Global electronic thesis and dissertation repositories – collection diversity and management issues

This article discovers the collection diversity of electronic thesis and dissertation (ETD) repositories based on key parameters such as regional distribution, subject classification, language diversity, etc. and identifies the critical management issues of the ETD repositories related to collection management, software management, content management and metadata policies. The ETD repositories were identified in the Directory of Open Access Repositories (OpenDOAR). The required data were manually collected from the OpenDOAR and websites of repositories to achieve the prescribed objectives of the study. The data were later tabulated, analysed and interpreted using simple arithmetic techniques.

The study was limited to the ETD repositories available in the OpenDOAR, and findings cannot be generalized across repositories and directories. It provides insights about ETD repositories worldwide, highlights their critical management issues and suggests mechanisms for their sustainable growth and development. This article is purely based on research and its findings are valid for scholars, faculty members, institutions – as well as administrators and managers of the ETD repositories.

Keywords

digital repositories; digital repositories management; electronic theses and dissertations; ETD repositories; collection diversity; OpenDOAR

Introduction

> 'Theses and dissertations are the most useful kinds of invisible scholarship and the most invisible kinds of useful scholarship because of their high quality and low visibility.'[1]

Electronic theses and dissertations (ETDs) are primary, rich, unique and valuable sources of scholarly information, which is the outcome of focused, extensive and in-depth research work of several years, involving intellectual labour by scholars and their supervisors. These ETDs were historically always kept under lock and key by vigilant information managers, possibly to avoid plagiarism and theft. Access to these valuable and scholarly sources was restricted to a few users within the four walls of the library of each institution, and most libraries do not lend theses and dissertations through inter-library loan. The closed access system affected their usage badly and these valuable sources mostly remained undiscovered, unutilized and uncited. The emergence of electronic sources, developments in open access (OA) and the creation of the digital repositories all make possible the best use of scholarly information sources including theses and dissertations. These repositories have become showcases of the intellectual achievements of scholars and their institutions by making their research output available globally in various forms, including ETDs. Since the digital repositories have started to archive ETDs, their usage statistics have been positively affected. The ETD repositories not only increase the visibility of ETDs but also increase institutional research impact and their ranking in the scholarly world. Data from these repositories suggest a dramatic increase in the use and citation of doctoral theses in current research activity.

FAYAZ AHMAD LOAN

Documentation Officer
Centre of Central
Asian Studies
University of Kashmir

UFAIRA YASEEN SHAH

PhD Scholar
Centre of Central
Asian Studies
University of Kashmir

Literature review

ETDs are a major topic of interest for researchers worldwide. A good number of studies have been conducted on ETDs. The current literature focuses on two aspects – the growth and development of ETDs and their management issues.

The origin of ETDs can be traced back to the first meeting held in Michigan, USA in 1987, organized by UMI and attended by representatives from Virginia Tech, the University of Michigan, and two small software companies—Toronto-based SoftQuad and Michigan-based Arbortext. Later, ETDs started to emerge in various institutions of the developed world. As a result, the National Digital Library of Theses and Dissertations (NDLTD) was established in 1996.[2] The NDLTD is a collaborative effort by the world's universities to create, archive, distribute and access ETDs. The ETD repositories flourished internationally and membership of the NDLTD increased significantly.[3] Zhang and associates[4] found a significant increase in the usage of ETDs in Korea by both national and international patrons. Sale[5] studied the impact of mandatory policies on ETD acquisition in Australia and found that only 15% of ETDs were deposited in repositories voluntarily, whereas mandatory policies increased the deposit rate to 100%. Sugita and Murakami[6] examined the university library policies on theses and dissertations in Japan and found that the university libraries had begun to deposit and disseminate ETDs to institutional, subject-specific and content-specific repositories.

> 'repositories of ETDs are now becoming common in universities of all countries across the world'

Ghosh[7] examined the developments in the ETDs scene in India to explore the possibilities for creating a national repository for the deposit, discovery, use and long-term care of research theses in an OA environment. The India University Grants Commission made it mandatory for all universities to deposit a copy of each thesis in the National ETD repository, Shodhganga, in 2009, however, the universities did not initially take it seriously.[8] Despite all issues, repositories of ETDs are now becoming common in universities of all countries across the world.

At the end of 2013, the Directory of Open Access Repositories (OpenDOAR) listed more than 1,400 repositories – of which more than 50% archived ETDs.[9] With the advancement of ETD repositories, many issues emerged. Looi and Yeng[10] proposed an ETD framework to capture and preserve the intellectual output of Malaysian universities and discuss various issues such as archiving, preservation, accessibility, scalability, security, searchability and copyright issues of ETDs. Jin[11] analysed the ETD repositories in China and identified 'grey issues' like metadata standards, submission issues, software selection, content formats, copyright protection, fair use, access and preservation, among other issues. Al Salmi[12] evaluated the status of ETDs in the university libraries of Gulf countries and concluded that university libraries in this region have the necessary infrastructure for ETD programmes, but they face technological, administrative and legal barriers. Yiotis[13] also found various grey issues relating to ETD repositories such as copyright, plagiarism, costs and preservation. Similarly, Juznic[14] argued that besides the preservation and availability of ETDs, plagiarism and other forms of cheating at all levels are also a big concern for universities.

> 'the metadata elements had a significant level of inconsistency and variation'

Park and Richard[15] assessed the metadata element sets of electronic theses and dissertations used at the Canadian academic institutional repositories. The results revealed that the metadata elements had a significant level of inconsistency and variation. Perrin and her associates[16] examined the problems that arose after the transition from a physical to the electronic collection of ETDs and presented documentation solutions for preservation and curation. Steele and Sump-Crethar[17] conducted a study on university repositories in the United States and provided valuable suggestions for bibliographic description and vocabulary control of ETDs. Ndungu[18] identified and analysed the

challenges faced in the bibliographic control of theses and dissertations in Kenya. The study found delays in the lack of consistency and uniformity in bibliographic records. Schopfel and Rasuli[19] argued the applicability of the concept of grey to ETDs and concluded that, '"greyness" remains a challenge for ETDs, a problem waiting for the solution through the application of the FAIR (findability, accessibility, interoperability, and reusability) principles'.

In the initial stage, many institutions started to establish ETD repositories worldwide after these developments and consequently created national ETD repositories like ETHOS (E-Thesis Online Service). These national-level repositories collaborated with the international ETD platforms such as NDLTD and OATD (Open Access Theses and Dissertations) to make the research in these ETDs more visible and useful. However, there emerged various issues with the growth and development of ETDs in collection management that needed to be addressed with the other developments.

Research design

The objectives of the study were to discover the collection diversity of the ETD repositories, based on key parameters such as regional distribution, subject classification, language diversity, etc. and to identify the critical management issues of the ETD repositories related to collection management, software management, content management, and metadata policies. The OpenDOAR was selected as a source for identifying the ETD repositories. All the repositories archiving ETDs (1,938) have been selected for the study. The required data were manually collected from the OpenDOAR and websites of these repositories in December 2017 to achieve the objectives of the study. The data were later tabulated, analysed and interpreted using simple quantitative techniques to reveal the findings.

'Next to journal articles, ETDs are the most frequent document type found in OA repositories'

Data analysis
Contents archived

Next to journal articles, ETDs are the most frequent document type found in OA repositories listed in OpenDOAR. Out of 3,504 repositories, 1,938 (55.30%) accept the submission of ETDs (Figure 1). The findings are consistent with the study conducted by Loan and Sheikh[20] on health and medical repositories, in which the results also reveal that the highest number of repositories store articles, followed by theses.
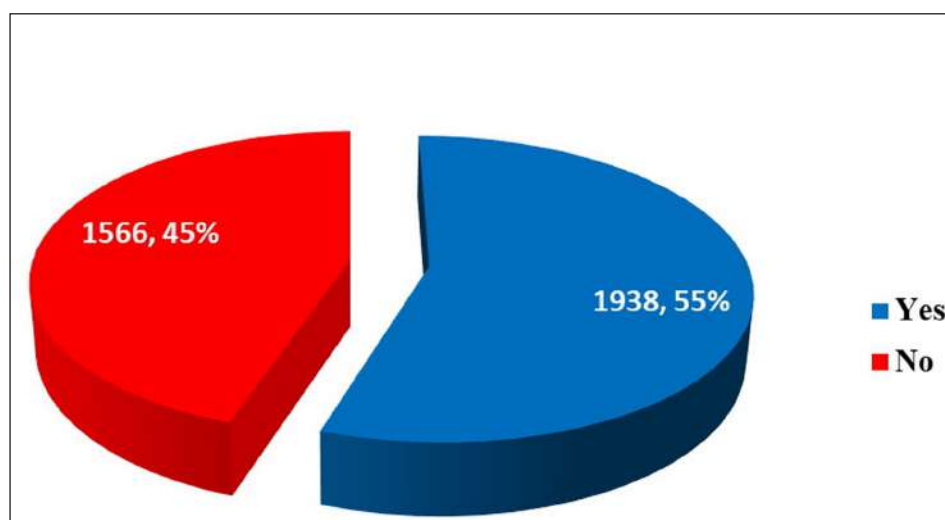

Figure 1. ETDs accepted by the repositories

Growth-rate

The number of ETD repositories from 2006 to 2017 shows a constantly increasing trend. In the year 2006, the number of ETD repositories was only 418 whereas the number had increased to 1,938 by the end of 2017. The year 2008 had the highest growth rate, of about 39% increase in the number of ETD repositories. The growth rate was high in the initial years, but in recent years it has gone slightly down (Figure 2). The findings are in tune with some of the earlier studies[21] wherein it was revealed that the number of repositories has increased exponentially since 2006.
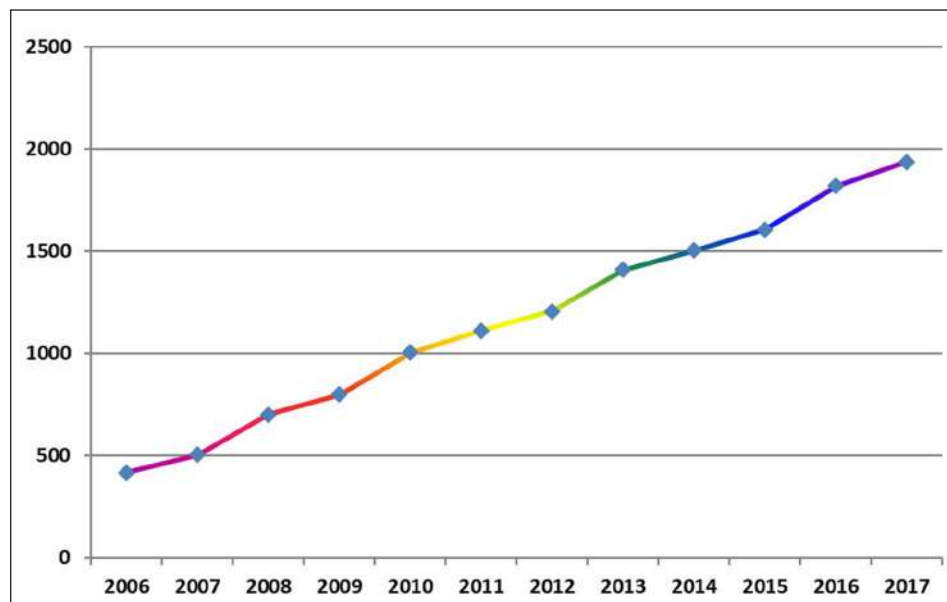

Figure 2. Growth of ETD repositories

Regional distribution

The highest number of ETD repositories is contributed by Europe (843, 43.5%), followed by Asia (408, 21.05%), North America (302, 15.58%), South America (213, 10.99%) and Africa (116, 5.99%) respectively (Table 1). The earlier study also found that Europe has contributed the most repositories, followed by North America, in OpenDOAR.[22]

S. No.	Continents	Number	Percentage
1.	Europe	843	43.50
2.	Asia	408	21.05
3.	North America	302	15.58
4.	South America	213	10.99
5.	Africa	116	5.99
6.	Australasia	43	2.22
7.	Central America	13	0.67
	Total	**1,938**	**100**

Table 1. Contribution of the continents in the ETD repositories

Countries contribution

The United States tops the list of countries by contributing about 12% (234) of the total repositories, followed by Germany with 7.38% (143) and Japan with 5.52% (107) respectively. Other countries especially France, UK, Spain, Turkey, Italy, Brazil and Indonesia also make significant contributions to the ETD repositories (Table 2). The findings are in tune with the study conducted by Loan and Sheikh[23] to a great extent. They revealed that the highest number of repositories is contributed by the USA, followed by Japan and the UK. However, developed countries contribute more than developing countries.

S. No.	Countries	Number	Percentage
1.	United States	234	12.07
2.	Germany	143	7.38
3.	Japan	107	5.52
4.	France	82	4.23
5.	United Kingdom	80	4.13
6.	Spain	68	3.51
7.	Turkey	60	3.10
8.	Italy	59	3.04
9.	Brazil	59	3.04
10.	Indonesia	58	2.99
11.	Others	988	50.99
	Total	**1,938**	**100**

Table 2. Contribution of the countries in the ETD Repositories

Content language

The ETD repositories archive content in 35 languages. Most of the repositories (67.75%, 1,313) accept contents written in English followed by Spanish (13.78%, 267), German (8.88%, 172) and French (7.43%, 144) respectively. It is also revealed that most of the repositories are multilingual, archiving content in more than one language (Table 3). English is the dominant language and the majority of the repositories archive content in the English language, which is also confirmed by the present study.

S. No.	Languages	Number	Percentage
1.	English	1,313	67.75
2.	Spanish	267	13.78
3.	German	172	8.88
4.	French	144	7.43
5.	Japanese	110	5.68
6.	Portuguese	108	5.57
7.	Chinese	70	3.61
8.	Italian	64	3.30
9.	Turkish	60	3.10
10.	Norwegian	44	2.27
11.	Russian	44	2.27
12.	Indonesian	43	2.22
13.	Swedish	43	2.22
14.	Ukrainian	32	1.65
15.	Arabic	31	1.60
16.	Polish	30	1.55
17.	Dutch	24	1.24
18.	Croatian	23	1.19
19.	Greek	17	0.83
20.	Others (15)	127	6.55

Table 3. Content-language of the ETD repositories

Classification of repositories

The repositories have been classified into four categories – institutional, disciplinary, aggregating and governmental. The majority of the ETD repositories are institutional (93.71%) whereas the disciplinary repositories and aggregating repositories contribute a very small percentage of 3.2 and 2.3 respectively (Table 4).

S. No.	Type	Number	Percentage
1.	Institutional	1,816	93.71
2.	Disciplinary	62	3.20
3.	Aggregating	45	2.32
4.	Governmental	15	0.77
	Total	**1,938**	**100**

Table 4. Classification of the ETD repositories

Subject Coverage

It has been found that most of the ETD repositories are multidisciplinary (71.93%) in nature, archiving ETDs of more than one subject area, whereas only 28.07% of the repositories are subject-specific, covering repositories on particular subjects only (Figure 3).
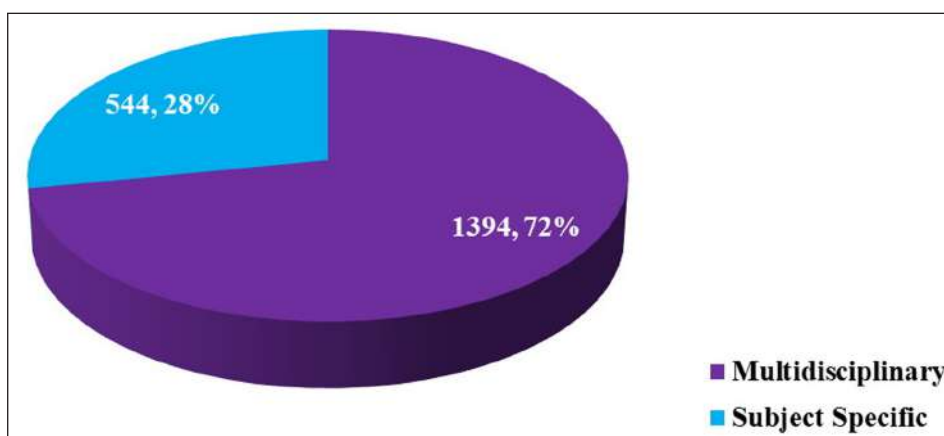

Figure 3. ETD repositories

Operational status

The operational status of the ETD repositories shows that almost 96% of them are operational, 2.53% are in trial mode, while only 1.6% are broken (not functional) (Table 5). The study conducted by Yaseen, Loan and Jan[24] also confirmed that more than 96% of all the ETD repositories are fully operational whereas a small percentage of repositories are available on a trial basis (1.9%, 25) and 2% are non-functional.

Type	Number	Percentage
Operational	1,857	95.82
Trial	49	2.53
Broken	32	1.65
Total	**1,938**	**100**

Table 5. Operational status of the ETD repositories

Software Used

DSpace is the most used software, operational in 50% of the ETD repositories. Other software brands used by the ETD repositories are EPrints (12.85%), Digital Commons (5.83%), OPUS (3.51%) and WEKO (2.89%) respectively (Table 6). DSpace is the first choice of administrators to manage content in digital repositories all over the world. In 2011, DSpace was used by more than 1,000 digital repositories[25] and since then the number is increasing constantly.

S. No.	Software	Number	Percentage
1.	DSpace	956	49.33
2.	EPrints	249	12.85
3.	Digital Commons	113	5.83
4.	Unknown	84	4.33
5.	OPUS	68	3.51
6.	WEKO	56	2.89
7.	Others	412	21.26

Table 6. Use of software in the ETD repositories

Policies

Policies are very important to the operation of the repositories. Metadata policies are the set of policies related to the information describing items in the repository. Preservation is a crucial element in the process of managing electronic information resources in digital repositories. Content submission policies provide information about the content that can be archived in the repositories. The data clearly shows that more than half of the repositories have explicitly undefined metadata policies (54.95%, 1065), content submission policies (51.39%, 996) and preservation policies (55%, 1,058), which is a very serious issue in the management of ETDs (Table 7).

'Preservation is a crucial element in the process of managing electronic information resources in digital repositories'

Policies	Metadata Policy	Content Submission Policy	Preservation Policy
Unknown	74 (3.82)	74 (3.82)	70 (3.61)
Unstated	15 (0.77)	13 (0.67)	163 (8.41)
Undefined	1,065 (54.95)	996 (51.39)	1,058 (54.59)
Not-analysed	538 (27.76)	538 (27.76)	538 (27.76)
Defined	246 (12.69)	317 (16.35)	109 (5.62)
Total	**1,938 (100)**	**1,938 (100)**	**1,938 (100)**

Table 7. Metadata policies in the ETD repositories

Discussion

Electronic theses and dissertations (ETDs) are the most frequent document type found in open access repositories after journal articles. They are accepted by more than 55% of the repositories in the OpenDOAR. ETDs are also perhaps the most important research products after journal articles. Journal articles are mostly 'shined and polished products' of ETDs. Therefore, most of the digital repositories enrich their collection by accepting the ETDs. The online availability of the ETDs is a very good sign for the optimum use of these resources. Researchers worldwide can take advantage of the research conducted at any institution in the world, along with other benefits. The growth of ETD repositories has also shown a constantly increasing trend since 2006.

Many countries created national repositories, like Shodhganga (India), to archive the ETDs of all disciplines at the national level and made regulations for scholars to compulsorily deposit the ETDs in the national repository to facilitate use and avoid plagiarism and duplication. Many institutions have created repositories to archive subject-specific ETDs as well. These efforts have increased the number of ETD repositories worldwide. Further, all the continents contribute to the ETD repositories as per their capacities and developed continents like Europe, and countries like the United States, top the list. However, the movement of archiving the ETDs in repositories is not limited to any specific region or country but has crossed boundaries. Many developing countries have followed the steps of developed nations to create digital repositories and archive contents, including ETDs.

These ETD repositories archive content in 35 languages and most of the repositories accept ETDs written in the English language. The valuable information can be in any language, not necessarily in English. Further, many countries in the world, like China and Iran, also prefer to research using their national languages. Therefore, the ETDs written in other languages may also be archived for optimum use by the present and future generations. The other positive signs are that the majority of the ETD repositories are created by reputed institutions, are archiving quality content and are fully operational. DSpace is the most used software, operational in 50% of the ETD repositories. Besides having extraordinary features, the DSpace community provides healthy support for creating digital repositories worldwide. It is the prominent reason that most of the ETDs have opted for DSpace for content management.

'ETD repositories archive content in 35 languages and most accept ETDs written in the English language'

The findings show that more than half of the repositories have explicitly undefined metadata policies, content submission policies, and preservation policies and these are very serious issues in the management of ETDs.

Conclusion

The ETD repositories have come up with many strengths, weaknesses, opportunities and challenges for the Library and Information Management profession. The strengths need to be fully utilized, the weaknesses need to be identified and overcome, the opportunities need to be elaborated and the challenges need to be addressed for the upgradation of services. The metadata policies, content submission policies and preservation policies have not been fully addressed as more than half of the repositories have explicitly undefined metadata policies, content submission policies and preservation policies. The ETD metadata scheme 'ETD-MS: an Interoperability Metadata Standard for Electronic Theses and Dissertations' has not been adopted by all the repositories. The growing landscape of the ETD calls for explicit content policies to inform users about their rights and reuse policies. Authors complained about the absence of adequate policies and infrastructure to handle the ETDs at the national level as early as 2007[26] and very little progress has been made since then. Other issues like archiving, preserving, cataloguing, harvesting, interoperability, copyright and plagiarism are also noteworthy issues for the ETD repositories which need immediate attention for their redressal.

'The growing landscape of the ETD calls for explicit content policies'

Abbreviations and Acronyms
A list of the abbreviations and acronyms used in this and other *Insight*s articles can be accessed here – click on the URL below and then select the 'full list of industry A&As' link: **http://www.uksg.org/publications#aa**

Competing Interests
The authors have declared no competing interests.

References

1. Peter Suber, "Open access," Cambridge, Mass: MIT Press (2012), **http://dash.harvard.edu/handle/1/10752204** (accessed 6 August 2020).

2. Edward Fox, Gail McMillan and J L Eaton, "The evolving genre of electronic theses and dissertations," *NDLTD Document Archive* (2019) **http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.32.9489&rep=rep1&type=pdf** (accessed 13 August 2020).

3. Susan Copeland and Andrew Penman, "The development and promotion of electronic theses and dissertations (ETDs) within the UK," *New Review of Information Networking* 10, no. 1 (2004), DOI: **https://doi.org/10.1080/13614570412331311978** (accessed 7 August 2020).

4. Zhang Yin, Lee Kyiho and Bum-Jong You, "Usage patterns of an electronic theses and dissertations system," *Online Information Review* 25, no. 6 (2001), DOI: **https://doi.org/10.1108/EUM0000000006536** (accessed 7 August 2020).

5. Arthur Sale, "The impact of mandatory policies on ETD acquisition," *D-Lib Magazine* 12, no. 4 (2006), **http://www.dlib.org/dlib/april06/sale/04sale.html**. DOI: **https://doi.org/10.1045/april2006-sale** (accessed 7 August 2020).

6. Izumi Sugita and Yuko Murakami, "Dissertations and theses in institutional repositories: Case study in Japan," (2007), **http://epc.ub.uu.se/etd2007/files/papers/paper-41.pdf** (accessed 7 August 2020).

7. Maitrayee Ghosh, "E-theses and Indian academia: a case study of nine ETD digital libraries and formulation of policies for national service," *The International Information & Library Review* 41, no. 1 (2009), DOI: **https://doi.org/10.1080/10572317.2009.10762794** (accessed 7 August 2020).

8. Dinesh Gupta and Neerja Gupta, "Analytical study of the ETD repositories and government initiatives for depositing ETDs in India," *Library Management* 35, no. 4/5 (2014), DOI: **https://doi.org/10.1108/LM-09-2013-0092** (accessed 7 August 2020).

9. Transito Ferreras-Fernandez, et al, "Providing open access to PhD theses: visibility and citation benefits," *Program: Electronic Library and Information Systems* 50, no. 4 (2016), DOI: **https://doi.org/10.1108/PROG-04-2016-0039** (accessed 7 August 2020).

10. Eng Ngah Looi and Suit Wai Yeng, "The inevitable future of electronic theses and dissertations within Malaysia context," *Lecture Notes in Computer Science* 2911, (2003): 340–350, DOI: **https://doi.org/10.1007/978-3-540-24594-0_35** (accessed 7 August 2020).

11. Yi Jin, "The development of the China networked digital library of theses and dissertations," *Online Information Review* 28, no. 5 (2004): 367–370, DOI: **https://doi.org/10.1108/14684520410564299** (accessed 7 August 2020).

12. Jamal Al Salmi, "Factors influencing the adoption and development of electronic theses and dissertations (ETD) programs with particular reference to the Arab Gulf States," *Information Development* 24, no. 3 (2008): 226–236, DOI: **https://doi.org/10.1177/0266666908094838** (accessed 7 August 2020).

13. Kristen Yiotis, "Electronic theses and dissertation (ETD) repositories," *OCLC Systems & Services: International digital library perspectives* 24, no. 2 (2008): 101–115, DOI: **https://doi.org/10.1108/10650750810875458** (accessed 7 August 2020).

14. Primoz Juznic, "Grey literature produced and published by universities: A case for ETDs," In *Grey Literature in Library and Information Studies*, eds D Farace and J Schöpfel (Munich: De Gruyter Saur, 2010), 39–51, DOI: **https://doi.org/10.1515/9783598441493** (accessed 7 August 2020).

15. Eun G Park and Marc Richard, "Metadata assessment in e-theses and dissertations of Canadian institutional repositories," *The Electronic Library* 29, no. 3 (2011): 394–407, DOI: **https://doi.org/10.1108/02640471111141124** (accessed 7 August 2020).

16. Joy M Perrin, Heidi M Winkler and Le Yang, "Digital preservation challenges with an ETD collection – A case study at Texas Tech University," *The Journal of Academic Librarianship* 41, no. 1 (2015): 98–104, DOI: **https://doi.org/10.1016/j.acalib.2014.11.002** (accessed 7 August 2020).

17. Tom Steele and Nicole Sump-Crethar, "Metadata for electronic theses and dissertations: A survey of institutional repositories," *Journal of Library Metadata* 16, no. 1 (2016): 53–68, DOI: **https://doi.org/10.1080/19386389.2016.1161462** (accessed 7 August 2020).

18. Miriam Wanjiku Ndungu, "Bibliographic control of theses and dissertations in Kenya," *Library Review* 66, no. 6/7 (2017): 523–534, DOI: **https://doi.org/10.1108/LR-06-2016-0050** (accessed 7 August 2020).

19. Joachim Schopfel and Behrooz Rasuli, "Are electronic theses and dissertations (still) grey literature in the digital age? A FAIR debate", *The Electronic Library* 36, no. 2 (2018): 208–219, DOI: **https://doi.org/10.1108/EL-02-2017-0039** (accessed 7 August 2020).

20. Fayaz Ahmad Loan and Shueb Sheikh, "Analytical study of open access health and medical repositories," *The Electronic Library* 34, no. 3 (2016): 419–434, DOI: **https://doi.org/10.1108/EL-01-2015-0012** (accessed 7 August 2020).

21. Aquil Ahmed, Sulaiman Alreyaee and Azizur Rahman, "Theses and dissertations in institutional repositories: an Asian perspective," *New Library World* 115, no. 9/10 (2014): 438–451, DOI: **https://doi.org/10.1108/NLW-04-2014-0035** (accessed 10 August 2020); Ufaira Yaseen, Fayaz Ahmad Loan and Nelofar Jan, "Open Access E-book Repositories: A Global Scenario," *Library Philosophy and Practice*, 1915 (2018), retrieved from **http://digitalcommons.unl.edu/libphilprac/1915** (accessed 10 August 2020).

22. Fayaz Ahmad Loan, "Open Access Digital Repositories in Asia: Current Status and Future Prospects," *International Journal of Information Science and Management* 12, no. 2 (2014): 35–45, retrieved from **https://ijism.ricest.ac.ir/index.php/ijism/article/view/389** (accessed 10 August 2020).

23. Loan and Sheikh, "Analytical study of open access health and medical repositories".

24. Yaseen, Loan and Jan, "Open Access E-book Repositories: A Global Scenario".

25. Fayaz Ahmad Lone and Mohammad Hanief Bhat,"Use of DSpace in Open Access digital repositories: A Global Perspective," *International Journal of Library and Information Management* 2, no. 1 (2011): 17–22.

26. Eun G Park, Young-Joon Nam and Sanghee Oh, "Integrated framework of electronic theses and dissertations in Korean contexts," *The Journal of Academic Librarianship* 33, no. 3 (2007): 338–346, DOI: **https://doi.org/10.1016/j.acalib.2007.01.010** (accessed on 10 August 2020).

Corresponding author:
Fayaz Ahmad Loan
Documentation Officer
Centre of Central Asian Studies
University of Kashmir, IN
E-mail: fayazlib@yahoo.co.in
ORCID ID: https://orcid.org/0000-0003-4556-8225

Co-author:
Ufaira Yaseen Shah
ORCID ID: https://orcid.org/0000-0003-3677-932X