

How Green is My Valley? Measuring Open Access Friendliness of Indian Institutes of Technology (IITs) through Data Carpentry

Parthasarathi Mukhopadhyay
Professor,
Department of Library and Information Science,
University of Kalyani
(psmukhopadhyay@gmail.com)

Abstract

This research study aims to develop a distributed weightage based ranking framework for measuring Open Access (OA) support/friendliness (Open Access Friendliness Indicator – OAFI) of a given Indian institute. It applies data carpentry tools and methods for gathering and extracting OA data from an array of diverse sources available against ODbL. The ranking framework has four primary areas (OA publications share, OA license share, OA citations share and OA altmetric share) and a total of ten factors under these selected areas. The distributions of weightage have been set on the basis of SWOC analysis of open access scenario in India and the product is a 100-point scale for measuring OA friendliness of a given institute. The framework has been tested with the publications data (1,59,107), citations data (21,69,395), and altmetric data (24,308) of the top Indian Institutes of Technology (16 IITs) that are listed in the top 100 of the overall category of the National Institutional Ranking Framework (NIRF), 2020. The final ranked list of the selected 16 IITs shows that in general older IITs are much ahead in terms of numbers (publications, citations, altmetric) but newer IITs are tenanted all top five positions in OAFI framework as factor formulae are ratio dependent. It also shows that Indian Institute of Technology Gandhinagar (established in 2008) occupied the top most position with leading scores in area I (OA publications share) and area III (OA citations share) and Indian Institute of Technology Bhubaneswar has obtained the highest value in area II – OA license share, and Indian Institute of Technology Ropar has topped the list in area IV – OA altmetric score share.

Keywords: Data carpentry, Open access, Open access friendliness, OAFI ranking framework, Indian Institutes of Technology (IITs)

1 Introduction

The concept of Open Access (OA) has changed the world of scholarly communication, and India is no exception to this new trend. The researchers and contributors in India are adopting different routes of OA since late 2000s to support OA philosophy and to get the OA citation advantages. But till date we don't have a ranking framework to measure OA support by a given Indian institute. A few researchers attempted to measure OA friendliness of institutes at the country/regional/global scale mainly by considering parameters like OA share in publications, number of repositories from the country/region listed in ROAR or OpenDOAR, number of documents in those listed repositories, number of OA journals listed in DOAJ, ranking of OA repositories, and adaptation of Creative Commons licenses by the researchers of a country/region during 2009-2014 (Aguillo et al., 2010; Archambault et al., 2014; Gómez et al., 2009), but in 2014 an edited book entitled "Open access indicators and scholarly communications in Latin America", and published by UNESCO (Alperin et al., 2014) first pointed out the need to consider citation data and altmetric data, apart from OA publications & OA licensing, in developing indicators to measure OA share in citations

Post-print

and altmetric scores. A team of researchers lead by Robinson-Garcia has attempted to develop a framework to measure OA support in global universities (963 universities across the world including institutes from India) by taking into consideration different OA routes (Gold, Green, Bronze, Hybrid). The publication dataset for the study has been created from Web of Science and Unpaywall has acted as OA related data source (Robinson-Garcia, Costas, et al., 2020; Robinson-Garcia, Leeuwen, et al., 2020).

In view of the above stated initiatives, this research study aims to develop a distributed weightage based ranking framework to measure OA friendliness of a set of given Indian institutions by taking into consideration four primary areas namely – OA publications share, OA licenses share, OA citations share and OA altmetric scores share. Each of these primary areas are subdivided again to pinpoint distribution of weightage, for example, the area weightage for “OA publications share” is subdivided into three groups like – OA publication share in terms of total publications, Green & Gold OA percentage in terms of total OA, and Green OA availability in repositories as best OA locations. The sample for this research study includes a set of 16 IITs that are listed in the “Overall” category of the National Institutional Ranking Framework (NIRF) 2020 list (so far the latest NIRF ranked list – see table 1).

Table 1: IITs in NIRF (2017-2020) – arranged by ranks in NIRF, 2020

SL	Name of IIT (year of establishment as IIT)	Ranks in NIRF lists (2017-2020)			
		NIRF 2017	NIRF 2018	NIRF 2019	NIRF 2020
01	Indian Institute of Technology Madras (1959)	02	02	01	01
02	Indian Institute of Technology Delhi (1961)	05	04	03	03
03	Indian Institute of Technology Bombay (1958)	03	03	04	04
04	Indian Institute of Technology Kharagpur (1951)	04	05	05	05
05	Indian Institute of Technology Kanpur (1959)	07	07	06	06
06	Indian Institute of Technology Guwahati (1994)	08	12	09	07
07	Indian Institute of Technology Roorkee (2001)	09	08	08	09
08	Indian Institute of Technology Hyderabad (2008)	26	22	22	17
09	Indian Institute of Technology (ISM) Dhanbad (2016)	53	27	25	22
10	Indian Institute of Technology Indore (2009)	24	28	NL	23
11	Indian Institute of Technology (BHU) Varanasi (2012)	70	NL	28	26
12	Indian Institute of Technology Gandhinagar (2008)	78	39	51	35
13	Indian Institute of Technology Ropar (2008)	32	NL	NL	39
14	Indian Institute of Technology Patna (2008)	83	69	58	54
15	Indian Institute of Technology Bhubaneswar (2008)	66	51	46	56
16	Indian Institute of Technology Mandi (2009)	37	NL	44	67

(NL: Not Listed, Year of establishment within parenthesis, Overall category in NIRF was introduced in 2017)

2 Data Carpentry: Tools and Techniques

Many of the early researchers in the domain of open access indicator reported the problems of obtaining OA related data against the publications of a given institute that can help to measure values beyond OA publications share (Bar-Ilan, 2008; Hajjem et al., 2005; Harnad et al., 2004, 2008). The present technological advancements like API based data wrangling, JSON-formatted data responses, availability of open source data carpentry software, emergence of large-scale ODbL-powered data sources like Unpaywall, Dimension, OpenCitations, Crossref, Altmetric.com and so on have changed the situation of OA data poverty faced by the early researchers. The in-

Post-print

depth data gathering and analysis as included in this research study are based on the open source data carpentry software called Openrefine and an array of diverse data sources (see section 4.3) as proposed by the domain researchers (Leeuwen et al., 2018; Maddi, 2020; Robinson-Garcia, Leeuwen, et al., 2020). The data carpentry methods like gathering of OA data through API call from the selected data sources, extraction of data from JSON responses and availability of deep faceting features in Openrefine allow this study to go beyond OA publications share in designing the ranking framework (Mukhopadhyay et al., 2021; Mukhopadhyay & Mukhopadhyay, 2021).

3 Objectives and Research Questions

The primary objective of this research study is to develop a ranking framework that can measure OA support/friendliness of a given institute by taking into purview a diverse set of data, and thereby can go beyond measuring only OA publications share. The ranking framework includes the factors that are considered important and supports distributed weightage of the identified factors on the basis of analysis of the inherent nature of the OA datasets gathered and extracted by using Openrefine from an array of identified data sources. In view of these stated objectives, the following two interlinked research questions are designed to trigger this research study in proper directions:

- RQ1: How, and to what extent, is it feasible to develop a distributed weightage based Open Access Friendliness Indicator (OAFI) for a given Indian institute by taking into consideration diverse areas like OA publication share, OA license share, OA citations share and OA altmetric share (as identified by researchers of the domain)? What should be the basis for distribution of weightage for these primary areas? Is it feasible to identify a set of important factors within each of these major areas? If yes, how to gather data related to the identified factors and how to design a scale of weightage for the identified factors in each of the primary area?
- RQ2: Can the open source data carpentry software (Openrefine) provide the facilities of deep faceting required for an in-depth data exploration like this? What should be the methods to gather OA related data required for this study from a set of diverse data sources and to extract required values for the identified areas and factors? How to merge institute-specific datasets into a consolidated project in Openrefine to support deep faceting against a specific type of institutes (like a project for all IITs)?

4 Methodology

The present research study aims to build an indicator to measure Open Access (OA) friendliness of an institution of Indian origin. The sample includes all Indian Institutes of Technology (IITs) that are included in the top 100 institutions of the Overall category of NIRF 2020 (the dataset has been released on 11th June 2020 and so far the latest available ranked list of Indian institutions by NIRF). The ‘Overall’ category is a sort of combined rank lists of all the participating institutions in different categories like University, Engineering, Medical, Law, Management, Architecture, Pharmacy, Law, Dental and College (started from 2017 onwards). As a result, the ‘Overall’ category represents the top brass of the Indian institutions included in different discipline-specific categories. The ‘Overall’ category of NIRF, 2020 includes a total of 200 institutions and is divided into three rank bands – rank band I (rank band 1-100), rank band II (rank band 101-150) and rank band III (rank band 151-200).

Post-print

4.1 Selection of Institutions

There is a total of 23 Indian Institutes of Technology (IITs) but only 16 are included in the elite group of NIRF, 2020 (top 100 of the ‘Overall’ category) with the rank ranges from 1 to 67. These 16 IITs (listed according to their ranks in the top 100 of NIRF, 2020 in the ‘Overall’ category) are - Indian Institute of Technology Madras (1); Indian Institute of Technology Delhi (3); Indian Institute of Technology Bombay (4); Indian Institute of Technology Kharagpur (5); Indian Institute of Technology Kanpur (6); Indian Institute of Technology Guwahati (7); Indian Institute of Technology Roorkee (9); Indian Institute of Technology Hyderabad (17); Indian Institute of Technology (ISM) Dhanbad (22); Indian Institute of Technology Indore (23); Indian Institute of Technology (BHU) Varanasi (26); Indian Institute of Technology Gandhinagar (35); Indian Institute of Technology Ropar (39); Indian Institute of Technology Patna (54); Indian Institute of Technology Bhubaneswar (56); and Indian Institute of Technology Mandi (67). The respective ranks in NIRF 2020 (Overall category) are indicated with parenthesis against each listed IIT.

4.2 Development of Primary Dataset

The primary dataset of publications by these 16 IITs has been created by searching Scopus database with a suitable query that includes affiliation id [AF-ID (*eight-digit ID number*)] of the institution, all document types and time range from 2010 to 2019 (10 years). Scopus allows downloading of only 2000 retrieved records in ‘csv’ format, therefore contributions of each IIT under study are retrieved against a given year (e.g. 2015) and then all 10 ‘csv’ files (resulted from 10 years) related to a given IIT are merged into a single ‘csv’ file through a suitable script. In some cases, the retrieved results, even in a given year, exceeded the download limit of 2000. The problem of this sort has been addressed by dividing the year into two groups - January to June and July to December to support the workflow of data collection. The final primary dataset includes 16 ‘csv’ files for 16 IITs listed in the top 100 of ‘Overall’ category of NIRF, 2020. A summary table (table 2) gives a panoramic view of the primary dataset.

Table 2: Primary dataset (arranged by total number of publications in descending order)

SL	Name of IIT	Total Publications (2010-2019)	Publications with DOI	Rank in NIRF 2020
01	Indian Institute of Technology Kharagpur	22365	20889	05
02	Indian Institute of Technology Bombay	21179	19378	04
03	Indian Institute of Technology Madras	20891	18989	01
04	Indian Institute of Technology Delhi	19930	18185	03
05	Indian Institute of Technology Roorkee	16000	14713	09
06	Indian Institute of Technology Kanpur	13197	12294	06
07	Indian Institute of Technology Guwahati	11975	11384	07
08	Indian Institute of Technology (ISM) Dhanbad	8527	7621	22
09	Indian Institute of Technology (BHU) Varanasi	6521	5941	26
10	Indian Institute of Technology Hyderabad	4061	3858	17
11	Indian Institute of Technology Indore	3456	3381	23
12	Indian Institute of Technology Bhubaneswar	2554	2410	56
13	Indian Institute of Technology Patna	2518	2385	54
14	Indian Institute of Technology Ropar	2049	1962	39
15	Indian Institute of Technology Gandhinagar	1977	1859	35
16	Indian Institute of Technology Mandi	1907	1829	67
Grand total of publications		1,59,107	1,47,078	

Post-print

The publications with DOI in the primary dataset has a very important role to play in developing the secondary dataset as the entire data wangling process stands on DOI based data fetching to collect OA related data, citations data and altmetrics data from data sources available through Open Data Commons Open Database License (ODbL).

4.3 Development of Secondary Dataset

The primary dataset, as prepared by merging publication datasets of all 16 IITs, is now ready for enrichment with data elements related to OA status, citations and altmetrics data (surprisingly almost 12 thousands papers (7.56%) do not have DOI published by IITs). The data sources available against ODbL for achieving the stated purposes are as follows:

Open Access Data: Unpaywall is presently the largest bibliographic storehouse for open contents (2,97,12,856+ records as on 31st July 2021 and counting). It harvests open contents from 50,000+ publishers and allows free access to the dataset on the top of REST/API call (version 2 with DOI endpoint - GET /v2/:doi) against a generous call limit 1,00,000 calls per day. The API call structure with the valid responses received is arranged in table 3.

Table 3: Unpaywall – API calls and result

API call structure for Unpaywall	No. of queries sent	Responses received
"https://api.unpaywall.org/v2/" + value + "?email=psmukhopadhyay@gmail.com"	1,47,078	1,44,969
value is DOI	publications with DOI	98.56% of publications with DOI

Citations Data: The open data revolution has paved the path of obtaining citation data from many sources against ODbL but the two most visible products are – i) Open Citation Corpus (opencitations.net); and ii) Dimensions (app.dimensions.ai). OpenCitations, a not-for-profit agency developed by JISC, UK, aims to provide open bibliographic and citation data by using semantic technologies like LOD (Linked Open Data) and it is quite easy to gather citations count through DOI based API call. There is no restriction or call limit for fetching data from open citation corpus. On the other hand, Dimensions is a profit making agency, but it allows free API calls to metrics datasets for obtaining citations related data against DOI or PMID. The response from Dimension includes a rich dataset that includes times_cited (all citations as on date), recent_citations (citations received in last two years), relative_citation_ratio (performance of an article in comparison with the citations received by other articles in the same domain of research) and field_citation_ratio (impact of an article in comparison with the articles in the same subject are having similar publication age). The API call structures for both the services are given in table 4.

Table 4: Citation data sources

Service name	API call structure	No. of queries sent	Responses received
OpenCitations	"https://opencitations.net/index/api/v1/citation-count/" + value	1,47,078	1,11,312
	value is DOI	publications with DOI	75.68% publications
Dimesions	"https://metrics-	1,47,078	137460

Post-print

	api.dimensions.ai/doi/" + value		
	value is DOI	publications with DOI	93.46% publications

As Dimensions dataset is much more comprehensive in terms of coverage, currency and number of data elements, this research study has decided to use citation data corpus from Dimensions.ai for further analysis.

Altmetric Data: The only altmetrics data source that provides free API call to gather altmetric data is altmetric.com (api.altmetric.com). A researcher can typically make 1200 API calls per day without a license key or 86400 daily calls against a license key (which can be obtained free of cost against application). It delivers a rich dataset in JSON format that includes many socio-academic webspaces like twitter, facebook, blogs, wikipedia, mendeley, citeulike, connotea and altmetric attention score. The API call structure and the coverage are included in table 5.

Table 5: Altmetric data – API calls and result

API call structure for Almetric	No. of queries sent	Responses received
"https://api.altmetric.com/v1/doi/" + value + "?key=Your-Key-Here"	1,47,078	24,308
value is DOI	publications with DOI	16.52% publications

4.4 Data Extraction

The value of a dataset lies in its extraction. The next important step in the methodology is data extraction from the datasets gathered in JSON format by using the data sources as mentioned in section 4.3. The open source data wrangling tool in use i.e. Openrefine provides a comprehensive parsing mechanism for JSON data called GREL (General Refine Expression Language). Generally, data extraction involves a series of GREL syntaxes to extract target data from a JSON formatted dataset depending on the needs of a researcher and the structure of datasets obtained. GREL supports many functions like string parsing & splitting, mathematical functions, format based functions (JSON, XML, HTML), array functions, date functions, Boolean functions, string functions and so on. A set of simple examples is illustrated in table 6.

Table 6: GREL – an example to extract data from Unpaywall dataset

Response from Unpaywall in JSON	GREL for data extraction	Extracted data
<pre>{ "best_oa_location": { "host_type": "publisher", "is_best": true, "license": "cc-by", }, "doi": "10.1186/1687-1499-2011-87", "has_repository_copy": true, "is_oa": true, }</pre>	value.parseJson().best_oa_location.host_type	publisher
	value.parseJson().best_oa_location.license	cc-by
	value.parseJson().has_repository_copy	true
	value.parseJson().is_oa	true
	value.parseJson().journal_is_in_doaj	true
	value.parseJson().journal_is_oa	true
	value.parseJson().oa_status	gold

Post-print

<pre> "is_paratext": false, "journal_is_in_doaj": true, "journal_is_oa": true, "oa_status": "gold", "published_date": "2011-09-05", "publisher": "Springer Science and Business Media LLC", } </pre>		
-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	--	--

In the same way, GREL syntaxes for data extraction applied in Openrefine to pull out required data values from JSON responses received from Dimensions.ai and Altmetric.com. For example, value.parseJson().times_cited extracts number of citations for a publication from Dimensions.ai and value.parseJson().score retrieves altmetric attention score from Altmetric. Com.

5 Discussion and Findings

This section of the research study starts with the larger picture of the OA contributions by all 16 IITs ranked in the top 100 of NIRF, 2020 (Overall category) under the generic scenarios section and followed by the institute-specific scenarios.

5.1 Generic Scenarios

An analysis of the extracted JSON data from the merged dataset (includes 1,44,969 responses received from Unpaywall against a set of 1,47,078 publications with DOIs) shows that 1,19,065 publications (82.13%) are ‘closed access’ and only 25,904 are OA publications (17.87%). These 25,904 OA publications are made available through four routes – green OA (9672 i.e. 37.34% of total OA contributions); gold OA (9526 i.e. 36.77% of total OA contributions); bronze OA (3685 i.e. 14.23% of total OA contributions); and hybrid OA (3021 i.e. 11.66% of total OA contributions). The general trend of OA publications from 16 IITs reveals that the green and the gold paths of OA are steadily increasing over the years (2010-2019) in comparison with the bronze and hybrid routes as illustrated in Figure 1 on the basis of the values extracted from the datasets as given in table 7. The tabulated data also reveal that there is a steady progress (though bit slow) over the ten years for percentage of OA in total publications (12.43% in 2010 as the lowest to 20.04% in 2018 as the highest) by these 16 top IITs. It also shows that whereas the ‘total publications’ has increased in the tune of 185% and ‘close access’ publications has grown by 163%, the ‘open access’ has registered a growth of 343% by considering the lowest and the highest recorded values in the respective categories. In the ‘open access’ categories, the gold path has the highest growth rate (1236%), followed by the hybrid (838%), green (170%), and bronze (128%) routes of OA (table 7).

Table 7: Growth of OA in IITs (for publications with open/close status from Unpaywall)

Year	Total publications by 16 IITs	Close Access Publications	Open Access Publications	Open Access Categories				% OA in Total publications
				Green	Gold	Bronze	Hybrid	

Post-print

2010	7,997	7,003	994	565	148	229	52	12.43%
2011	8,865	7,641	1,224	583	204	276	161	13.81%
2012	8,959	7,650	1,309	515	292	290	212	14.61%
2013	11,103	9,364	1,739	788	420	339	192	15.66%
2014	12,942	10,849	2,093	908	684	296	205	16.17%
2015	14,361	11,702	2,659	901	1,029	353	376	18.52%
2016	16,667	13,387	3,280	1,135	1,253	442	450	19.68%
2017	19,416	15,593	3,823	1,353	1,612	452	406	19.69%
2018	21,862	17,481	4,381	1,400	1,978	524	479	20.04%
2019	22,797	18,395	4,402	1,524	1,906	484	488	19.31%

It is also clear that there is an increasing awareness of open access licensing systems amongst contributors from these 16 IITs as OA resources with formal licenses have increased from 199 in 2010 to 2,068 in 2018 and 2,065 in 2019. Almost half of the OA resources are now available with some kind OA licenses (10,973 publications out of 25,904 – 42.36%). Another interesting positive point is that the most favourite licensing agreement for contributors from these IITs are – CC-BY (6875 contributions – 62.65%), CC-BY-NC-ND (2071 contributions – 18.87%), and CC-BY-NC (564 contributions – 5.14%).

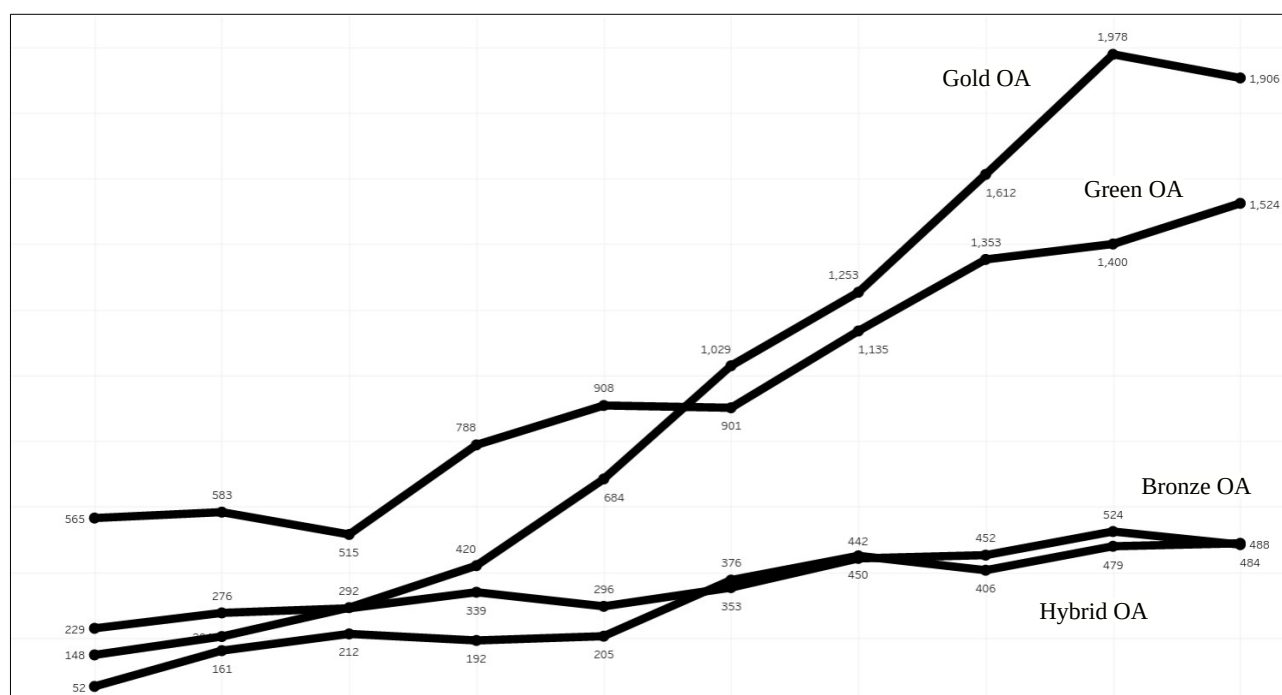


Figure 1. Growth of four OA types in 16 IITs over the years (2010-2019)

The route of green OA is slightly ahead in comparison with gold path as far as 16 top IITs are concerned. It includes a total of 9672 number of contributions published in 2451 number of sources in different documentary forms like articles (7385 contributions – 76.07%), conference papers (1836 – 18.98%), reviews (197 contributions), book chapters (162 contributions), editorial (34

Post-print

contributions), letters (21 contributions), notes (15 contributions), books (14 contributions) and others (8 contributions). The 7385 number of journal papers available through green OA route appeared in 1695 ‘close access’ journals and interestingly the majority of the green OA contributions made by these top 16 IITs do not have any form of open access licenses attached with these resources e.g. 7053 contributions out of 7385 (95.50%) are without any formal OA licenses. The rest 4.5% of the journal papers available in green OA route are having formal OA licenses like CC-BY-NC-ND (151 articles), CC-BY (62 articles), CC0 (50 articles), CC-BY-NC-SA (46 articles), CC-BY-NC (13 articles), IMPLIED-OA (9 articles) and CC-BY-SA (1 article). On the other hand, after the journal papers, the next highest green OA contributions came from conference papers (18.98%) but 1727 number conference papers available in green OA channel do not have any OA license attached (94.06%) and the rest 5.94% contributions are attached with OA licenses like CC-BY (68 conference papers), CC-BY-NC-ND (26 conference papers), CC0 (6 conference papers), CC-BY-NC-SA (5 conference papers), CC-BY-ND (2 conference papers), CC-BY-NC (1 conference paper) and PD – Public Domain (1 conference paper). As a whole 9201 contributions out of 9672 green OA resources from 16 IITs (95.13%) are not attached with formal OA licensing system and rest are distributed with 8 types of creative common licenses of which CC-BY-NC-ND is the mostly adopted OA license type. Indian Institute of Technology Bombay of the top 16 IITs made the highest contribution in green OA route with 1781 contributions (18.41% of green OA publications from 16 IITs). A total of 3840 number of green OA publications from 16 IITs have received altmetric score (39.70% of publications in the green OA category). The highest altmetric score (AAA score 5092.72) is obtained by a green OA publication from Indian Institute of Technology Kanpur. The highest number of citations (1881) is secured by a green OA publication from Indian Institute of Technology Madras published in a close access journal namely *Reviews of Modern Physics* in 2013 and made available through a repository but without any formal OA license. Another interesting observation in the green OA domain may be the study of availability and distribution of green OA publication in different institutional repositories as the open access repositories are natural destinations for such publications. The Unpaywall dataset provides a data element namely ‘repository_institution’ (name of the repository archiving the publication) under the groups ‘best_oa_location’ and ‘oa_locations’. An analysis of this data element shows that a total of 7779 number of publications out of 9672 green OA resources (80.43%) are available through 310 repositories distributed across the globe (as best OA locations). The top five most favourite repositories for researchers of 16 IITs are - Cornell University – arXiv (3439 publications), arXiv.org (1310 publications), pubmedcentral.nih.gov (352 resources), CiteSeerX.psu (327 resources) and The Pennsylvania State University - CiteSeer X (137 publications). A deeper look into the ‘repository_institution’ dataset shows that a total of 276 publications (out of 7779 green OA resources available through repositories - 3.55%) are archived in Indian repositories. The top five Indian repositories are attached with the following institutions (with archived publications) - Indian Institute of Technology Hyderabad (118), Indian Institute of Technology Bombay (43), Indian Institute of Technology Madras (25), Indian Institute of Technology Kanpur (19), Indian Institute of Technology Guwahati (5). This proves that institutional open access repositories in India are not quite favourite destinations for the scholars of the top 16 IITs. The state of licensing for these 7779 numbers of green OA publications is not quite encouraging – only 434 green OA objects are attached with any sort of OA licenses (5.58%) and the variant CC-BY-NC-ND is the most popular license agreement for green OA contributors (171 out of 434 green OA with licenses). The best performance related to archiving of green OA in the own institutional repository comes from a relatively younger IIT – Indian Institute of Technology Hyderabad. It has a total of 504 green OA publications, out of which 394 are available through repositories (78.17%), and 94 contributions out of 394 archived objects in repositories are made available through open access repositories of the institution (<http://raiith.iith.ac.in/>).

Post-print

The 9,526 number of gold OA publications appeared in a total of 1,137 Open Access journals (by the definition of gold OA) and are published by 450 number of publishers. The top 5 favourite OA journals of all 16 IITs are - Scientific Reports (568 publications); Journal of High Energy Physics (494 publications); Procedia Engineering (405 publications); IOP Conference Series: Materials Science and Engineering (356 publications); and Physics Letters, Section B: Nuclear, Elementary Particle and High-Energy Physics (335 publications). The top five most productive OA publication agencies for all 16 IITs are - Elsevier B.V. (1221 publications); Institute of Physics Publishing (683 publications); Elsevier Ltd (643 publications); Nature Publishing Group (573 publications); and Springer Verlag (509 publications). The license scenario in gold OA is comparatively better than the green OA route possibly due to the fact that established OA journals selected by the members of 16 IITs are having the OA licensing system in place and published articles inherited that OA license. A total of 7524 gold OA contributions (78.98%) are having formal OA licenses and distributed in 9 categories of licenses – CC-BY (5378 contributions), CC-BY-NC-ND (1565 contributions), CC-BY-NC (402 contributions), CC-BY-NC-SA (78 contributions), IMPLIED-OA (51 contributions), CC-BY-SA (28 contributions), CC-BY-ND (11 contributions), CC0 (8 contributions) and PD (3 contributions). The highest contribution to gold OA comes from Indian Institute of Technology Madras (1457 contributions i.e. 15.29% of gold OA publications from 16 IITs) and the least from Indian Institute of Technology Ropar (118 contributions). One of the relevant questions for this category is that how many of these gold OA contributions are listed in DOAJ (Directory of Open Access Journals). The good part is that 6938 number of publications have appeared in DOAJ listed journals (72.83%) whereas 2588 contributions are in other than DOAJ listed sources (27.17%). A total of 901 number of journals out of 1137 open access journals (79.24%) are listed in DOAJ. A total of 3292 number of gold OA publications from 16 IITs have received altmetric score (34.52% of 9526 publications in the gold OA category). The highest altmetric score (AAA score 1665.668) obtained by a gold OA publication is from Indian Institute of Technology Bombay. The highest number of citations (1044) in the gold OA category is received by a paper published in 2019 from Indian Institute of Technology Bombay in a DOAJ-listed OA journal namely Physical Review X.

The bronze OA category (3685 in number - constitutes 14.23% of total OA contributions) is distributed in 973 sources published by 262 agencies. The top five favourite publishing agencies for the 16 IITs in bronze route are Springer India (346 publications), Springer (279 publications), Indian Academy of Sciences (273 publications), Oxford University Press (106 publications) and Institute of Physics Publishing (75 publications). The top five destinations for bronze OA are Sadhana - Academy Proceedings in Engineering Sciences (292 papers), Journal of Chemical Sciences (221 papers), Journal of Earth System Science (188 papers), Bulletin of Materials Science (118 papers), and Current Science (108 papers). As expected, none of these contributions in bronze OA route is attached with open license. The highest number of citations (376) in the bronze OA category is received by a paper published in 2017 from Indian Institute of Technology Guwahati in a close access journal namely British Journal of Pharmacology. A total of 678 number of bronze OA publications from 16 IITs have received altmetric score (18.40% of 3685 publications in the bronze OA category). The highest altmetric score (AAA score 1318.976) is obtained by a journal article published in a close access journal namely Science in 2015 from Indian Institute of Technology Bombay.

The hybrid OA resources (3021 in number - constitutes 11.66% of total OA contributions) are made available through 590 sources published by 130 agencies. The top five hybrid OA publishers are - American Physical Society (1046 contributions), Royal Society of Chemistry (182 contributions),

Post-print

Elsevier Ltd (174 contributions), Elsevier B.V. (111 contributions) and Institute of Physics Publishing (105 contributions). The five most favourite journals in hybrid OA route are - Physical Review D (495 papers), Physical Review Letters (352 papers), Physical Review D - Particles, Fields, Gravitation and Cosmology (202 papers), Physical Review C (83 papers), Defence Science Journal (168 papers). Majority of the hybrid OA publications are having licenses attached with these resources (2978 out of 3021 – 98.57%). As expected many of the resources are available through CC-BY (1364 resources – 45.80%) and publisher-specific license agreements (1043 resources – 35.02%). The top two licensing mode in hybrid OA is followed by CC-BY-NC-ND (309 resources – 10.37%), CC-BY-NC (147 resources – 4.97%), IMPLIED-OA (120 – 4.03%), CC-BY-NC-SA (25 resources – 0.84%), PD (13 resources -0.44%), CC-BY-SA (3 resources – 0.10%) and CC-BY-ND (1 only). A total of 1754 number of hybrid OA publications from 16 IITs have received altmetric score (58.06% of 3021 publications in the hybrid OA category). The highest altmetric score (AAA score 4789.5239999999) is obtained by a journal article published in a close access journal namely Physical Review Letters in 2016 from Indian Institute of Technology Gandhinagar. The highest number of citations (5686) in the hybrid OA category is also received by the same paper which secured the highest altmetric score in this category published from Indian Institute of Technology Gandhinagar in 2016. This journal paper made open access through publisher hosted repository is incidentally the top cited journal amongst all the groups (close & open) and all the OA categories.

A total of 1,20,967 publications (out of 1,47,078 publications with DOIs) i.e. 82.25% contributions have received at least 1 citation as per the citation corpus from Dimensions but 1,20,070 contributions are having OA status recorded in the dataset from Unpaywall. The range of citations received by these contributions vary widely from 1 citation at the lowest to 5686 citations at the highest (as on 31st July 2021). The publications from ‘close access’ group is 98,198 out of 1,20,070 contributions that are having access status and have received citations (81.78%) and in ‘open access’ group 21,872 such publications out of 1,20,070 publications have received citations (18.22%). The distributions of these 18.22% publications in different OA routes are – green (8411 publications – 38.28% of contributions that received citations in OA group), gold (7853 publications – 35.74%), bronze (2850 publications – 12.97%), and hybrid (2758 publications – 12.55%). Interestingly, when the citation cut-off value sets at higher end to identify high impact publications by top 16 IITs, the ratio 81.78% (close) : 18.22% (open) changes quickly to - 71.14 % (close) : 28.86 % (open) [when citation cut off is ≥ 100 citations]; 54.90 % (close) : 45.10 % (open) [when citation cut off is ≥ 250 citations]; 42.24 % (close) : 57.76 % (open) [when citation cut off is ≥ 500 citations]; and 32.05 % (close) : 67.95 % (open) [when citation cut off is ≥ 750 citations]. The citation cut off when sets at ≥ 1000 mark, only 9 out of 52 publications (with citations 1000 and more) belong to ‘close access’ group and the rest 43 publications in OA group are having the following distribution pattern – hybrid (33 publications), green (6 publications), and gold (4 publications) with the close to open ratio – 17.31 % (close) : 82.69 % (open). The close to open access ratio for publications with extremely high citations depicts a completely reverse scenario in comparison to the generic citation analysis i.e 81.78% (close) : 18.22% (open) ratio (table 8).

Table 8: Citations ‘score range’ blocks and access types of publications

Citations score range	Close Access publications	Open Access publications	Open Access categories			
			Green	Gold	Bronze	Hybrid
1 to 99	96,788	21,300	8,206	7,726	2,812	2,556

Post-print

100 to 249	1,197	397	148	101	31	117
250 to 499	164	108	43	19	7	39
500 to 749	31	20	7	2	Nil	11
750 to 999	9	4	1	1	Nil	2
1000 to 5700	9	43	6	4	Nil	33
Total	98,198	21,872	8,411	7,853	2,850	2,758

The citation corpus of Dimensions.ai also provides two other important measures – ‘recent citations’ (the number of citations that were received by a publication in the previous two calendar years) and Field Citation Ratio (FCR) that is obtained “by dividing the number of citations a paper has received by the average number received by documents published in the same year and in the same Fields of Research (FoR) category” (<https://dimensions.freshdesk.com/support/solutions/>). In these two cases it has also been observed that OA publications are performing gradually better than the ‘close access’ publications with the increasing cut off values. For example, the threshold value ≥ 100 for recent citations, qualifies 550 publications of which 310 ‘close access’ and 240 ‘open access’ whereas the cut off mark ≥ 500 retrieves 54 publications consist of 8 ‘close access’ and 46 ‘open access’ published resources. Similarly, the FCR cut off value ≥ 25 skims up 1043 ‘close access’ and 678 ‘open access’ publications but the higher values like ≥ 50 lists 211 ‘close access’ and 244 ‘open access’ publications or the cut off value ≥ 100 shows 34 ‘close access’ and 122 ‘open access’ publications.

The socio-academic webspace also tells a story on the basis of altmetric data. A total of 24,308 contributions (16.52% of total publications with DOI) have received altmetric score (dataset available in table 5) ranging from 0.25 (at the lowest scale) to 5092.72 (at the highest scale). A total of 24,133 publications out of 24,308 (99.26%) are having recorded status statement for access types (close or open) and 175 publications are not available with access status. These 24,133 publications having altmetric score and access status shows that 14,569 contributions are in ‘close access’ group (60.37%) and 9,561 contributions are available through OA channels (39.63%). The OA resources that have received altmetric attention score (9,561 resources) are distributed in different routes like - green (3840 contributions), gold (3292 contributions), hybrid (1751 contributions), and bronze (678 contributions). Majority of the resources having altmetric score are within the score range from 0.25 to 99.999 (23,883 of 24,133 – 98.95%). An analysis of contributions having high altmetric score (≥ 100) shows that there is a total of 250 publications out of 24,133 (1.03%) having altmetric score ≥ 100 and the majority of such contributions are available through OA routes (207 in open access vs 43 in close access). If we set the scale in higher magnitude (altmetric score ≥ 500) then only 61 contributions pass that mark and all are available through OA routes (hybrid – 50; gold – 6; green – 4; and bronze- 1). A detail account of altmetric score ranges and OA types is given in table 9.

Table 9: Altmetric ‘score range’ blocks and access types of publications

Altmetric score range	Close Access publications	Open Access publications	Open Access categories			
			Green	Gold	Bronze	Hybrid
0.25 to 99.999	14,526	9,357	3,794	3,255	661	1,647
100 to 499	43	146	42	31	16	57
500 to 999	Nil	35	1	3	Nil	31

Post-print

1000 to 1,999	Nil	20	Nil	3	1	16
2000 to 2,999	Nil	03	1	Nil	Nil	2
3000 to 5100	Nil	03	2	Nil	Nil	1
Total	14,569	9,564	3,840	3,292	678	1,754

The purpose of this section is to understand the nature and the features of open access in the top most educational system of India (IITs) so that the strength, weakness, opportunities and challenges (SWOC) can be pinpointed for building an open access friendliness indicator. It is quite clear from the trend analysis as performed in this section that for a given institute the OA publication share, OA availability through proper open access licensing, citation share of OA publications and altmetric scores of OA publications can play major roles to understand the views, policies and attitude of an institute towards open access. The first two areas (OA publications share & OA licensing) can give us a direction about the growth of OA in general for an institute and its awareness about the importance of OA licensing to ensure rights of the stakeholders. The other two major areas (OA citation share & OA altmetric score share) may throw lights on the issues related with the availability of quality publications through OA channels. It is also important to note that each of these four major areas of consideration are again influenced by number of interconnected factors. A summary of such observations is tabulated as a decision table (table 10) for developing a distributed weightage based formula for open access friendliness indicator (OAFI) as discussed later in section 5.2.

Table 10: Decision table for identifying factors on the basis of SWOC

Area	Observations		Important factors	Area weightage
OA publications	Strength	Steady and continuous growth of OA over the years in all IITs	1. Overall OA share in total publications; 2. Gold & Green OA share as preferable routes; and 3. Green OA access through open access repositories (as best OA locations)	50
	Weakness	The highest growth recorded as 20.04%, not even 25% of the total publications		
	Opportunities	Availability of OA infrastructure for repository setup in all IITs		
	Challenges	To achieve OA share at least 50% by 2030		
OA licensing	Strength	Increasing awareness of OA licenses (almost 43% OA contributions are now available through OA licenses)	1. Share of all OA publications with licenses; and 2. Availability of gold & green OA publications with licenses	25
	Weakness	Still, more than 50% of resources are made available without OA licenses		
	Opportunities	Almost 63% of licensed OA contributions are made available though liberal OA		
	Challenges	Ideally, all OA resources need to be made available through OA licenses and the target needs to be set as 100% licensed OA by the year 2030		

Post-print

OA citations	Strength	Significant number of OA resources belong to the group of highly cited publications	1. Overall citations share by OA publications;	15
	Weakness	Overall citation ratio still shows 82% (close) : 18% (open)	2. Citations share by OA resources in moderately cited publication group;	
	Opportunities	Increasing number of OA resources are entering to highly cited publications group	3. Citations share by OA resources in highly cited publication group	
	Challenges	To increase overall citation ratio from 80/20 to 50/50 ratio by the year 2030		
OA socio-academic	Strength	Highly discussed publications in socio-academic webspace are almost all OA	1. Share of OA publications in total altmetric score of a given institution considering all publications; and	10
	Weakness	Around 16.50% resources have received altmetric attention and overall altmetric score ratio still shows 60% (close) : 40% (open)	2. Share of OA publications in altmetric score of a given institution considering publications more than average altmetric score	
	Opportunities	Number of OA resources in the group of highly discussed publications in socio-academic webspace are increasing steadily		
	Challenges	To increase overall altmetric score ratio from 60/40 to 25/75 ratio by the year 2030		

The area weightage values as given in table 10 are based on the priority order that requires to be followed by an academic institute in a developing country where the concept of open access is a relatively new one. The priority order is set according to the importance of the area (and the factors associated with that area) for ensuring success of open access in terms of availability, visibility and impact in the following order – 1) to ensure OA availability (accessibility of publications by a given institute in open access channels with a target to achieve close-open ratio at least 50:50 by 2030); 2) to support gold & green paths of OA, where the contributors/institutes have provisions to set rights for the stakeholders, in comparison to bronze & hybrid paths; 3) to pledge for proper OA licensing for every OA resources; 4) to make contributors aware of citation advantages for OA resources and to encouraging availability of quality publications through OA channels; and 5) to build awareness related to the new channels of visibilities of scholarly resources in the socio-academic webspace.

5.2 Indicator for Open Access Friendliness (OAF)

The scenarios and trends as discussed in the previous section make it evident that the formula for measuring Open Access Friendliness (OAF) needs to consider four major areas – i) OA share in publications; ii) compliance of OA publications with the formal licensing; iii) share of citations received by OA publications; and iv) attention received by OA publications from socio-academic web space. The first two parameters indicates growth and awareness of the researchers and the authority of a given institute, and the next two factors emphasis on the impact of OA publications in the academic spheres. Each of these major areas are again subdivided into groups to ensure proper distribution of weightage for the factors within the major ares. The selection of major areas and

Post-print

divisions of these areas by groups, and the weightage allocation for respective groups are based on reasons mentioned in table 10 under section 5.1. As a whole, the proposed groupings within the four major areas and the distribution of weightage in a 100 point scale is explained in table 11.

Table 11: 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: 25%)	OA license share (Group weightage: 15%)		Gold & Green license share (Group weightage: 10%)
OA impact: Citations (Area weightage: 15%)	OA citation share kind I (Group weightage: 5%)	OA citation share kind II (Group weightage: 5%)	OA citation share kind III (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 scope of the major areas and their respective weightage allocation is self explanatory but it is necessary to provide the scope and details of the formula for groups as mentioned under an area. The ten factors for calculating Open Access Friendliness Indicator (OAFI) are envisioned as below:

OA share: Represents percentage of OA share in the total publications during the period under study (i.e. 2010 to 2019). Here, total publication means the number of publications of a given institute that have received status (close/open) from Unpaywall. The value is obtained through the formula – Total OA publications during the period divided by the Total publications (with status) during the period and then multiplied by the weightage 25.

Gold & Green share: Provides measure to indicate gold and green OA share in total OA publications as these two routes of OA are generally considered more appropriate (dominated by the contributors) in comparison to bronze and hybrid routes (dominated by the publishers). The formula to measure value is - Total gold & green OA publications during the period divided by the Total OA publications during the period and multiplied by the weightage 15.

Repository share: The natural destination of a green OA publication is an open access repository (OAR). An OAR is considered as 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 own institutional repositories. The indicator is calculated by summing up total green OA papers of an institute available through repositories (as best OA location) and dividing the sum by total green OA papers made available by that institute, and then multiplied by the weightage value 10 (product is rounded into two decimal places). The range of years obviously the period of study (here 2010 to 2019).

OA license share: After the OA share, the next most important issue in open access movement is availability of resources against formal OA licenses (mainly Creative Commons licenses). This group 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 15.

Gold & Green license share: Similarly, the formula for this group is - Total gold & green OA publications with formal OA licenses (during the period) divided by the Total gold & green

Post-print

publications during the period and then multiplied by the allocated weightage value 10. The licensing of gold & green OA has given emphasis over the bronze & hybrid OA as because the later two routes of OA in general do not extend much scope to contributors in selecting OA license.

OA citation share kind I: It measures 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 dividing 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, sum of all citations received by only OA papers is done (B) and then it is divided by sum of all citations received for all publications during the period of study (A), and finally multiplied by the allocated weightage value for the group i.e. 5 ($B/A \times 5$).

OA citation share kind II: The average citation value as obtained in the previous group is utilized here as 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 – ($B/A \times$ weightage value for the group). The weightage value for this group is 5.

OA citation share kind III: This group emphasizes on the ratio of sum of ‘recent citations’ (the number of citations that were received in the last two calendar years) gathered by OA papers and sum of recent citations received by all papers in the top 1% of the papers of a given institute. A total of 1,37,460 contributions (out of 1,47,000 publications having DOI) are having recent citation data from Dimensions.ai. The top 1% publications of an institution is obtained through arranging publications by their citations in descending order and then selecting 1% from top (round up value).

OA altmetric share kind I: It measures the impact of OA publications of a given institute in terms of altmetric scores. This factor is important to consider due to the increasing influence of socio-academic webspaces on 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, sum of altmetric attention scores received by only OA papers is done (B), and then it is divided by sum of altmetric attention score by all publications of the institute during the period of study (A), and finally multiplied by the allocated weightage value for the group i.e. 5 ($B/A \times 5$).

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 through dividing sum of altmetric attention scores (A) by all publications with altmetric score (B). The average value is then utilized as cut off point for further processing. It goes for summing up 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 – ($D/C \times 5$). The weightage value for this group is 5.

This ten-factor based 100 point distributed weightage driven scale determines open access friendliness indicator (OAFI) value for each of these 16 IITs.

5.3 Institutional Scenarios

This part of the research report exemplifies the application of the OAFI (Open Access Friendliness Indicator) ranking framework with two IITs and then produces the ranked list covering all 16 IITs. These two IITs are selected from the list of top 16 IITs on the basis of their year of establishment – the oldest one i.e. Indian Institute of Technology Kharagpur (IIT, KGP, 1951) and the newest one i.e. Indian Institute of Technology Mandi (IIT, Mandi, 2009). The Indian Institute of Technology (ISM) Dhanbad (2016) or the Indian Institute of Technology (BHU) Varanasi (2012) are not considered in the newest category as these two are upgraded IITs from old institutions.

Table 12: Calculation of scores for Area I & Area II for two selected IITs

Area I: OA publications (weightage 50) & Area II: OA licensing (weightage 25)							
Factors		Values (round)		Factors		Values (round)	
SL	Element	IIT, KGP	IIT, Mandi	SL	Element	IIT, KGP	IIT, Mandi
A	Total publications	22,365	1,907	E	Total Gold & Green OA	2,179	343
B	Total publications with DOI	20,889	1,829	F	Total Green OA	1,168	179
C	Publications with close/open status from Unpaywall	20,633	1,822	G	Green OA available via repositories (as the best OA locations)	920	150
D	Total OA publications	2,899	390	H	Licensed OA resources	1,021	159
Factor 1: OA share (D/C * 25)		3.51	5.35	K	Licensed Gold & Green OA	824	143
Factor 2: Gold & Green OA share (E/D * 15)		11.27	13.19	Factor 4: OA license share (H/D * 15)		5.28	6.12
Factor 3: Repository share (G/F * 10)		7.88	8.38	Factor 5: Gold & Green license share (K/E * 10)		3.78	4.17
Area I scores		22.66	26.92	Area II scores		9.06	10.29

It is quite obvious from table 12 (depicting OA publications and OA license share) that the oldest and the largest in the IIT system i.e. IIT, KGP is much ahead of the newest IIT, Mandi in terms of numbers (total publications, OA publications, OA licenses etc) during the period of study (2010-2019) but the newest IIT has performed better than the oldest one in terms of ratio (OA publications share and OA license share). The same trend continues in OA citations share and OA altmetric share as reflected in table 13.

Table 13: Calculation of scores for Area III & Area IV for two selected IITs

Area III: OA impact - citations (weightage 15) & Area II: OA impact - altmetric (weightage 10)							
Factors		Values		Factors		Values	
SL	Element	IIT, KGP	IIT, Mandi	SL	Element	IIT, KGP	IIT, Mandi
A1	Total publication with	20,640	1,814	M	OA Publications with recent	51	6

Post-print

	citation status				citations >Lrc		
A2	Total publications with citation ≥ 1	18,528	1,588				
B	Total citations received by all resources that are having citations status	3,11,732	20,705	N	Total recent citations received by OA resources with recent citations >Lrc	9,191	842
C	Average citations per publication (Avg= B/A1)	15	11	P	Total publications with altmetric score	3.012	364
D	OA publications with citation ≥ 1	2,591	352	Q	Sum of altmetric scores for all publications with score >0	13,919	1,570
E	Citations received by OA resources with citation ≥ 1	49,477	5,880	R	Average altmetric score publication (AAvg= Q/P)	4.62	4.31
F	All publications with citation >Avg (Avg is the average citation value 'C' here)	5,726	537	S	Number of OA publications with altmetric score	889	143
G	Citations received by all resources with citation >Avg	2,35,389	15,798	T	Sum of altmetric scores for OA publications with score >0	7,749	989
H	OA Publications with citation >Avg	768	131	U	All publications with altmetric score >AAvg (AAvg is the average altmetric score here)	421	58
J	Citations received by OA resources with citation >Avg	38,341	4,870	V	Sum of altmetric scores for all publications with score >AAvg	10,746	1292
K	All publications with 'recent citations' >Lrc (Lrc is the lowest recent citations in the set of top 1% publications by recent citations - rounded)	207	19	X	Number of OA publications with altmetric score >AAvg	165	28
L	Total recent citations received by all resources with recent citation >Lrc	24,971	1,894	Y	Sum of altmetric scores for OA publications with score >AAvg	6,887	888
Factor 6: OA citation share kind I (E/B * 5)		0.79	1.42				
Factor 7: OA citation share kind II (J/G * 5)		0.81	1.54	Factor 9: OA altmetric share kind I (T/Q * 5)		2.78	3.15
Factor 8: OA citation share kind III (N/L * 5)		1.84	2.22	Factor 10: OA altmetric share kind II (Y/V * 5)		3.20	3.44
Area III scores		3.44	5.18	Area IV scores		5.98	6.59

The OAF indicator (OAFI), a sum of area I (three factors related to OA publications share), area II (two factors related to OA license share), area III (three factors related to the OA citations share) and area IV (two factors related to OA altmetric share) shows the newest IIT (IIT, Mandi) with an OAFI value 48.98 (out of 100) is more open access affable than the oldest IIT (IIT, KGP) with score 41.14 (out of 100). The ranked list of top 16 IITs in terms of OAFI shows a similar trend, where all top five IITs in the OAFI ranked list are comparatively newer as far as year of establishment is concerned (table 14). There may be many reasons for this but one particular fact is

Post-print

quite obvious that the OA culture (sharing, repository archiving and licensing of publications) have been adopted by these newer IITs right from the beginning of their journey in scholarly communication process. For example, Indian Institute of Technology Gandhinagar has scored 5 out of 5 in factor 8, which indicates how many publications in top 1% publications of a given institute (arranged by citations) are available through OA channels. Similarly, Indian Institute of Technology Mandi has scored 13.19 (out of 15) in factor 2 related with Gold & Green OA share and Indian Institute of Technology Bhubaneswar has secured 10.85 (out of 15) in factor 4 related to the OA license share. The final ranked list of 16 IITs under study by their respective OAF scores (table 14) shows that the Indian Institute of Technology Gandhinagar has secured the top positions in two areas (area I – OA publications share and area III – OA citations share), whereas Indian Institute of Technology Bhubaneswar has obtained the highest value in area II – OA license share, and Indian Institute of Technology Ropar has topped the list in area IV – OA altmetric score share.

Table 14: Ranked list of 16 top IITs by OAFI

Name of IIT	Area I (50)	Area II (25)	Area III (15)	Area IV (10)	OAF (100)	Rank
Indian Institute of Technology Gandhinagar	29.17	10.57	13.31	9.64	62.69	1
Indian Institute of Technology Bhubaneswar	24.44	17.63	6.06	8.35	56.48	2
Indian Institute of Technology Indore	26.93	15.07	6.42	7.98	56.40	3
Indian Institute of Technology Ropar	25.72	8.46	10.33	9.76	54.27	4
Indian Institute of Technology Hyderabad	25.61	7.60	9.81	8.72	51.74	5
Indian Institute of Technology Madras	23.67	11.55	5.88	8.48	49.58	6
Indian Institute of Technology Mandi	26.92	10.29	5.18	6.59	48.98	7
Indian Institute of Technology Bombay	24.91	9.38	6.46	8.14	48.89	8
Indian Institute of Technology Kanpur	25.21	7.74	4.21	8.17	45.33	9
Indian Institute of Technology Guwahati	22.57	11.98	4.65	5.80	45.00	10
Indian Institute of Technology Delhi	22.79	10.77	2.77	6.90	43.23	11
Indian Institute of Technology (BHU) Varanasi	22.63	11.39	2.49	5.74	42.25	12
Indian Institute of Technology Kharagpur	22.66	9.06	3.44	5.98	41.14	13
Indian Institute of Technology Patna	22.48	11.22	3.84	3.31	40.85	14
Indian Institute of Technology Roorkee	21.44	10.86	1.00	6.39	39.69	15
Indian Institute of Technology (ISM) Dhanbad	19.55	13.63	1.79	3.56	38.53	16

Conclusion

No ranked list is free from criticisms, and that holds equally true for this attempt to develop a ranking framework for OA support by Indian institutions engaged in higher education and research by taking top 16 IITs as samples. However, the effort of this research, as the first ever attempt to rank Indian institutes by their OA friendliness, is important in view of the following parameters – i) use of OA related, citations related and altmetric related datasets available through ODbL; ii) deep dive in OA related data with the help of data wrangling software like obtaining data through API call and extracting later as required from JSON-formatted data; iii) deep faceting of gathered datasets to measure previously unreported datasets like All publications >> OA publications >> Gold & Green OA publications >> Green OA publications >> Green OA available through repositories >> Facet by the name of the institutes; iv) inclusion of diverse datasets to get the picture beyond OA publication share such as OA citations share overall and also OA share in the highly cited article group; v) inclusion of altmetric data in the ranking framework to measure the popularity of OA publications of a given institute in the socio-academic webspace. The methodology as adopted here, and the ranking framework as developed, can now be extended easily for measuring and comparing OAF scores for other Indian and global institutes by using the data wrangling software, namely Openrefine.

References

- Aguillo, I., Ortega, J., Fernández, M., & Utrilla, A. (2010). Indicators for a webometric ranking of open access repositories. *Scientometrics*, 82(3), 477–486. <https://doi.org/10.1007/s11192-010-0183-y>
- Alperin, J. P., Babini, D. S. de, Fischman, G., & Consejo Latinoamericano de Ciencias Sociales. (2014). *Open access indicators and scholarly communications in Latin America*. <http://biblioteca.clacso.edu.ar/clacso/se/20140917054406/OpenAccess.pdf>
- Archambault, É., Amyot, D., Deschamps, P., Nicol, A., Provencher, F., Rebout, L., & Roberge, G. (2014). Proportion of open access papers published in peer-reviewed journals at the European and world levels—1996-2013. *Undefined*. <https://www.semanticscholar.org/paper/Proportion-of-Open-Access-Papers-Published-in-at-Archambault-Amyot/e75c3b9849995470af9c227cba3899f1e8e02919>
- Bar-Ilan, J. (2008). Informetrics at the beginning of the 21st century—a review. *Journal of Informetrics*, 2(1). <https://doi.org/10.1016/j.joi.2007.11.001>
- Gómez, 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. <https://doi.org/10.1108/10650750910961884>
- Hajjem, C., Harnad, S., & Gingras, Y. (2005). Ten-year cross-disciplinary comparison of the growth of open access and how it increases research citation impact. *IEEE Data(Base) Engineering Bulletin*, 28(4).
- Harnad, S., Brody, T., Vallières, F., Carr, L., Hitchcock, S., Gingras, Y., Oppenheim, C., Hajjem, C., & Hilf, E. R. (2008). The access/impact problem and the green and gold roads to open access: an update. *Serials Review*, 34(1), 36–40. <https://doi.org/10.1016/j.serrev.2007.12.005>
- Harnad, S., Brody, T., Vallieres, F., Carr, L., Hitchcock, S., Gingras, Y., Oppenheim, C., Stamerjohanns, H., & Hilf, E. R. (2004). The access/impact problem and the green and gold roads to open access. *Serials Review*, 30(4). <https://doi.org/10.1016/j.serrev.2004.09.013>

Post-print

- Leeuwen, T. van, Meijer, I., Yegros-Yegros, A., & Costas, R. (2018). *Developing indicators on open access by combining evidence from diverse data sources*.
<https://arxiv.org/abs/1802.02827>
- Maddi, A. (2020). Measuring open access publications: a novel normalized open access indicator. *Scientometrics*, 124(1), 379–398. <https://doi.org/10.1007/s11192-020-03470-0>
- Mukhopadhyay, P., Mitra, R., & Mukhopadhyay, M. (2021). Library carpentry: towards a new professional dimension (part i – concepts and case studies). *SRELS Journal of Information Management*, 58(2), 67–80. <https://doi.org/10.17821/srels/2021/v58i2/159969>
- Mukhopadhyay, P., & Mukhopadhyay, M. (2021). Library carpentry: towards a new professional dimension (part ii – automatic authority control to enhance retrieval). *SRELS Journal of Information Management*, 58(3), 135–155.
<https://doi.org/10.17821/srels/2021/v58i3/163890>
- Robinson-Garcia, N., Costas, R., & van Leeuwen, T. N. (2020). Open Access uptake by universities worldwide. *PeerJ*, 8, e9410. <https://doi.org/10.7717/peerj.9410>
- Robinson-Garcia, N., Leeuwen, T. N. van, & Torres-Salinas, D. (2020). Measuring open access uptake: data sources, expectations, and misconceptions. *Scholarly Assessment Reports*, 2(1), 15. <https://doi.org/10.29024/sar.23>

Post-print