further data wrangling.

Development of secondary dataset

The secondary dataset is developed by using REST/API-based content negotiation methods. A set of three ODbL data sources were selected in view of their REST/API support, extent of items of information, and capability to provide responses in JSON format (Java Script Object Notation – a light-weight online data exchange format) against API calls. These three ODbL-based data sources are: Unpaywall (to fetch the OA status of a publication); Dimensions (to get the citation status of a given publication); and Altmetric.com (to obtain the altmetric attention score of a given publication). All of these data sources can provide responses in JSON format to DOI-based API calls. OpenRefine, an open-source data carpentry tool, is configured to manage the content negotiation processes with the selected ODbL-based data sources for developing the secondary dataset by deploying GREL (General Refine Expression Language) based API call syntax and extracting information from JSON-formatted responses.

Open access data: Unpaywall is a storehouse of open access content (3,12,78,623+ records as of February 16th, 2022 and counting) that has been developed by "OurResearch" (https://unpaywall.org/). It offers open access content from 50,000 publishers and allows 1,00,000 API calls per day freely via REST/API. The REST/API based data fetching shows that a total of 98.40% of the publications of the 26 state universities are covered by the Unpaywall (1,38,158 publications out of 1,40,407 total publications).

Citation data: Dimensions (dimensions.ai) is a profit-making agency but allows access to publication metrics through REST/API calls. It was developed by

Table 2 — Primary dataset (arranged by number of publications with DOI in descending order)

SL.	Name of state universities	Total publications(2012-2021)	Publications with DOI	NIRF Rank 2021
1	Jadavpur University	17,743	16,504	14
2	Anna University	19,511	16,048	25
3	Panjab University	11,765	11,072	38
4	Calcutta University	11,771	10,885	11
5	Savitribai Phule Pune University	8,464	7,784	20
6	Bharathiar University	8,141	6,820	22
7	Delhi Technological University	6,572	6,195	54
8	University of Madras	5,617	4,978	47
9	Guru Nanak Dev University	5,227	4,917	85
10	King George`s Medical University	4,939	4,559	60
11	Bharathidasan University	5,136	4,497	90
12	Osmania University	5,249	4,450	62
13	Cochin University of Science and Technology	4,476	4,106	65
14	Andhra University	4,734	3,717	48
15	Mysore University	4,240	3,614	34
16	University of Kashmir	3,888	3,481	76
17	Alagappa University	3,957	3,462	57
18	Sri Venkateswara University	3,988	3,204	92
19	Madurai Kamaraj University	3,349	3,072	83
20	Gauhati University	3,505	3,006	67
21	Guru Gobind Singh Indraprastha University	3,237	2,946	99
22	Mumbai University	3,025	2,648	96
23	Mahatma Gandhi University	2,899	2,636	52
24	Kerala University	2,825	2,514	43
25	Calicut University	1,875	1,690	95
26	Gujarat University	1,838	1,602	62
	Grand total of publications	**1,57,971**	**1,40,407**	

Digital Science (https://www.digital-science.com/products/). A total of 1,40,407 quarries have been sent through OpenRefine, out of which 98.21% (1,37,897) of the publication's citation status has been received from Dimensions.

Altmetric data: A researcher can get altmetric attention scores for publications (with DOI) free of charge through REST/API calls from Altmetric.com (an API key is required but available through the application). The JSON-based response not only provides altmetric attention scores but also includes other socio-academic information from Twitter, Facebook, blog postings, Wikipedia, Mendeley, CiteULike, Connotea, and so on. A total of 19.25% of publications (27,030 out of 1,40,407 publications) have received an altmetric attention score.

Data extraction

The value of data comes through its extraction for the purpose of generating required information. This study has developed a set of GRELs in OpenRefine for the extraction of different items of information related to the OA status of a given publication. A few examples related to the deployment of GRELs in extracting information from JSON responses are given in Table 3.

Generic scenarios

An analysis of the extracted JSON data from the merged dataset (a total of 1,38,158 responses have been received from Unpaywall against a total of 1,57,971 publications) shows that 75.43% publications (1,04, 213 out of 1,38,158) are "closed access" and only 24.57% are "open access" publications (33,945 out of 1,38,158). These 33,945 OA publications are distributed among four OA routes: Gold OA (19,288 i.e., 56.82% of total OA contributions); Green OA (6, 040 i.e., 17.79% of total OA contributions); Bronze OA (5,231 i.e., 15.41% of total OA contributions); and Hybrid OA (3,386 i.e., 9.97% of total OA publications). In the last ten years, the highest number (14,302) of OA has been published in the year 2021. The lowest percentage (21.96%) of OA was in 2017 and the highest percentage (27.70%) of OA was in 2019.

Figure 1 shows that in the last ten years (2012–2021), gold OA publications have increased more

Table 3 — GREL- examples to extract data from Unpaywall's response

Response from Unpaywall in JSON	GREL for data extraction	Extracted data
{"doi":"10.48175/ijarsct-5853",..... "genre": "journal-article", "is_paratext": false, "published_date": "2022-07-22", "year": 2022, "journal_name": "International Journal of Advanced Research in Science, Communication and Technology", "journal_issns": "2581-9429", "journal_issn_l": "2581-9429", "**journal_is_oa**": true, "**journal_is_in_doaj**": false, "**publisher**": "Naksh Solutions", "**is_oa**": true, "**oa_status**": "gold", "**has_repository_copy**": false, ...}	value.parseJson().journal_is_oa	True
	value.parseJson().journal_is_in_doaj	False
	value.parseJson().publisher	Naksh Solutions
	value.parseJson().is_oa	True
	value.parseJson().oa_status	Gold
	value.parseJson().has_repository_copy	False

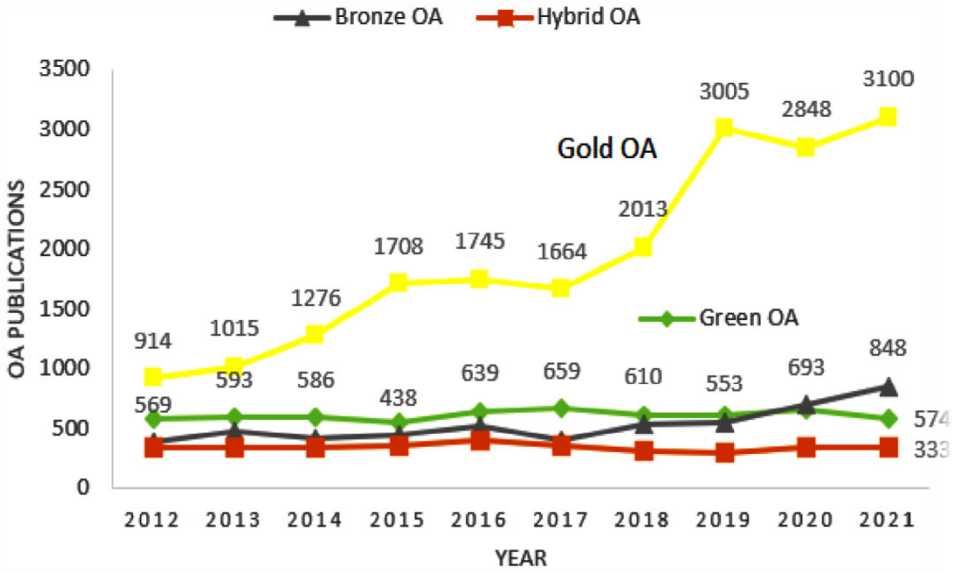

Fig. 1 — Growth of four OA types in 26 State Universities over the years (2012-2021)

than green OA, bronze OA, and hybrid OA. It is important to note that in these 26 state universities, a total of 19,580 OA publications are archived in different repositories worldwide (57.68% of total OA publications). These archived publications belong to different categories, like the Gold OA (52.22%); the Hybrid OA (61.34%); the Bronze OA (26.57%), and as expected for Green OA this ratio is 100% (Figure 2). An overview of the growth of OA publications in the state universities under study is reflected in Table 4.

Licensing is an important element in the OA-based scholarly communication process. Ideally, each OA publication should be supported by a suitable license to ensure the degrees of freedom associated with the publication. Unpaywall provides an item of information in response if licensing statement is available, e.g., *"license": "cc-by"*. The awareness of the licensing of OA publications in the 26 state universities is shown in Table 5.

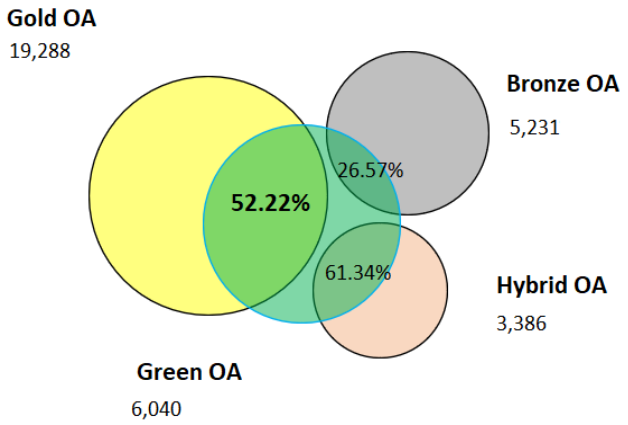

Fig. 2 — Numbers of OA in different repositories (2012-2021)

In total, 59.83% of OA publications (20,309 out of 33,945) have been published under OA licensing. The tabulated data shows that over the last ten years, the percentage of licenses used in total OA publications (54.86% in 2012, the lowest, to 66.04% in 2021, the highest) by those 26 top state universities has steadily increased. In the licensing scenario, 60.89% (12,367 of total licensed publications) have been published under the most liberal licensing routes like CC-BY (60.03%–12,191 contributions), CC-BY-SA (0.53%-108 publications), CC0 (0.18%–37 contributions) and Public Domain (0.15%–31 contributions), while the remaining (39.11%–7,942 of total licensing publications) have been published in less liberal licensing routes like CC-BY-NC-ND (20.18%-4,099 contributions), CC-BY-NC (7.64%-1,551 contributions), CC-BY-NC-SA (6.12%-1,243 contributions), Elsevier-Specific: OA user license (1.58%-318 publications), Publisher-specific license (1.54%-313 publications), ACS-Specific (1.13%-230 publication), Implied-OA (0.72%-146 contributions) and CC-BY-ND (0.21%-42 resources).

Table 4 — Growth of OA in 26 state universities

Year	Total publications by 26 state universities	Closed access Publications	OA publications	Open Access Categories				% OA (of total publications)
				Green OA	Gold OA	Bronze OA	Hybrid OA	
2012	9,477	7,277	2,200	569	914	377	340	23.21%
2013	10,277	7,874	2,403	593	1,015	465	330	23.38%
2014	11,553	8,952	2,601	586	1,276	409	330	22.51%
2015	12,176	9,127	3,049	549	1,708	438	354	25.04%
2016	13,219	9,919	3,300	639	1,745	515	401	24.96%
2017	13,984	10,913	3,071	659	1,664	402	346	21.96%
2018	15,274	11,805	3,469	610	2,013	531	315	22.71%
2019	16,087	11,631	4,456	604	3,005	553	294	27.70%
2020	16,954	12,413	4,541	657	2,848	693	343	26.78%
2021	19,157	14,302	4,855	574	3,100	848	333	25.34%

Table 5 — Growth of OA licenses in 26 State Universities

Year	Total OA publications by state universities	OA published without a license	OA published with a license	License categories		% of license in total OA publications
				Most liberal license	Less liberal license	
2012	2,200	993	1,207	732	475	54.86%
2013	2,403	1,052	1,351	878	473	56.22%
2014	2,601	1,043	1,558	1,054	504	59.90%
2015	3,049	1,239	1,810	1,144	666	59.36%
2016	3,300	1,291	2,009	1,091	918	60.88%
2017	3,071	1,090	1,981	999	982	64.51%
2018	3,469	1,277	2,192	1,259	933	63.19%
2019	4,456	2,252	2,204	1,331	873	49.46%
2020	4,541	1,750	2,791	1,728	1,063	61.46%
2021	4,855	1,649	3,206	2,151	1,055	66.04%

Gold OA

Figure 1 shows that for these 26 top state universities, most of the OA publications have been distributed along the gold OA route. It includes a total of 19,288 documents in 2,134 open access journals in six documentary formats like: journal-article (19,227 contributions, 99.68%), book-chapter (52 contributions, 0.27%), journal-issue (4 documents), journal (3 documents), monograph (1 document) and proceedings-article (1 document).

The top five open access journals for the state universities under study are: *Acta Crystallographica Section E* (663 contributions, 3.44%), *Scientific Reports* (572 contributions, 2.97%), *PLoS ONE* (491 publications), *Journal of High Energy Physics* (425 publications) and *International Journal of Innovative Technology and Exploring Engineering* (395 publications). A total of 72.1% gold OA publications have appeared in 1,617 DOAJ listed journals.

As for licensing, 80.72% (out of a total of 15,570 documents) of gold OA publications have been distributed under 11 types of licenses. The most popular Creative Commons licenses are CC-BY (9,662 contributions, 62.06% of total license publications) and CC-BY-NC-ND (3,423 contributions, 21.98%). The total gold OA articles of the 26 state universities have been distributed by 637 publishing agencies. The top five publication agencies are: Elsevier BV (3,250 contributions, 16.85%), Springer Science and Business Media LLC (2,305 contributions, 11.95%), MDPI AG (995 articles), International Union of Crystallography (IUCr) (913 articles), and IOP Publishing (877 articles).

The highest number of gold OA publications came from Panjab University (1,870 contributions, i.e., 9.70% of gold OA publications from 26 state universities) and the least from Calicut University (206 contributions). A total of 5,858 gold OA publications from 26 state universities have received Altmetric scores (30.37% of 19,288 total gold OA publications). The highest altmetric attention score (AAA score of 1152.635) was obtained by a gold OA publication from Bharathidasan University in the year 2013. The highest number of citations (727) in the gold OA category was received by a paper published in 2013 by the University of Madras in a DOAJ listed OA journal, namely the *International Journal of Biological Sciences,* which was published by Ivyspring International Publisher.

Green OA

The second largest number (6,040) of OA publications among the 26 state universities has been published via the Green OA route. The total green OA publications have been published in 1,581 'close access' journals in 8 documentary forms like journal-article (5,609 contributions– 92.86%), proceedings-article (224 contributions– 3.71%), book-chapter (16 contributions), book (22 contributions), monograph (11 contributions), reference-book (7 contributions), other (3 contributions) and posted-content (3 contributions). In the licensing scenario, 82.73% of articles have been published without any legal licensing. Only 17.27% of articles have been attached with 9 types of licensing, like CC-BY (391 documents), CC-BY-NC-SA (279 documents), CC-BY-NC-ND (190 documents), CC-BY-NC (96 documents), IMPLIED-OA (52 documents), CC0 (23 documents), PUBLIC DOMAIN (7 documents), CC-BY-ND (4 documents), CC-BY-SA (1 document).

The King George`s Medical University has the highest number of 813 (contributions of 13.46% of Green OA publications from 26 state universities) of the green OA productive institutions. The Panjab University has achieved the highest Altmetric attention score (AAA score of 1445.014) under Green OA publication. The highest number of citations (4997) secured by a Green OA publication from King George`s Medical University has been published in a close-access journal *Lancet* in 2015 and made available through a repository but without any formal license. The Unpaywall datasets provide institutional repository names under "best_oa_location".

An analysis of datasets shows that 91.37% (n=5,519) of Green OA has been presented in 302 different repository institutions worldwide. The top five institutional repositories are: Cornell University – arXiv (1,894 publications), PubMed Central-Europe PMC (1,739 publications), Research Square (86 publications), CERN-Zenodo (72 publications), Mysore University Library-MyPrints@UOM (66 publications). A study of the "repository_institutions" data shows that a total of 207 publications (out of 5,519 Green OA resources available through repositories—3.75%) are archived by institutional repositories of Indian origin. The top five repositories are attached to Mysore University Library (66 OA publications), Cochin University of Science and Technology (59 OA publications), Bangalore University (34 OA publications), CSIR-The National Physical Laboratory (29 OA publications), and the Indian Institute of Science (18 OA publications).

Bronze OA

The Bronze OA resources (5,331 in total, 15.41% of total OA publications) are distributed among 1,509

journals covering 284 publishers. The top five publications agencies among the top 26 state universities are: Springer Science and Business Media LLC (1,486 publications), Elsevier BV (727 publications), Wiley (261 publications), Oxford University Press (OUP) (255 publications) and Informa UK Limited (250 publications). The top five bronze OA publication journals are *SN Applied Sciences* (181 papers), *Bulletin of Materials Science* (142 papers), *Journal of Chemical Sciences* (142 papers), *Journal of Earth System Science* (141 papers), and *Asian Journal of Chemistry* (106 papers).

All the documents in the bronze OA publications were published without any licensing. The highest number of citations (6,665) of bronze OA publications was received in a paper in 2016 published by Savitribai Phule Pune University. Among all the bronze OA publications, 27.09% (out of 5,231 Bronze OA publications) have been published with an Altmetric Score. The highest altmetric attention score (AAA score of 1441.990) has been obtained in a paper published in a close-access journal, namely *Monthly Notices of the Royal Astronomical Society* in 2015 by Savitribai Phule Pune University.

Hybrid OA

The hybrid OA resources (3,386 contributions, 9.97% of total OA publications) are made available through 712 journals by 159 publishers. The top five favourite publication agencies for the universities under study are: the American Physical Society (APS) (748 contributions), Springer Science and Business Media LLC (565 contributions), Elsevier BV (537 contributions), Science Publications (158 contributions) and *Asian Journal of Chemistry* (107 contributions).

The top five favourite Hybrid OA publication journals are: *Physical Review D* (332 papers), *Physical Review Letters* (249 papers), *Physical Review C* (155 papers), *Journal of Computer Science* (112 papers) and *Asian Journal of Chemistry* (107 papers). On average, 92.56% (3,135 out of 3,386) of the publications of the 26 state universities in hybrid OA group have received a Creative Commons license. The most popular license is CC-BY (2,138 resources), apart from others such as CC-BY-NC-ND (486 resources), CC-BY-NC (292 resources), CC-BY-NC-SA (91 resources), IMPLIED-OA (86 resources), PUBLIC DOMAIN (24 resources) CC-BY-SA (16 resources), CC-BY-ND (2 resources). The highest altmetric attention score (AAA score of 8866.192000000097) has been obtained by a journal article published in a close-access journal, namely *The Lancet* in 2018 from Panjab University.

A total of 1,07,168 publications (out of 1,40,407 publications with DOI), i.e., 77.72% of the contributions have received at least 1 citation as per the citation corpus from Dimensions, but 1,06,790 contributions having OA status have been recorded in the datasets from Unpaywall. The range of the citations received by these contributions varies widely from 1 citation at the lowest to 6665 citations at the highest (as on January 13th, 2022). The publications from the 'close access' group were 80,937 out of 1,06,790 contributions that had received citations (75.79%), and in the 'open access' groups, 25,853 contributions (24.21% of 1,06,790) had received citations. The citations received by OA publications have been distributed into four OA routes: Gold OA (13,981 publications-54.08% of the contributions that received citations in the OA group), Green OA (5,202 publications-20.12%), Bronze OA (3,920 publications-15.16%), and Hybrid OA (2,750 publications-10.64%).

The proportion of 'close vs open' in terms of citations changes considerably with the citation cut-off value (Table 6); the ratio of 75.79% (close); 24.21% (open) changes to 63.96% (close); 36.04% (open) when the citation cut-off is > = 100 citations; 39.10% (close): 60.90% (open) when the citation cut-off is >= 250 citations; 33.33% (close); 66.67% (open) when the citation cut-off is >= 500 citations;

Table 6 — Citations 'score range' blocks and OA categories

Citations score range	Closed Access publications	Open Access publications	Open Access categories			
			Green	Gold	Bronze	Hybrid
1 to 99	80,142	25,298	5,109	1,3813	3,827	2,549
100 to 249	724	408	75	141	62	130
250 to 499	61	95	11	25	21	38
500 to 749	6	12	2	2	3	5
750 to 999	3	8	1	Nil	3	4
1000 to 6665	1	32	4	Nil	4	24
Total	80,937	25,853	5,202	13,981	3,920	2,750

27.27% (close); 72.73% (open) when the citation cut-off is > = 750 citations. When the citation cut-off is set to >=1000, only 1 out of 33 publications (with 1000 or more citations) belongs to a closed access group, while the rest 32 publications in an open access group are distributed as follows: Hybrid (24 publications), Bronze (4 publications), and Green (4 publications).

A total of 27,030 publications (contributions 19.25% out of 1,40,407 publications with DOI) have received altmetric scores ranging from 0.25 (at the lowest) to 8866.192 at the highest scale. A total of 99.40% (n = 26,869) of publications out of 27,030 publications have got "open/closed status" from Unpaywall. Of those 26,869 papers, 57.11% of contributions are closed access and 42.89% of contributions are open access. A total of 11,524 open access papers have been distributed through various OA routes (Table 7), including Gold (5,858 articles), Green (2,634 articles), Hybrid (1,615 articles), and Bronze (1,417 articles).

Indicator for Open Access Friendliness (OAF)

The present section is concerned with the use of a proposed framework[3] named Open Access Friendliness Indicator (OAFI). The OAFI framework is based on four primary areas of OA: *i) OA publications share; ii) licensing of open access publications; iii) share of citations received by OA publications; and iv) the attention that OA publications have received from socio-academic webspace.* The first two factors depend on the awareness of authors and related affiliation institutions. The last two factors depend on the impact of OA publications in academic spheres. Each of these areas is also subdivided into groups. A total of nine factors and a 100-point weightage scale have been used for this present study in calculating OAFI, deviating slightly from the OAFI framework as proposed by Mukhopadhyay[3] (Table 8).

The nine OA friendliness factors that form the assessing framework are described here:

OA share: The OA share represents the percentage of OA publications (from Unpaywall's response of open/closed status) during 2012–2021 for a given state university. The value is obtained following the formula: total OA publications during the period divided by the total publications (with access status from Unpaywall) during the period under study and then multiplied by the group weightage of 25.

***Green and Gold OA share*:** This indicator represents the percentage of green and gold OA publications of an individual institution during 2012–2021. The Green and Gold OA routes are considered because these two routes are more liberal than the other two (Bronze and Hybrid) routes. The value is obtained following the formula: total gold and green OA publications during the period divided by the total OA publications during the period and multiplied by the group weightage of 15.

***Repository share*:** The natural destinations of Green OA publications are repositories. This indicator is based on individual institutions' Green OA publications, which are presented in an institutional repository worldwide. The value is obtained following

Table 7 — Altmetric 'score range' blocks and OA categories

Altmetric score range	Closed Access publications	Open Access publications	Open Access categories			
			Green	Gold	Bronze	Hybrid
0.25 to 99.999	15,314	11,284	2,595	5,806	1,385	1,498
100 to 499	28	173	27	43	25	78
500 to 999	1	40	7	7	5	21
1000 to 1,999	2	25	5	2	2	16
2000 to 2,999	Nil	1	Nil	Nil	Nil	1
3000 to 8867	Nil	1	Nil	Nil	Nil	1
Total	15,345	11,524	2,634	5,858	1,417	1,615

Table 8 — Areas, groups and factors for calculating Open Access Friendliness Indicator (OAFI)

Areas (weightage)	Groups within the areas (with distributed weightage)		
OA publications (Area weightage: 50%)	OA share (Group weightage: 25%)	Gold & Green share (Group weightage: 15%)	Repository share (Group weightage: 10%)
OA licensing (Area weightage: 30%)	OA license share (Group weightage: 20%)	Gold & Green license share (Group weightage: 10%)	
OA impact: Citations (Area weightage: 10%)	OA citation share kind I (Group weightage: 5%)	OA citation share kind II (Group weightage: 5%)	
OA impact: Altmetric (Area weightage: 10%)	OA altmetric share kind I (Group weightage: 5%)	OA altmetric share kind II (Group weightage: 5%)	

the formula: summing up the total green OA papers of an institute available through repositories (from the best OA location value of Unpaywall), dividing the sum by the total green OA papers made available by that institute, and then multiplying by the group weightage value of 10.

***OA license share*:** Another important factor is the share of total OA publications with a license (mainly Creative Commons license) in OA publications. This formula is: total OA publications with formal OA licenses (during the period) divided by the total OA publications during the period and then multiplied by the weightage value of 20.

***Gold and Green license share*:** The formula for this group is: total gold and green OA publications with formal OA licenses (during the period) divided by the total gold and green publications during the period and then multiplied by the allocated weightage value of 10.

***OA citation share kind I*:** It measures the impact of OA publications of a given institute in terms of citations. It first goes for summing up all citations received by all publications of a given institute during the period of study (A). The average citation per publication (round up) is then calculated by dividing the sum of all citations by total publications (it actually helps in calculating indicator value for the next group i.e., OA citation share kind II). Similarly, the sum of all citations received by only OA papers is done (B) and then it is divided by the sum of all citations received for all publications during the period of study (A), and finally multiplied by the allocated group weightage value of 5 - (B÷A*5).

***OA citation share kind II*:** The average citation value as obtained in the previous group is utilized here as a cut-off point. It first sums up all citations received by only those publications that are having citations greater than (>) the average citation value (A) and then sums up all citations received by OA publications with citations greater than the average citation value (B). The group value is obtained by the formula – (B÷A*weightage value for the group). The weightage value for this group is 5.

***OA altmetric share kind I*:** It measures the impact of OA publications of a given institute in terms of altmetric attention scores. This factor is important to consider due to the increasing influence of socio-academic web spaces on the scholarly communication process. It starts with summing up altmetric attention scores received by the publications of an institution during the period of study (A). Similarly, the sum of altmetric attention scores received by only OA papers is done (B), and then it is divided by the sum of altmetric attention scores by all publications of the institute during the period of study (A), and finally multiplied by the group weightage value of 5 (B÷A*5).

***OA altmetric share kind II*:** It measures the share of OA in publications with high-value altmetric scores. The average altmetric attention score per publication (A/B - roundup) is calculated by dividing the sum of altmetric attention scores (A) by all publications with altmetric scores (B). The average value is then utilized as a cut-off point for further processing. It goes for summing up the altmetric attention score received by those publications that are having scores greater than (>) the average altmetric attention score (C) and then sums up altmetric attention scores of OA publications with the altmetric score greater than the average altmetric attention score (D). The group value is obtained by the formula – (D÷C*5). The weightage value for this group is 5.

Institutional Scenarios

This section explains the application of OAFI (open access friendliness indicator) for 26 state universities. This process produces an open access friendliness ranking of 26 NIRF-enlisted (top 100 overall category) state universities. For example, two state universities are selected from the list based on their year of establishment. The oldest one is Calcutta University (CU), established in 1857, and the newest one is Guru Gobind Singh Indraprastha University (GGSI, established in 1998. The three state universities, i.e., Calcutta University, University of Madras, and Mumbai University, were established in the same year, but here we have considered Calcutta University as it has a higher NIRF rank.

The OAFI scores of the oldest and the newest state universities are shown in Tables 9 and 10. The combined scores under three areas (area I:OA publications, area II:OA impact by citations, and area IV:OA impact by altmetric as stated in Table 9 and 10) achieved by the University of Calcutta (CU, est. 1857) are higher than the scores achieved by the newest established university (Guru Gobind Singh Indraprastha University (GGSI, est. 1998). On the other hand, the newest state university (Guru Gobind Singh Indraprastha University, GGSI, 1998) scores

Table 9 — OAF scores for area I & area II for two selected state universities

Area I: OA publications (weightage 50) & Area II: OA licensing (weightage 30)

Factors		Values (round)		Factors		Values (round)	
SL	Element	CU	GGSI	SL	Element	CU	GGSI
A	Total publications	11,771	3,237	E	Total Gold & Green OA	1,660	407
B	Total publications with DOI	10,885	2,946	F	Total Green OA	598	58
C	Publications with close/open status from Unpaywall	10,784	2,904	G	Green OA available via repositories (as the best OA locations)	570	50
D	Total OA publications	2,324	540	H	Licensed OA resources	1,255	330
Factor 1: OA share (D/C * 25)		5.39	4.65	K	Licensed Gold & Green OA	1,009	294
Factor 2: Gold & Green OA share (E/D * 15)		10.71	11.31	Factor 4: OA license share (H/D * 20)		10.80	12.22
Factor 3; Repository share (G/F * 10)		9.53	8.62	Factor 5: Gold & Green license share (K/E * 10)		6.08	7.22
Area I scores		25.63	24.58	**Area II scores**		16.88	19.45

Table 10 — OAF scores for area III & area IV for two selected state universities

Area III: OA impact - citations (weightage 10) & Area IV: OA impact - altmetric (weightage 10)

Factors		Values		Factors		Values	
SL	Element	CU	GGSI	SL	Element	CU	GGSI
A1	Total publication with citation status	10,676	2,894	K	Total publications with altmetric score	2,534	342
A2	Total publications with citation >=1	8,416	2,139	L	Sum of altmetric scores for all publications with score >0	15,918	1,106
B	Total citations received by all resources that are having citations status	1,08,482	21,223	M	Average altmetric score publication (AAvg= Q/P)	6.28	3.23
C	Average citations per publication (Avg= B/A1)	10	7	N	Number of OA publications with altmetric score	1,022	148
D	OA publications with citation >=1	1,846	398	O	Sum of altmetric scores for OA publications with score >0	14,268	604
E	Citations received by OA resources with citation >=1	31,005	4,668	P	All publications with altmetric score >AAvg (AAvg is the average altmetric score here)	299	68
F	All publications with citation >Avg (Avg is the average citation value 'C' here)	2,888	774	Q	Sum of altmetric scores for all publications with score >AAvg	15,613	1,068
G	Citations received by all resources with citation >Avg	86,403	17,105	R	Number of OA publications with altmetric score >AAvg	185	40
H	OA Publications with citation >Avg	676	171	S	Sum of altmetric scores for OA publications with score >AAvg	14,050	576
J	Citations received by OA resources with citation >Avg	26,086	3,914	Factor 9: OA altmetric share kind I (O/L * 5)		4.48	2.73
Factor 6: OA citation share kind I (E/B * 5)		1.43	1.10				
Factor 7: OA citation share kind II (J/G * 5)		1.51	1.14	Factor 10: OA altmetric share kind II (S/Q * 5)		4.50	2.70
Area III scores		2.94	2.24	**Area IV scores**		8.98	5.43

higher in area II (OA licensing) indicating better OA license awareness amongst faculty and researchers.

The final OAF ranking (Table 11) of the top 26 state universities shows that King George`s Medical University has got the top position in two areas (area I-OA publication and area III-OA impact – citations). On the other hand, Panjab University has scored the top position in area II (OA licensing) and Delhi Technological University has got the highest score in area IV (OA impact – altmetric). The OAF ranked list reveals that OA awareness in the state universities across the country is increasing over the years, irrespective of regions, age of a university or discipline-specific nature. The good side of this ranked list is to know that 92.30% of state universities (24 out of 26 state universities) are passing the 50 mark on this 100-point scale to measure OA friendliness.

Table 11 — Ranked list of 26 State universities by OAF

Name of State Universities	Area 1 (50)	Area 2 (30)	Area 3 (10)	Area 4 (10)	OAFI (100)	Rank
King George`s Medical University	35.28	16.57	7.70	9.41	68.96	**1**
Panjab University	27.20	23.13	6.10	9.44	65.87	**2**
Gauhati University	29.86	20.78	5.09	8.85	64.57	**3**
Mysore University	31.07	21.74	3.53	6.56	62.90	**4**
University of Madras	27.81	21.60	2.23	7.45	59.09	**5**
Savitribai Phule Pune University	26.57	17.07	5.93	9.30	58.87	**6**
Madurai Kamaraj University	26.07	21.70	2.06	8.54	58.38	**7**
University of Kashmir	27.91	19.43	2.90	8.00	58.23	**8**
Andhra University	27.73	17.48	2.92	7.99	56.12	**9**
Bharathidasan University	28.54	17.14	2.36	7.53	55.56	**10**
Bharathiar University	27.06	17.94	2.15	7.96	55.12	**11**
Delhi Technological University	24.75	16.46	3.97	9.63	54.81	**12**
Calcutta University	25.63	16.88	2.94	8.98	54.43	**13**
Cochin University of Science and Technology	26.96	16.61	2.56	8.08	54.21	**14**
Mahatma Gandhi University	24.06	18.53	2.39	9.15	54.13	**15**
Osmania University	27.38	18.38	3.01	5.13	53.89	**16**
Sri Venkateswara University	26.37	18.58	1.98	6.25	53.18	**17**
Alagappa University	26.68	17.23	1.61	7.40	52.92	**18**
Guru Nanak Dev University	25.07	18.77	1.98	6.99	52.81	**19**
Calicut University	26.63	16.87	2.41	6.07	51.98	**20**
Guru Gobind Singh Indraprastha University	24.58	19.45	2.24	5.43	51.69	**21**
Kerala University	24.59	18.61	2.56	5.66	51.42	**22**
Anna University	24.13	18.06	1.47	6.98	50.64	**23**
Jadavpur University	25.14	15.64	2.05	7.35	50.19	**24**
Mumbai University	25.43	15.30	2.28	6.81	49.81	**25**
Gujarat University	24.62	20.71	1.92	1.59	48.84	**26**

Conclusion

This large-scale study to assess the OA friendliness of a set of selected institutions necessitated three major components: the availability of data sources and a tool to handle large datasets; mechanisms for auto-fetching of JSON-formatted datasets from the selected sources; and extraction of values required for calculation. The data carpentry based methodology as adopted by this research study made it possible to fetch and manage OA status data, citation data, and altmetric attention scores mechanically from Unpaywall.org, Dimensions.ai, and Altmetric.com respectively. The open-source data wrangling tool OpenRefine is a powerful tool to fetch, extract, and manage the value, volume, and variety of the datasets as gathered by following the stated methodology, and without OpenRefine, this study simply could not be possible. The data carpentry methods, tools, and sources are all set to change the way LIS professionals conduct data-intensive research studies like scientometrics, informetrics, and bibliometrics studies. It may be stated that this methodology and the proposed OAFI framework can be extended to measure the OA friendliness of different categories of institutions in future studies related to assessing OA affinities, and such OA friendliness scores can be taken into consideration for measuring the performance of an educational or research institute.

References

1 Mering M and Hoeve C D, A brief history to the future of open access, *Serials Review,* 46(4) (2020) 300–304.

2 Roy B K, Biswas S C and Mukhopadhyay P, Open access repositories for Indian universities: towards a multilingual framework, *IASLIC Bulletin,* 61(4) (2016) 150–161.

3 Mukhopadhyay P, 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,* Ess Ess, (2022) 67–89. doi: 10.5281/zenodo.6511080.

4 Naika M & Pathak S K, India's open access future, *Nature India,* (2020). doi:10.1038/nindia.2020.125.

5 Open Access - Global Trends and Rankings. *wizdom.ai* Available at https://www.wizdom.ai/dashboards/open-access (Accessed February 15, 2022).

6 Aguillo I F, Ortega J L, Fernandez M & Utrilla A M, Indicators for a webometric ranking of open access repositories, *Scientometrics,* 82(3) (2010) 477–486.

7 Alperin J P, Babini D and Fischman G, *Open access indicators and scholarly communications in Latin America*, (CLACSO) 2014.

8 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*,* (2013).

9 Archambault É, Amyot D, Deschamps P, Nicol A, and Provencher F*,* Proportion of Open Access Papers Published

in Peer-Reviewed Journals at the European and World Levels—1996–2013, European Commission, (2014).

10 Gomez N, Bustos-Gonzalez A, Santillan-Aldana J & Arias O, Open access indicators and information society: The Latin American Case, *OCLC Systems & Services: International digital library perspectives,* 25(2) (2009) 82–92.

11 Maddi A, Measuring open access publications: a novel normalized open access indicator, *Scientometrics,* 124 (2020) 379–398. doi: 10.1007/s11192-020-03470-0

12 Piwowar H, Priem J, Larivière V, Alperin J P, Matthias L, Norlander B, Farley A, West J and Haustein S, The state of OA: a large-scale analysis of the prevalence and impact of Open Access articles, *PeerJ,* 6 (2018) e4375. doi: 10.7717/peerj.4375

13 Piwowar H, Priem J & Orr R, The Future of OA: A large-scale analysis projecting Open Access publication and readership, *BioRxiv* (2019) 795310. doi:10.1101/795310

14 Robinson-Garcia N, Costas R & van Leeuwen T N, Indicators of Open Access for universities, *arXiv preprint arXiv:1906.03840,* (2019).

15 Robinson-Garcia N, Costas R & Leeuwen T N van, Open Access uptake by universities worldwide, *PeerJ,* 8 (2020) e9410. doi: 10.7717/peerj.9410

16 Martín-Martín A, Costas R, van Leeuwen T & Delgado López-Cózar E, Evidence of open access of scientific publications in Google Scholar: A large-scale analysis, *Journal of Informetrics,* 12 (3) (2018) 819–841. doi: 10.1016/j.joi.2018.06.012

17 Bhat M H, Open access publishing in Indian premier research institutions, *Information Research: An International Electronic Journal,* 14 (3) (2009) 3.

18 Piryani R, Dua J & Singh V K, Open Access Levels and Patterns in Scholarly Articles from India, *Current Science,* 117 (9) (2019) 1435–1440.

19 Nazim M, Analysing Open Access Uptake by Academic and Research Institutions in India, *DESIDOC Journal of Library & Information Technology,* 41(2) (2021) 108–115.

20 Roy A & Mukhopadhyay P, Measuring Open Access Friendliness of Indian Central Universities through Data Carpentry, *SRELS Journal of Information Management,* 59 (3) (2022) 131–139.