



Some issues in the *Shodhganga* – A theses repository from India

Murari P. Tapaswi

Former Chief Scientist & Head, KRC, CSIR-National Institute of Oceanography, Goa,
Present Address: 4 Singbal Bldg, Tonca, Caranzale, 403002 Goa, India, Email: tapaswimurari@gmail.com

Received: 23 May 2023; revised: 08 June 2022; accepted: 25 June 2023

Shodhganga theses repository, a cooperative effort between universities and INFLIBNET has a digital collection of over 460000 theses. Metadata from *Shodhganga* on the theses on Marathi literature was downloaded to study the trends in this area of knowledge. While carrying out this study, many flaws in the metadata and relevant files on *Shodhganga* were noticed. Time lag in the submissions of theses by the participating Universities, duplicate records for one document, non-adherence to the cataloguing standards while entering metadata, absence of abstracts and apt keywords for the document being described, attributing theses to wrong departments/subject areas, garbled text in metadata and files, identification of similarity between items conferred with a degree and inconsistencies in the rendering of full-text file names and linking incorrect documents were noticed. This study showcases such specific examples from within the repository and suggests some solutions and recommendations to avoid the flaws in the repository.

Keywords: Theses, *Shodhganga*, Repository, Marathi

Introduction

The University Grants Commission (UGC) in the year 2009 published a Notification in the Gazette of India on 'Minimum Standards and Procedure for Award of M.Phil./Ph.D. Degrees Regulations' and the same was revised in the year 2016¹. Information and Library Network (INFLIBNET) Centre was chosen as an aggregator to receive theses from Indian universities for developing a national ETD repository. INFLIBNET developed a repository by the name '*Shodhganga*' to upload these theses and make them available to the entire scholarly community to (1) avoid duplication and repetition in research, (2) raise the standard and quality of research by making them available online, (3) provide financial assistance to digitize backlists, and (4) have a control on the plagiarism². As of today, there are over 460000 theses on this repository and the collection grows almost every day.

Since its launch, *Shodhganga* has become very popular among the research community and has also received international awards and recognition³. In India, this is a unique example of a cooperative effort in developing a database repository by higher education institutions. To capture data on the time scale of the Ph.D. research process, the repository encourages researchers, guides, university representatives, research cell, library, and INFLIBNET Centre to upload metadata

and submit theses to *Shodhganga*, or share the details with the INFLIBNET for upload.

The successful existence of this repository for over a decade and ever-increasing size necessitates its critical evaluation for its value addition. This study showcases some observations and puts forward recommendations for increasing the richness of the repository. These cases are identified from a limited dataset downloaded from *Shodhganga*. It would be worth examining other records for needed possible corrections, if any.

Review of literature

The unique database has provided an opportunity for researchers in library science to carry out bibliometric studies using data from *Shodhganga*⁴⁻⁶ in the recent past.

Nicholas, *et al*⁷ have grouped repositories into institutional, subject, and format. According to their grouping, *Shodhganga* is a 'format repository' that stores outputs of a particular type such as electronic theses. A framework, FAIREST principles for the assessment of research repositories⁸ suggests parameters such as findability, accessibility, interoperability, reusability, engagement, social connections, and trust. All these parameters focus on the usefulness of the repository. However, the study presumes that the metadata of the records and its link

to the full-text are in order and therefore do not look into the quality of the metadata and the content, which is equally important.

In another study⁹ comparing metadata creation strategies by three institutions, the authors highlight the importance of effective metadata management as the input of metadata within DSpace, the software used for *Shodhganga* too, with mistakes that are especially expensive to correct. Irrespective of repository metadata, clean metadata is a basic requirement of all databases. Besides the most common parameters used for database evaluation are timeliness, uniqueness, validity, and richness of metadata¹⁰⁻¹⁴.

Timeliness evaluates the timely availability of the information, uniqueness checks the occurrence of duplicate records, validity looks into the quality of the records concerning the cataloguing standards, error-free data, orthographical errors (system of spelling in a language - nonstandard usage of spellings), etc. Whereas richness deals with the accuracy in document analysis while describing metadata, the ability to finely describe a resource (with keywords, abstract, etc), index on all possible query fields for better retrieval, etc.

Methodology

A subset of records on Marathi literature and linguistics was downloaded from *Shodhganga* by searching on the Departments of Marathi and also finding out relevant literature in the Schools of language and literature of the respective Universities during 16-18 April 2023. The Marathi language is largely spoken in the State of Maharashtra and the border areas in the adjoining states. The research work done on this area of study was expected to be in the Universities from this region. However, stray theses on this subject were also located in the universities from Aligarh, Ahmedabad, Baroda, Bathinda, Delhi, Gandhinagar, Hyderabad, Jhunjhunu, and Mysore.

In all, data on 1849 theses were collected. They were then sorted out based on the conventional classification of Marathi literature. During the process some records were identified as not a part of Marathi Literature but some other subject. This resulted in identifying 262 records containing research on the 'Saint and religious literature in Marathi'.

The full-text files as needed were opened for some records to have clarity on the metadata. During this process, the anomalies in the full-text files were

noticed. All the anomalies that were observed during the process were compiled. They were further grouped based on the quality parameters as suggested in previous studies¹⁰⁻¹². A critical examination that provided insight into *Shodhganga* has been recorded in the observations. The author also extends recommendations to overcome the observed problems based on his experience in maintaining repositories.

Observations

The observations are recorded based on the quality parameters such as timeliness, uniqueness, validity, and richness as suggested in earlier studies¹⁰⁻¹².

Timeliness

The importance of the availability of the information as soon as it is due is highly valued. The UGC regulation¹ makes it obligatory for the degree-awarding institution to submit an e-copy of the Ph.D. thesis to INFLIBNET for hosting the same on *Shodhganga* before the announcement of the award. The records of *Shodhganga* provide data on the date of completion of work as well as the date since when the item is available. A time lag was identified by computing the difference between the date of completion and its availability (Table 1). The data in Table 1 indicates about 50% of the theses are made available within two years of their completion of work. It may be possible that the process of the award of a degree gets delayed in the universities and therefore the availability of theses after completion of work to the extent of 50% is reasonable but submissions beyond this period could certainly be classed as 'delayed' submission.

There are four Universities, viz, Dr. Babasaheb Ambedkar Marathwada University (BAMU), Aurangabad, Savitribai Phule Pune University (SPPU), Shivaji University (Shivaji), Kolhapur and Swami Ramanand Teerth Marathwada University (SRTMU), Nanded conferring over 100 degrees after 2009 for studies in Marathi literature and linguistics. A radar graph (Fig. 1) on these 4 universities indicates that the SRTMU is much quicker in uploading their theses, followed by BAMU. Graphical analysis for other universities was not carried out simply because the number of theses on Marathi literature added by them to *Shodhganga* is less than 100 since 2009.

The University of Mumbai in the State of Maharashtra is one of the oldest universities (166 years!). However, only 20 theses on Marathi literature were identified to have been added. As of the date of

collection of data for this study, *Shodhganga* had <1500 theses from this university. In the recent past, a piece of news has appeared¹⁵ on the students losing their precious time during research since uploading theses by this university is not being done.

Similarly, the Department of Marathi in the Rashttrasant Tukadoji Maharaj Nagpur University on *Shodhganga* is non-existent. A webpage¹⁶ on the University website indicates over 115 PhDs have been awarded by the Department of Marathi.

Uniqueness

Uniqueness indicates an absence of duplicate entries. Within a subset of 262 theses on 'Saint and religious literature in Marathi' at least 12 duplicate (4.6%) records (Table 2) are noticed. This appears to be on the higher side for a quality database. Moreover, such instances are noticed for the theses from 6 universities, so the problem appears to be universal.

Cataloguing Standards

The metadata entered in the database must follow certain standards for bibliographic description. While the data entry standards for the *Shodhganga* have been showcased², they do not seem to have adhered to while populating records. This study, therefore, tries to correlate the standards as described in the User's

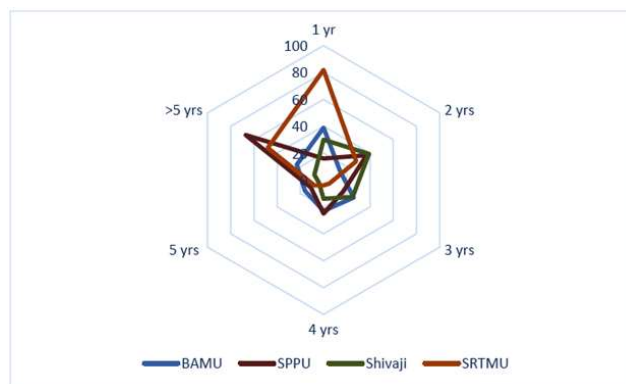


Fig. 1—Time lag for submission of theses to *Shodhganga*

Table 1—Time lag for submission of theses to *Shodhganga*

University	Upload time (years)							Theses after 2009	Theses before 2010	Theses without date	Total theses
	0	1	2	3	4	5	>5				
BAMU	12	39	14	26	23	16	23	153	182	0	335
Goa	0	1	1	1	0	0	0	3	9	0	12
Gondwana	0	0	5	1	0	2	0	8	0	0	8
KUD	2	1	3	2	0	1	0	9	13	0	22
North_Maha Baroda	2	8	13	17	10	2	2	54	0	0	54
Amravati	0	0	1	0	1	3	6	11	70	0	81
SPPU	2	16	37	17	25	11	67	175	369	15	559
Shivaji	6	30	39	25	14	4	8	126	213	0	339
SNDT	1	3	7	2	2	1	4	20	83	0	103
Solapur	8	4	0	0	0	0	0	12	0	0	12
SRTMU	31	82	28	5	4	8	48	206	93	0	299
TiMaVi	5	12	0	0	0	1	0	18	0	1	19
Mumbai	0	4	9	1	0	0	5	19	2	0	21
Deccan Coll PG & Res Institute	0	1	1	0	0	0	0	2	0	0	2
EFLU, Hyderabad	0	0	0	0	0	0	0	0	1	0	1
University of Mysore	0	0	0	0	0	1	0	1	0	0	1
Central Univ of Punjab	1	0	0	0	0	0	0	1	0	0	1
Central Univ of Gujarat	0	0	1	0	0	0	0	1	0	0	1
Gujarat University	0	0	0	0	0	0	0	0	1	0	1
Shri Jagdishprasad Jhabarmal Tibarewala University	7	7	3	0	1	0	0	18	0	0	18
JNU	0	0	0	0	0	0	1	1	3	0	4
Total	77	208	162	97	80	50	164	838	1042	16	1896

Table 2 — Occurrence of duplicate records from a subset on 'Saint and religious literature in Marathi'

University	Researcher	URL	URL	Reason
Dr. Babasaheb Ambedkar Marathwada University	Bokil, Suman	http://hdl.handle.net/10603/117580	http://hdl.handle.net/10603/123006	Spelling difference in title, Author name rendering
Dr. Babasaheb Ambedkar Marathwada University	Mundhe, Anil Vyankatrao	http://hdl.handle.net/10603/150463	http://hdl.handle.net/10603/132794	Spelling difference in title, Author name rendering
Karnatak University	Khairnar, S R	http://hdl.handle.net/10603/21696	http://hdl.handle.net/10603/100070	Spelling difference in title, Author name rendering
Sant Gadge Baba Amravati University	Mundhe G	http://hdl.handle.net/10603/184999	http://hdl.handle.net/10603/227105	Spelling difference in title, Author name rendering
Savitribai Phule Pune University	Keskar, Vijay Vishnupant	http://hdl.handle.net/10603/3812	http://hdl.handle.net/10603/156266	Spelling difference in title, One entry under SPPU and the other under Univ Pune
Savitribai Phule Pune University	Pathak, Yashwant Tryambak	http://hdl.handle.net/10603/153738	http://hdl.handle.net/10603/175482	Spelling difference in title, One entry under SPPU and the other under Univ Pune
Savitribai Phule Pune University	Kale, Shailendra Pandrinath	http://hdl.handle.net/10603/169005	http://hdl.handle.net/10603/206723	Spelling difference in title, Author name rendering
Savitribai Phule Pune University	Koli, Sanjaykumar Shankar	http://hdl.handle.net/10603/3775	http://hdl.handle.net/10603/163095	Spelling difference in title, Author name rendering
SNDT Womens University	Harmalkar, K P	http://hdl.handle.net/10603/161206	http://hdl.handle.net/10603/153748	Spelling difference in title
SNDT Womens University	Belsare, Kishori	http://hdl.handle.net/10603/253530	http://hdl.handle.net/10603/110406	The data entry operator didn't check the previously existing record
Swami Ramanand Teerth Marathwada University	Khair Pratap Jalindhar	http://hdl.handle.net/10603/139155	http://hdl.handle.net/10603/138945	The data entry operator didn't check the previously existing record
Swami Ramanand Teerth Marathwada University	Nivruttirao, Bodke Nita	http://hdl.handle.net/10603/249419	http://hdl.handle.net/10603/242227	The data entry operator didn't check the previously existing record

Guide and Tutorials for *Shodhganga*². Here are some observations on the basic metadata elements:

Name of Researcher/Guide: INFLIBNET recommends the name of the researcher and guide be rendered as Surname, First name Middle name. It also exhibits a couple of examples. A large number of the data records that were downloaded do not adhere to these standards. Some examples of metadata in the records and correct rendering are showcased (Table 3). Interestingly, in certain cases, the person entering data has failed to identify the Surname from the name elements of the researcher and guide in the document. In several records, suggested punctuation has not been adhered to. In case the thesis has a co-guide, both the names should have been entered separately opening an additional row but the name of the co-guide has been entered along with the guide's name itself (last row in Table 3).

Shodhganga showcases² examples of author entries that have surname elements. However, a large number of south Indian names and some north Indian names do not have a surname element. While no such names have been noticed in the data so collected during this study, it is hoped that the names without the Surname

element have been entered appropriately in other records.

Title and Alternative Title: The data pulled out from *Shodhganga* was of the theses of Marathi literature. The titles of these theses are in the Devanagari script. Metadata for most of these theses have been entered in transliterated form (in Roman script). While doing so, no standards of transliteration appear to have been followed. Transliteration of Devanagari text in Roman script without the use of any standards has made reading these titles difficult. A variety of spellings for a single re-occurring word 'vaaGmay' (वाङ्मय) in the metadata of the *Shodhganga* repository for theses on Marathi literature have been showcased (Table 4).

In one of the entries (<http://hdl.handle.net/10603/144811>) the first characters of words (Shakespearecha Marathi atakanvar alela arinam) in the title have not been typed making it meaningless. In another case (<http://hdl.handle.net/10603/253826>), the metadata entry operator has taken the liberty to replace the word (Panchatantra) in the title with another term (Panchyajanya). Such gross errors result in senseless reading and adversely reflect the richness of the database.

Table 3 — Names of the Researcher/Guide in the metadata

University	URL	Researcher	Expected Rendering	Guide	Expected Rendering
Swami Ramanand Teerth Marathwada University	http://hdl.handle.net/10603/241835	Divakar, Kulkarni Martand	Kulkarni, Martand Divakar	Shankar, Sadhu Yeshwant	Sadhu, Yashwant Shankar
Swami Ramanand Teerth Marathwada University	http://hdl.handle.net/10603/242921	Ganesh, Shinde	Shinde, Ganesh	Madhav, Gadekar	Gadekar, Madhav
Swami Ramanand Teerth Marathwada University	http://hdl.handle.net/10603/244388	Shivajirao, Jadhav Sambhaji	Jadhav, Sambhaji Shivajirao	Madhavrao, Kadam Jagdish Pundlikrao,	Kadam, Jagdish
Swami Ramanand Teerth Marathwada University	http://hdl.handle.net/10603/242938	Champati, Hankare Santosh	Hankare, Santosh Champati	Suryawanshi Nanasaheb	Suryawanshi, Nanasaheb Pundalikrao
Swami Ramanand Teerth Marathwada University	http://hdl.handle.net/10603/242257	Shrinivasrao, Nerkar Shamrao (नेरकर शामराव श्रीनिवासराव)	Nerkar, Shamrao Srinivasrao	Prakash, Medakkar (मेदककर प्रकाश)	Medakakar, Prakash
Dr. Babasaheb Ambedkar Marathwada University	http://hdl.handle.net/10603/99277	Nerkar, Shamrao Shreenivasrao	Nerkar, Shamrao Shreenivasrao	Vyawahare, Sharad	Vyawahare, Sharad
Sant Gadge Baba Amravati University	http://hdl.handle.net/10603/180630	Mahale C D	Mahale, Chaya Dileep	Kolate R	Kolte, Ramakant
Solapur University	http://hdl.handle.net/10603/73421	Gurav Devidas Shrimant	Gurav, Devidas Shrimant	Pujari, Suhas Govindrao	
Savitribai Phule Pune University	http://hdl.handle.net/10603/185123	Eknath, Dongare Mangal	Dongare, Mangal Eknath	Deshmukh, Lata Keshav	
Shivaji University	http://hdl.handle.net/10603/146829	B N, Turanbekar	Turambekar, Balwant Natha	Nasirabadkar, L R	
Shri Jagdishprasad Jhabarmal Tibarewala University	http://hdl.handle.net/10603/396062	Bagul Pradnya Bhagwan	Bagul, Pradnya Bhagwan	Pratibha Shankar Ghag And Gorwadkar Vinod Govind	Ghag, Pratibha Shankar; Gorwadkar, Vinod Govind

Table 4 — Variations that show how the word 'vaaGmay' (वाङ्मय) (or its variants) has differently been spelled (Roman Transliteration) in the Marathi literature records

Sl. No.	Variation	Sl. No.	Variation
1	vadamahin	12	vagndmay
2	vadamay	13	vagnmy
3	vadayin	14	Vandamay
4	vadgamaya	15	vandmay
5	vadgmaya	16	vangamay
6	vadgmyain	17	vangmay
7	vadmay	18	vanmay
8	vadmya	19	wandgmay
9	vadumay	20	wanghmay
10	vadymay	21	wangmay
11	vagmay	22	wangmy

A case has been identified (<http://hdl.handle.net/10603/110267>) wherein the data entry has been carried out by wrongly translating the title of the thesis from Marathi to English.

There are a few records, where the title in Devanagari script has been entered along with the title

in Roman script. The input form of *Shodhganga* provides a field for 'Alternative title'. This could have been used for entering metadata in the original script. An alternative title is defined¹⁷ as equivalent to the title in another language.

Abstract and keywords: The elements such as abstract and keywords finely describe the resource and add value to the database¹². The records for the theses on Marathi Literature in *Shodhganga* hardly have an abstract. In almost all university records, 'abstract' has not been added to the metadata. In a large number of cases, the thesis for which the record is being created does not have an 'abstract' page. Describing the documents by writing abstracts and defining keywords is the job of a traditional cataloguer. So, in the absence of an abstract in the thesis, the cataloguer is expected to write an abstract by going through the contents and other preliminary pages. While this is the general case, some sporadic theses are noticed to have an abstract that has not been entered in the metadata of the thesis (Table 5).

Table 5 — Records in *Shodhganga* having Abstracts but not added to the metadata field

University	URL
Savitribai Phule Pune University	http://hdl.handle.net/10603/350081
Dr. Babasaheb Ambedkar Marathwada University	http://hdl.handle.net/10603/368175
Dr. Babasaheb Ambedkar Marathwada University	http://hdl.handle.net/10603/246246
Dr. Babasaheb Ambedkar Marathwada University	http://hdl.handle.net/10603/132794
Swami Ramanand Teerth Marathwada University	http://hdl.handle.net/10603/241810
Swami Ramanand Teerth Marathwada University	http://hdl.handle.net/10603/86293

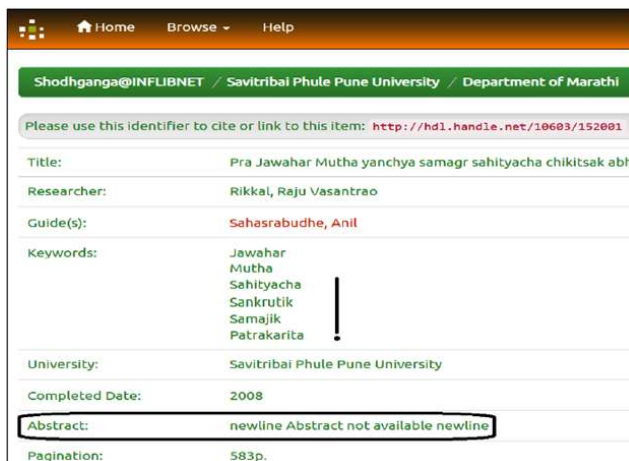


Fig. 2 — Representative example of the absurd and non-standard keywords rendering. Also, metadata for abstract field contains information not expected

The abstracts field of the metadata at times has a text indicating the availability of abstract in full-text. But the purpose of this field is not to indicate that. In a large number of cases, this field is filled with the term 'newline', whose meaning and purpose is unknown.

While some records' metadata are populated with keywords, they appear to be absurd, non-standard, and irrelevant (Fig. 2). INFLIBNET¹⁸ suggests the use of Library of Congress Subject Headings (LCSH) for additional keywords besides the predefined three levels of a subject from the dropdown. The use of LCSH appears to be not worthwhile for the studies on Marathi literature and might be true for other subjects that have regional influence.

Other metadata: Irregularities are often noticed in the data entry (or no entry) of metadata especially for the fields such as completion date (guideline¹⁸ suggests typing year alone) of the thesis, pagination, etc.

Rendering the wrong department and language of the thesis: As indicated earlier, data that existed in the Department of Marathi in various Universities was downloaded. It was then realized that some records though grouped under the Department of Marathi, the theses do not study Marathi literature. The attribution of theses to the Department of Marathi is probably done because those theses are in the Marathi language. However, the language of the work need not necessarily be the topic of study. The wrong entries that were identified are listed in (Table 6).

A separate metadata field exists to write the language of the thesis. The selection of the language is from the drop-down menu¹⁸. In many records, the language field is filled with 'other'. While some records do have 'Marathi' as the language.

Garbled metadata/files

Clean metadata and flawless files in attachments also add up to the richness of the database. However, it has been noticed that in some situations, the metadata itself has garbled text and/or the files that are attached to the metadata record. Representative examples identified with this limited study are listed in (Table 7).

Similarity

The very first two introductory paragraphs of the User Guide² explain the purpose of the repository - to avoid the duplication and repetition of the research work and raise a hope that the *Shodhganga* would help in overcoming the duplication. However, similarity in the title and contents clearly show overlapping work for which degrees are awarded. Just within a small subset of theses, two instances have been noticed (Table 8). INFLIBNET also provides access to the similarity check software (commonly termed a plagiarism checker) to the member libraries. The similarity check software probably has a language constraint, in a sense, it cannot find similarities in non-English text and therefore the similarity in titles and topics remained untraced.

Inconsistency in full-text files

The User Guide² suggests splitting full-text of the thesis into different files before it is added to the record. The file naming conventions have also been standardized to have uniformity among the database records, conveying the content of the file, etc. However, instances have been noticed that the participating universities are casual in adhering to the guidelines and

Table 6 — Theses attributed to the Department of Marathi but in reality, degrees offered in other subjects/disciplines

University	Year	Title	URL	Expected Department*
SNDT Womens University	1993	Bharataril Prathamik Staravaril Anoupacharik	http://hdl.handle.net/10603/173557	Education
SNDT Womens University	1992	Raigad jilhyatil satvichya 7 vi chya vidyartchya galteecha chikitsak abhyaas	http://hdl.handle.net/10603/154051	Education
SNDT Womens University	1976	Bruhanamubee mahanagarapaalikeecya praathamik shaalaateel shishikaannnya vyaavasaayik kautubik saamaajik va aathik paristhiteeca cikitsapuun abhyaas	http://hdl.handle.net/10603/153769	Education
SNDT Womens University	1991	Pandharpurchi vaari ek samajshastriya abhyas	http://hdl.handle.net/10603/161280	Sociology
SNDT Womens University	1999	Hindustani shastriya sangitatil geybandhancha sangit Ratnakaracha sandarbhat vivechak abhyas	http://hdl.handle.net/10603/119787	Music/Fine Arts
Sant Gadge Baba Amravati University	2008	Kanishta Mahavidyalayatil Vidhyarthyanchi Bhavanik Budhimata Ani Paryavarniya Mudde Yanchyatil Sahasambadhache Vishleshanatmak Adhyayan	http://hdl.handle.net/10603/20286	Education
Sant Gadge Baba Amravati University	2008	Bharat Bangladesh Sambadha	http://hdl.handle.net/10603/20284	Political Science
Sant Gadge Baba Amravati University	2008	Banjara Samajachya Badalya Rajkiya Janivanche Vishleshanatmak Adhyayan	http://hdl.handle.net/10603/20285	Political Science
Sant Gadge Baba Amravati University	2002	Tarunya Pithikewar Shankhbhashmachya Abhyash (तरुन्यपीटिकेवर शंखभष्माचा अभ्यास)	http://hdl.handle.net/10603/12806	Ayurved
Sant Gadge Baba Amravati University	2008	Krushni Sudharna	http://hdl.handle.net/10603/26284	Geography
Sant Gadge Baba Amravati University	2008	Akola Zilyatil Beldar Jamaticha Abhyas	http://hdl.handle.net/10603/23798	Sociology
Sant Gadge Baba Amravati University	2009	Samved anni sanskrit strotravadgmyatil sangit: tulnatmak adhyayan	http://hdl.handle.net/10603/22727	Sanskrit
Shivaji University	1978	Bhudargad talukyatil shet majuranacha samajik aarthik abhyas	http://hdl.handle.net/10603/144493	Economics
University of Mumbai	2010	Bhandara Jilhyatil Bhikaryanchi Avastha Maharashtra Rajya_Ek Vyashti adhyayan	http://hdl.handle.net/10603/311194	Sociology
Swami Ramanand Teerth Marathwada University	2004	Amravati zilhyache krida sanchalnalaya ani sharirak shikshan mahavidyalaye ek chikitsak abhyas	http://hdl.handle.net/10603/241797	Physical Education

*Based on the title page, the topic of study

Table 7 — Garbled metadata, files

University	URL	Observation
Dr. Babasaheb Ambedkar Marathwada University	http://hdl.handle.net/10603/150463	Garbled text in files
Dr. Babasaheb Ambedkar Marathwada University	http://hdl.handle.net/10603/132794	Garbled text in files
Savitribai Phule Pune University	http://hdl.handle.net/10603/268892	Garbled metadata in the title
Solapur University	http://hdl.handle.net/10603/73421	Garbled text in files
Swami Ramanand Teerth Marathwada University	http://hdl.handle.net/10603/244733	Garbled metadata in keywords
Swami Ramanand Teerth Marathwada University	http://hdl.handle.net/10603/241797	Garbled metadata in keywords

Table 8 — Theses having similarities in Title and Contents

Case	Thesis 1			Thesis 2			Thesis 3		
	University	URL1	Title (Year of completion)	University	URL2	Title (Year of completion)	University	URL3	Title (Year of completion)
1	Swami Ramanand Teerth Marathwada University	http://hdl.handle.net/10603/240142	Jatyavaril ovigitatun disun yenare kautubmik bhavdarshan (2004)	Swami Ramanand Teerth Marathwada University	http://hdl.handle.net/10603/244733	Jatyavaril ovigitatun disun yenare kautubmik bhavdarshan (2012)	-	-	-
2	Dr. Babasaheb Ambedkar Marathwada University	http://hdl.handle.net/10603/96260	Mahanubhav niyatalkatil lekhanacha chikitsakh abhyas 1955 te 2000 (2007)	Shivaji University	http://hdl.handle.net/10603/146187	Mahanubhaviy niyatkalikancha vangmayin abhyas(2008)	Savitribai Phule Pune University	http://hdl.handle.net/10603/259571	Mahanubhav niyatkalik_ek chikitsakh abhyas (2013)

Table 9 — Irregularities in the full-text attachments

University	URL	Irregularity
Goa University	http://hdl.handle.net/10603/13170	File with the name Chapter_1 contains Chapter 7
Karnatak University	http://hdl.handle.net/10603/105841	File with the name Chapter_1 contains complete thesis other than preliminary pages
Savitribai Phule Pune University Solapur University	http://hdl.handle.net/10603/268892	The file with the name 'Abstract' opens the acknowledgments page
Shri Jagdishprasad Jhabarmal Tibarewala University	http://hdl.handle.net/10603/396062	Single file with all pages of the thesis
		03_certificate of supervisor.pdf file has the name of the researcher and not the supervisor!

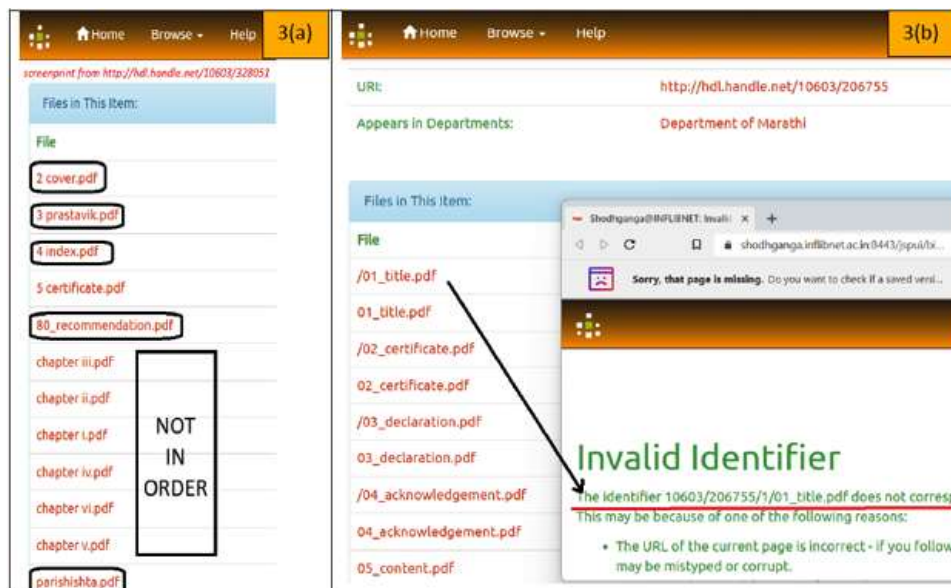


Fig. 3 — (a) Inconsistency in the file naming and sequence of files. (b) File names prefixed with a slash (/) do not open.

giving file names inconsistently (Fig. 3a), plugging files without a conventional sequence, opening different content than what it is named, etc. Some examples are also showcased in (Table 9).

A very large number of records from SPPU have dummy full-text files, for unknown reasons, with

a slash (/) before the file name (Fig. 3b). On click, they don't open and give an error as 'Invalid Identifier'. Of course, SPPU is not the only University with this problem. This has also been noticed sporadically in the records of other Universities too.

Recommendations

Timeliness

INFLIBNET should monitor the timeliness and insist on participating Universities for prompt uploads. Upload should at least be done in the year of the award of the degree. This would largely resolve the duplication of the study areas and save the time of budding researchers. Linking timely submissions to the ranking and accreditation will bridge the gap to a considerable extent and better address the issues arising from similarity.

Duplicate records and metadata verification:

The stepwise analysis of metadata entry will have to be studied to find out the genesis of the duplicate records to curb the problem. Importing data using a .csv file is the most convenient and easy option which is permitted by the DSpace software. By this, the generation of duplicate records would reduce almost to zero. The verification and editing process of the DSpace records is time-consuming and therefore expensive⁹. From the number of errors that have been pointed out from the small dataset, it is presumed that no verification/authentication has been done at any stage. The input if done from a .csv file, helps in quick verification too before upload as the data is in the columnar format where the error prominently stands out. Once the metadata is imported, then each record is edited to add the full-text files related to a particular record.

Cataloguing standards

INFLIBNET encourages universities to allow the researchers, guides, departments, etc., responsible for the entry of metadata in *Shodhganga*. However, based on the repeated occurrence of similar errors in multiple records hint that the entry is mostly done by the designated single person. If the data entry work is further distributed as indicated above, the number of errors will increase multi-fold making the verification and validation (if done) process time-consuming. There is a need to follow up with the designated person to curb repeated errors until the work is carried out to a certain level of standards. An option of batch upload (using a .csv file) accelerates the process, reduces the chances of errors, and provides an opportunity for quick verification in the .csv file itself. The introduction of an automated data validation system is the third option that INFLIBNET should give a try.

Fields in the metadata such as year (award, completion), pagination, keywords, and language of

the thesis should be made mandatory as in many cases these are observed to be empty. It is not known whether Marathi has been added at a later stage in the drop-down options for the language of the thesis field in the input form. *Shodhganga* must give an option to select from all languages that are in the Eighth Schedule of the Constitution of India, if not yet done.

Transliteration vs translation:

There are various methods of transliteration (https://en.wikipedia.org/wiki/Devanagari_transliteration). Among these, the Harvard-Kyoto system appears to be the simplest and is being used commonly. The use of a standard format makes it easier to read and convey meaning.

The language of the database is English. So, basically, should there not be a translated title for the document that is in non-English? The users who do not know the language of the document, but have an interest in the subject will not know if such a work exists. Consideration should be given to writing the title in the original language in the script of that language in the title metadata field and translated title in the alternative title field. Transliteration only helps the pronunciation of non-English terms without understanding their meaning. This recommendation needs discussion.

Keywords

Assignment of keywords using tools such as LCSH and/or by the author himself/herself is debated for over a decade¹⁹. Findings support that both, the author-assigned keywords and controlled vocabularies complement one another. *Shodhganga* recommends the use of LCSH for deriving keywords. This method is hardly satisfying for the concepts and ideas of regional importance, for example, Indian literature. In the absence of suitable terms in LCSH (or even otherwise), it is good to derive keywords from postulated elementary categories (Discipline, Entity, Action, Property, Place, and Time) as suggested in the POstulate-based Permutated Subject Indexing (POPSI)²⁰.

Full-text content

If the abstract is a part of the thesis, the data entry could be speeded up by using an abstract written by the researcher. However, in the absence of this, it is the responsibility of the indexer that (s)he writes an abstract for the work so that the information seeker gets an idea about the thesis from the metadata itself.

Unicode - the standard in keying the non-Roman script came into place in the year 1991. The thesis written in non-English script should be mandatorily typed using Unicode. This would resolve the problem of garbled full-text files. It is strange to see that the researchers are still typing their theses in non-standard encoding fonts.

The UGC Notification¹ should also make it mandatory for the universities to have all the theses in a standard format that include abstract too and non-English text typed in Unicode. The standard format should also suggest that the title page should contain the subject area for which the thesis is being submitted. This should largely resolve the problem of understanding the subject of the thesis during metadata entry and all the theses written in Marathi will not be ascribed to the subject of Marathi literature (or the Department of Marathi).

There appears to be a revision in the uploading file rules and naming convention. The 'Creation of metadata and uploading theses'¹⁸ page makes it mandatory to give a file name as '80_Recommendation' to the chapter that contains recommendations, conclusions, summaries, or future findings. The term "Future findings" is non-standard. It is probably being used for the concept of "Areas of study for the future" which at times is spelled out at the end of current studies. Moreover, the file name 'Recommendations' for the chapters that do not ascribe recommendations (but the thesis ends with a chapter on the conclusion, summary, etc.) itself is incorrect, misleading, and deviating from the principle of mnemonics that have been rightly argued on page 6 of the User Guide². The reason of this deviation is not known. All academics and researchers know very well that the last chapter of the thesis has all this. It is strongly proposed to drop the idea of this file as recommendation.

Shodhganga is one of the success stories in India. It contains rich information and benefits every individual in the academic and research field in general and budding researchers in particular. Error-free metadata and full-text will make it a richer database repository for other countries to imitate.

References

- 1 UGC, HRDC. Minimum standards & procedure for award of M.Phil./ Ph.D. degree, Regulation, 2009 (Amendment 5 July 2016). The Gazette of India, Part III Section 4(278) (5 July 2016). 1-12. [https://www.ugc.gov.in/pdfnews/4952604_UGC-\(M.PHIL.-PH.D-DEGREES\)-REGULATIONS,-2016.pdf](https://www.ugc.gov.in/pdfnews/4952604_UGC-(M.PHIL.-PH.D-DEGREES)-REGULATIONS,-2016.pdf).
- 2 Manojkumar, K and Arora, J, *Shodhganga: Digital Repository of Indian Electronic Theses and Dissertations - User's Guide and Tutorials*, (INFLIBNET), (ND) 1-8. <https://Shodhganga.inflibnet.ac.in:8443/jspui/sguserguide>
- 3 Anon, *Shodhganga* NDLTD Award to the Director and Scientist of INFLIBNET Centre on Aug 2017 (Washington). Available at https://Shodhganga.inflibnet.ac.in/newmoredetails/Shodhganga_Album/NDLTD/index.html (Accessed on 20 May 2023).
- 4 Paul S and Maity B K, Research trends in library and information science in India during 2011 to 2018. In: *Information Systems Management [Working Title]* (IntechOpen; London), 2023. <https://www.intechopen.com/online-first/86502>
- 5 Ghatowar N K and Kalita D, An empirical evaluation of research on "library management" at the doctoral level in India: A study of the Last 50 years from 1971 to 2020. *Open Information Science*, 7(1) (2023). DOI: 10.1515/opis-2022-0138
- 6 Kousar A, Reference accuracy in Indian library and information science theses, *Journal of Indian Library Association*, 59(1) (2023) 25-30. <https://www.ilaindia.net/jila/index.php/jila/article/view/1440>
- 7 Nicholas D, Rowlands I, Watkinson A and Jamali H R, Digital repositories ten years on: What do scientific researchers think of them and how do they use them? *Learned Publishing*, 25(3) (2012) 195-206. DOI: 10.1087/20120306
- 8 d'Aquin M, Kirstein F, Oliveira D, Schimmler S and Urbanek S, FAIREST: A framework for assessing research repositories. *Data Intelligence*, 5(1) (2023) 202-241. https://doi.org/10.1162/dint_a_00159
- 9 Chapman J W, Reynolds D and Shreeves S A, Repository metadata: Approaches and challenges, *Cataloging & Classification Quarterly*, 47(3-4) (2009) 309-325, DOI: 10.1080/01639370902735020
- 10 Rittberger M and Rittberger W, Measuring quality in the production of databases. *Journal of Information Science*, 23(1) (1997) 25-37. <https://doi.org/10.1177/016555159702300103>
- 11 Beall J and Hearn S, Introduction: Bibliographic database quality, *Cataloging & Classification Quarterly*, 46(1) (2008) 1-3, DOI: 10.1080/01639370802182711
- 12 Oury C, Assessing the ISSN register: Defining, evaluating, and improving the quality of a shared international bibliographic database, *Cataloging & Classification Quarterly*, 55(7-8) (2017) 588-605, DOI: 10.1080/01639374.2017.1354115
- 13 Kiraly P, Measuring metadata quality. Ph.D. Thesis. Georg-August-Universität Göttingen. (2019) 168pp. DOI: 10.13140/RG.2.2.33177.77920
- 14 Kiraly P, Measuring metadata quality. <https://www.rd-alliance.org/sites/default/files/attachment/peterkiraly-measuring-metadata-quality-rda14-poster.pdf>
- 15 Pandit N, Research scholars in a fix after MU delays theses uploading on *Shodhganga*, *Hindustan Times*, 2 May 2023. <https://www.hindustantimes.com/cities/mumbai-news/mumbai-university-negligence-phd-students-forced-to-change-research-topic-after-university-fails-to-upload-theses-on-digital-repository-101682966289051.html> (Accessed on 20 May 2023).

- 16 Anon. Rashtrasant Tukadoji Maharaj Nagpur University. <https://nagpuruniversity.ac.in/index.php/department/Contentpage/index/R1BObW9laFBHdjVYZWxmTTZDTT0%3D/RXY4aHZ1dENTZnhTWIZZPQ%3D%3D/en> (Accessed on 20 May 2023).
- 17 Anon. 6.3. Other title information. In: *Cataloger's Reference Shelf. CONSER Cataloging Manual*. https://www.itsmarc.com/crs/mergedprojects/conser/conser/module_6_3_ccm.htm (Accessed on 20 May 2023).
- 18 Anon. Creation of metadata and uploading theses into *Shodhganga*. https://Shodhganga.inflibnet.ac.in/newmoredetails/metadata_thesis.html (Accessed on 20 May 2023).
- 19 Strader C R, Author-assigned keywords versus Library of Congress Subject Headings, *Library Resources & Technical Services*, 53(4) (2009) 243-250. <https://doi.org/10.5860/lrts.53n4.243>
- 20 Anon. Unit 17. Techniques for subject indexing. IGNOU, 2017 84-86. <http://hdl.handle.net/123456789/33116>