An Affirmative Approach of Evaluating Open Access Repositories in India

Aditi Roy
Research Scholar, Department of Library & Information Science, University of North Bengal
Raja Rammohunpur West Bengal, India
Email- rs_aditi@nbu.ac.in

Dr. Saptarshi Ghosh
Professor, Department of Library & Information Science, University of North Bengal, Raja Rammohunpur, West Bengal, India
Email: sghosh@nbu.ac.in

Abstract:

The aim of the study is to find out the impact of Open Access Repositories (OAR) enlisted in Directory of Open Access Repositories (OpenDOAR) and Registry of Open Access Repositories (ROAR) with a comparative analysis of both the authoritative resources. For this study a total of 35 OARs enlisted in OpenDOAR out of which 24 is also registered in ROAR is selected for analysis by Web analytics tool Nibbler to evaluate the performances of the websites. "Ranking Web of World Repositories" (RWWR) ranking with its' search and retrieval strategies are also taken in consideration for this study.

The findings revealed that OpenDOAR and ROAR has their own pros and cons in aggregating data on Open Access Repositories. Although OpenDOAR claims that the listing of repositories in its directory are 'carefully reviewed and processed by a member of our editorial teams' but during this study it was found out that the claim is inaccurate in terms of URL specification, functionality and so on. ROAR registry includes a pool of features out of which some are malfunctional. However, ROAR has certain advantages over OpenDOAR, such as export facility, geographical identifiers (latitude and longitude) and total record count of the OARs.

Additional analytics using Nibbler showed that the best performing OAR was 'Knowledge Repository of Indian Institute of Horticultural Research' among the 35 studied and INFLIBNET is at top of the popularity rank.

This paper may help the administrators of the websites to improve the performance of the repositories and tries to draw the attention of the users to the advantages and drawbacks of OpenDOAR and ROAR.

Keywords: *Open Access Repositories (OARs), OpenDOAR, ROAR, Impact Analysis, Web Analytics Tools*

Introduction: Open Access Repositories (OARs) act as a gateway of information to researchers as it maximizes the visibility and impact of research outputs of different institutions as well as organizations[1]. Directory of Open Access Repositories (OpenDOAR) and Registry of Open Access Repositories (ROAR) are seemingly the most used directories of open access resources. OpenDOAR provides listing of the repositories using quality assurance parameters[2]. It is maintained by SHERPA Services, based at the Centre for Research Communications at the *University of Nottingham*[2.] ROAR is a searchable database with direct user submission. ROAR indexes the repository growth, contents, geographical location etc[3].

This study aims at finding the impact of the Open Access Repositories enlisted in OpenDOAR. Impact Analysis[11] is used in Computer Science, we preferred to implement impact analysis on Open Access Repositories (OAR) as OpenDOAR and ROAR those have most of the times being hyped by the authors in numerous literally warrants without measuring its true impact**.**

The **"Ranking Web of World Repositories" (RWWR)** is an initiative of ranking websites of world repositories operated by 'Cybermetrics Lab', a research group under CSIC, Spain (CSIC: Consejo Superior deInvestigaciones). The Cybermetrics Lab includes Quantitative Studies on scholarly activity along with scientific communication, resources and impact of Open Access Initiatives. The webometric ranking of the world repositories is based on some indicators such as the visibility of the repositories on the web, the transparency about the resources available, evaluation of the impact of the repositories and analysis of the usage of information through web data mining of log files[4].

Being the selective authoritative sources, both OpenDOAR and ROAR function as indicative services for the academic community. As the pivotal tertiary sources of information, these services warrant to administrative audit for their effectivity of the content they are holding. Hence, this study tries to a) evaluate the efficacy of the data on Open Access Repositories enlisted both in OpenDOAR and ROAR, b) a webometric measure of the performance and functionality of the OA repositories using an analytics tool and furthermore, the study evaluates the searching and retrieval facilities of 'Ranking Web of World repositories' collection.

Methodology:

The prime objective of the study is to analyze the Open Access Repositories in India registered in OpenDOAR and ROAR. The study is constituted with three parts,

- First part is the **web survey and analysis** of the OpenDOAR (**https://v2.sherpa.ac.uk/opendoar/search.html**) and ROAR (**http://roar.eprints.org/cgi/search/advanced**) and then comparing the data available in OpenDOAR and ROAR.

  The total number of OARs in OpenDOAR is 102 out of which two OARs are registered twice- 1. RAIITH (Research Archive of Indian Institute of Technology Hyderabad) and 2. NIRTIR (National Institute for Research in Tuberculosis Institutional Repository). To avoid cross classification these were eliminated and thus the total count of OARs in India stands 100 i.e., our population, with confidence interval 90%, margin of error 5% and population proportion 5%, that have been calculated for sample size determination by using the following formula[10],

$$n' = \frac{n}{1 + \frac{z^2 \times \hat{p}(1-\hat{p})}{\varepsilon^2 N}}$$

Where, z is the z score
ε is the margin of error
N is the population size
p̂ is the population proportion

Thus, the sample size stands as 35. These samples of OARs were chosen by using Simple Random Sampling method from OpenDOAR. Among these 35 OARs chosen from OpenDOAR for this study, 24 are also registered in ROAR.

- Second part is to analyze the **repositories' performance** using a web analytics tool i.e., **NIBBLER**[9] **(https://nibbler.silktide.com/)**. All the tests were done in between February 7-11, 2022.
- And lastly an **evaluative analysis** of **RWWR browsing and retrieval services** and **Transparent Ranking**[5] provided by RWWR is also included in the study with specific emphasis to the **Indian repositories registered in OpenDOAR and ROAR.**

Data collection and analysis:
List of repositories

Authoritative resource	Functional Repository			Non-Functional or Not Accessible Repositories	Total	Percentage (%)
	Total	Repository with wrong URL	Authorization Required			
OpenDOAR	91	22		9	100	91%
ROAR	82	32	1	21	103	79.6%

TABLE 1: List of repositories

Though the search executed in OpenDOAR and ROAR with the same parameter "Country- India" but the retrieval of OAR list is different from both the authoritative resources. After deduplication of repository list the total count of OARs in OpenDOAR is 100 and in ROAR the total number of OARs was 124 on 8/02/2022 which was exported as multiline CSV record and after deletion of duplicate record the number reduced to 103. One OAR from ROAR named "*Mahatma Gandhi University - Online THESIS Search*" requires authorization to navigate the website. It has been observed that in maximum cases the repositories overlap both in OpenDOAR and ROAR, but a countable few are there having exclusive inclusion in either OpenDOAR or ROAR.

State-wise distribution of OpenDOAR Data:

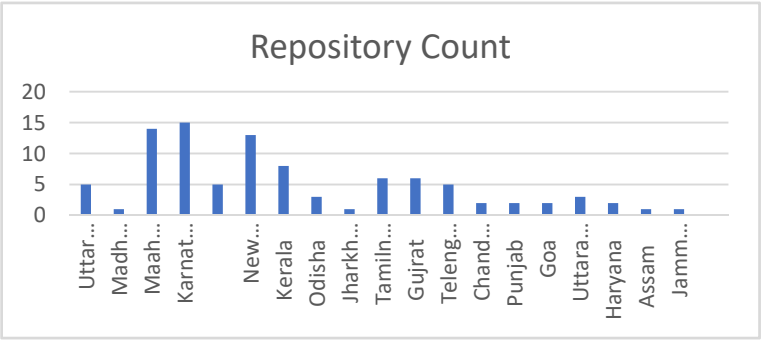

Figure: Country-wise Repository Count

As OpenDOAR does not provide the repository details with any geographical identifiers except country, so this study observed the state-wise distribution of OARs in India enlisted in the respective directory. The highest number of repositories are in Karnataka. Zone-wise the lowest performance is from the North-East region of India, there is only 1 repository from Assam enlisted in OpenDOAR.

Comparative analysis of OpenDOAR and ROAR:

Parameter	**OpenDOAR**	**ROAR**
URL	Present	Present
Export	Absent	Present
Interoperability	Present	Present
OAI-PMH	Present	Present
Record Count	Absent	Present
Geographical Identifiers	Absent	Present
Country	Present	Present
Subject Specific	Present	Absent
Record Creator	Absent	Present
Visualization	Static	Dynamic
Proliferation/User Engagement	User initiated but moderated	Only User initiated

TABLE 2: Comparative analysis of OpenDOAR and ROAR

The comparison between directory and registries invariably invites some bit of discrepancies while using predefined parameters. It digs out the pros and cons of the two authoritative resources on Open Access.

ROAR in comparison to OpenDOAR appears robust in data dissemination as it provides total record counts, geographical identifiers (latitude and longitude) as well as varied forms of data visualization.

On the other hand, OpenDOAR stands apart from ROAR in terms of subject representation, whereas the visualization of data is static in manifestation and export facility is absent which is considered as one of the prime features of any web directory for its evaluation.

Repositories performance analyzed using Nibbler:

Repository Name	Overall Score	Accessibility[1]	User Satisfaction[2]	Technology[3]
Knowledge Repository of Indian Institute of Horticultural Research (E-Repository@IIHR)	9.4	10	9.3	9.1
DSpace at Vidyanidhi	9.3	10	8.5	9.3
AIJR Preprints	8.7	9.8	7.9	8.9
CSIR-NCL Digital Repository	8.4	9.7	7.3	9.2
Institutional Repository - University of North Bengal (Institutional Repository NBU)	8.4	9.5	7.6	9.1
E Knowledge Center	8.1	10	7.3	9
Digital repository of West Bengal Public Library Network	8	9.8	8	9.1
DRS at National Institute Of Oceanography (DRS@nio)	7.9	9.1	7.2	8.1
Digital Knowledge Repository of Central Drug Research Institute (DKR@CDRI)	7.9	8.7	7.6	8.6
Electronic Theses and Dissertations at Indian Institute of Science (edt@IISc)	7.8	9.1	7.1	8.2
DSpace @ GGSIPU	7.7	8.1	7.5	5.9
DSpace@IMSC	7.7	9.3	8.1	8.3
DSpace at Indian Institute of Management Kozhikode (DSpace@IIMK)	7.5	9.7	6.9	8.6
DSpace@INFLIBNET	7.5	9.5	7.2	8.4
Dspace at IIT Bombay (DSpace@IITB)	7.5	9.5	7.3	9.1
DIR@IMTECH	7.2	9.3	6	7.3
Digital repository of Cochin University of Science & Technology (Dyuthi)	7.2	8.2	4.8	7.8
EPrints@IIT Delhi	7.1	9.2	7	8.2
National Aerospace Laboratories Institutional Repository (CSIR-NAL)	6.9	8.1	5.8	8.4

Dspace@NITR	6.9	8.9	6.5	8
Indian Academy of Sciences: Publications of Fellows	6.8	8.1	5.6	7.9
Electronic Theses and Dissertations of The Tamil Nadu Dr. M.G.R. Medical University	6.7	8.1	5.5	7.9
AMU Repository (Knowledge Repository)	6.6	8.3	5.5	8.2
Bhogawati Mahavidyalaya Institutional Repository	6.6	9.1	7.8	8.1
ethesis@nitr	6.5	8.2	5.2	7.5
Eprint@NML	6.4	7.9	4.8	7.2
Eprints @MDRF	6.4	8.2	4.8	7
DigitalLibrary@CUSAT	6.2	7	5	6.8
ARIES, Digital Repository	5.9	8	4.8	7.5
IR@NPL	5.8	6.4	5	6.6
IR@CECRI	5.7	6.4	4.9	6.6
IR@CGCRI	5.7	6.4	4.8	6.6
NEERI Institutional Repository (IR@NEERI)	5.7	6.4	4.9	6.6
IR@NEIST CSIR North East Institute of Science and Technology Open Access Institutional Repository (IR@NEIST)	5.6	6.4	4.8	6.6
IR@Central Leather Research Institute (IR@CLRI)	5.2	6.4	4.8	6.5

TABLE 3: NIBBLER Results

Nibbler Web Analytics tool measures performance of websites (non-IP address based) with four broad categories, namely, Accessibility, Experience, Technology and Marketing. Based on these four parameters Nibbler calculates overall scores out of 10. This study excluded Marketing Category as OARs are not directly associated with marketing. Hence the overall score derived by Nibbler might invite pitfalls in the score depicted. Most of the cases the marketing category provided marginal value in comparison to Open Access Repository sites corresponding to a business site.

The three parameters taken into consideration for the study comprise sub-parameters which were also been tested. Due to paucity of space and unnecessary cluttering of data those tests results are furnished as add-ons. The terminology used in this study as 'User Satisfaction' instead of 'Experience' (in NIBBLER) is rationalized because of the sub-parameters under experience mainly deals with user-interface and users' information behavior.
The superscript used in Table 4 under the categories represent the presence of sub-parameters under those categories.

Accessibility[1]	**User Satisfaction[2]**	**Technology[3]**
URL Format	Twitter	Printability
Headings	Printability	Metatags
Internal Links	URL Format	URL Format
Mobile	Popularity	Headings
Page Title	Internal Links	Internal Links
	Server Behavior	Server Behavior
	Amount of Content	Images
	Images	Mobile
	Mobile	
	Freshness	

TABLE 4: Nibbler Evaluation Parameters

Under Accessibility the five tests were done by Nibbler which corresponds to Information Access by the end user. Out of these five sub-parameters mobile accessibility, internal links and URL format are considered as the prime criteria for evaluation. The three prime criteria visible under accessibility are also manifested both in User Satisfaction and Technology. Up-to-date ness or freshness of the website comes under user satisfaction seems to be pivotal indicator for any website as it satisfies "Principle of Currency". The Technology category is divided into eight sub-parameters which manifest administrative metadata and the backend technology and algorithm used for development of the sites. Under this category the most important sub-parameters seem to be metatags and headings as they are intermingled with each other.

Ranking of the repositories:

Sl. No.	Repository Name	Popularity	RWWR RANK
1	DSpace@INFLIBNET	24953	Not available
2	Dspace at IIT Bombay (DSpace@IITB)	26342	2429
3	EPrints@IIT Delhi	28127	923
4	Electronic Theses and Dissertations at Indian Institute of Science (edt@IISc)	44283	907
5	DSpace @ GGSIPU	45070	Not available
6	Dspace@NITR	51795	1660
7	ethesis@nitr	53030	699
8	AMU Repository (Knowledge Repository)	58544	3885
9	DSpace at Indian Institute of Management Kozhikode (DSpace@IIMK)	69386	3060
10	Knowledge Repository of Indian Institute of Horticultural Research (E-Repository@IIHR)	121845	Not available
11	DigitalLibrary@CUSAT	128467	Not available
12	Digital repository of Cochin University of Science & Technology (Dyuthi)	130659	Not available
13	Indian Academy of Sciences: Publications of Fellows	160787	1040
14	CSIR-NCL Digital Repository	204059	3688
15	DSpace at Vidyanidhi	234029	Not available
16	DRS at National Institute Of Oceanography (DRS@nio)	255396	1557
17	National Aerospace Laboratories Institutional Repository (CSIR-NAL)	387243	3885
18	DSpace@IMSC	409222	Not available
19	Electronic Theses and Dissertations of The Tamil Nadu Dr. M.G.R. Medical University	432662	Not available
20	Digital Knowledge Repository of Central Drug Research Institute (DKR@CDRI)	436420	Not available
21	Institutional Repository - University of North Bengal (Institutional Repository NBU)	512101	1286
22	AIJR Preprints	704805	Not available
23	DIR@IMTECH	734263	Not available

24	Eprint@NML	868426	1176
25	ARIES, Digital Repository	1025304	Not available
26	IR@CECRI	3401402	2306
27	IR@CGCRI	3401402	3196
28	IR@Central Leather Research Institute (IR@CLRI)	3401402	3829
29	NEERI Institutional Repository (IR@NEERI)	3401402	3524
30	IR@NEIST CSIR North East Institute of Science and Technology Open Access Institutional Repository (IR@NEIST)	3401402	3557
31	IR@NPL	3401402	3387
32	Eprints @MDRF	5738489	3361
33	E Knowledge Center	6012460	Not available
34	Bhogawati Mahavidyalaya Institutional Repository	Not yet been ranked by Alexa	Not available
35	Digital repository of West Bengal Public Library Network	Not yet been ranked by Alexa	Not available

TABLE 5: Popularity Ranking (Nibbler) Vs. RWWR Transparent Ranking

The above table shows comparison between the ranking provided by Nibbler and RWWR. In NIBBLER's terminology the ranking is 'Popularity' which is none but the Alexa Global Ranking. RWWR ranking is the transparent ranking of World repositories. From the total population of 100 repositories, this study sampled out 35 OARs of which only 20 are listed in RWWR transparent ranking list till February 2022. The institutional repository of National Institute of Technology Rourkela topped in RWWR (out of 20) with the world rank 699. Inflibnet stands as the mostly used OAR in India out of studied 35 repositories according to **Alexa Global Ranking** but is not ranked under RWWR. Among the population of a total 100 Open Access Repositories enlisted in OpenDOAR, this study found that "Krishikosh" is the top Indian OAR with a rank of 47 in the RWWR ranking.

Results and discussion:

- From this study it has been observed that the percentage of functional repository is much more in OpenDOAR than in ROAR.

- OpenDOAR provides data on four categories 1. Repository Information, 2. Organization, 3. Open Access Policies and 4. Record Details. No data under "Open Access Policies" category is found except AIJR Preprints.
- Although OpenDOAR claims that the OARs enlisted in OpenDOAR are "*carefully reviewed and processed by a member of our editorial teams*" many functional OARs are unreachable because of wrong URLs. So, URL validation or regular updating of the URLs are suggested.
- It is traced that all the OARs enlisted under OpenDOAR, the bulk 'last update' is 12th January, 2022 whereas Nibbler shows different updation dates. The study identified that many of the repositories were not updated even for last ten years.
- Due to the facility of direct user submission in ROAR registry, it is visible that many of the OARs are enlisted multiple times resulting a total list of 124 Open Access Repositories in India. By de duplication of the data, the number is reduced to 103. Many of the URLs provided in the record details of the OARs seemed to be inaccurate.
- ROAR (http://roar.eprints.org/cgi/roar_graphic?cache=4568690) website provides additional features such as graphical representation of the data, which could not be visualised.
- Apart from transparent ranking, the RWWR provides continents as well as country-wise searching facilities. While using those facilities by the authors, the retrieved result always showed "***There are no date formats found in the db***".
- The ranking by Nibbler (source Alexa Global Rank) and the RWWR transparent ranking are poles apart from each other. The methodology used in both the cases are different; RWWR uses Visibility, Transparency and Excellence of the repositories whereas NIBBLER(Alexa) compares the popularity of the websites in respect to other websites to provide the rank depending on the traffic of the websites.

Conclusion:

This paper highlights the advantages and drawbacks of the Directory of Open Access Repositories (OpenDOAR) and the Registry of Open Access Repositories (ROAR) which may help the editorial teams and administrators of the respective directory services to evaluate and improve the deliverables and facilities provided by them. Implementation of web analytics tool for the study fosters to evaluate the performance of the websites of OARs and the results may guide the administrators to identify the weaknesses of the OARs in real time for ratifications of the pitfalls embedded therein. OARs are the contemporary knowledge carriers and thus

required to be in motion barring its stillness. This paper approaches the SWOT analysis of Open Access Repositories by redefining the 'Threat' as challenges that requires to be overcome for the generation of social goods.

References:

1. *Dattatraya, K. (2019). Institutional Repository in Open DOAR: Status Quo India. Library Philosophy and Practice, 5(1), 1–9.*
2. *https://v2.sherpa.ac.uk/opendoar/about.html*
3. *http://roar.eprints.org/information.html*
4. *https://repositories.webometrics.info/en/About_Us*
5. *https://repositories.webometrics.info/en*
6. *https://v2.sherpa.ac.uk/opendoar/search.html*
7. *http://roar.eprints.org/cgi/search/advanced*
8. *https://nibbler.silktide.com/*
9. *Chakravarty, R. (2019). National ETD repository evaluation using web analyser: A webometric analysis of Shodhganga, India. International Journal of Web-Based Learning and Teaching Technologies, 14(1), 54–68. https://doi.org/10.4018/IJWLTT.2019010104*
10. *https://www.calculator.net/sample-size-calculator.html*
11. *https://www.guru99.com/impact-analysis-software-testing.html*