

University of Nebraska - Lincoln

DigitalCommons@University of Nebraska - Lincoln

---

Library Philosophy and Practice (e-journal)

Libraries at University of Nebraska-Lincoln

---

2021

## Digital Archiving of Grey Literature in Universities: a case study of Jawaharlal Nehru University, New Delhi

Inugurthi Narasaiah

*Jawaharlal Nehru University, New Delhi, narasaiah@gmail.com*

Manorama Tripathi

*Jawaharlal Nehru University, New Delhi, manoramatripathi2@yahoo.com*

Shipra Awasthi

*Jawaharlal Nehru University, New Delhi, shipra2010@gmail.com*

Follow this and additional works at: <https://digitalcommons.unl.edu/libphilprac>



Part of the [Library and Information Science Commons](#)

---

Narasaiah, Inugurthi; Tripathi, Manorama; and Awasthi, Shipra, "Digital Archiving of Grey Literature in Universities: a case study of Jawaharlal Nehru University, New Delhi" (2021). *Library Philosophy and Practice (e-journal)*. 4820.

<https://digitalcommons.unl.edu/libphilprac/4820>

# Digital Archiving of Grey Literature in Universities: a case study of Jawaharlal Nehru University, New Delhi

Inugurthi Narasaiah  
Deputy Director, ACH  
Jawaharlal Nehru University, New Delhi  
[narasaiah@gmail.com](mailto:narasaiah@gmail.com)

Manorama Tripathi  
Librarian  
Jawaharlal Nehru University, New Delhi  
[manoramatripathi2@yahoo.com](mailto:manoramatripathi2@yahoo.com)

and  
Shipra Awasthi  
Assistant Librarian  
Jawaharlal Nehru University, New Delhi  
[shipra2010@gmail.com](mailto:shipra2010@gmail.com)

## Abstract

The present paper describes the importance of Grey literature (GL) in scholarly communication. It highlights the initiatives taken at the Central Library of Jawaharlal Nehru University, New Delhi, to digitize and preserve the GL. It dwells upon how 18,000 theses and dissertations were digitized, and 7500 documents related to people's movements in the pre and post-independence era comprising 3lac pages were digitized and archived. It elaborates on the challenges encountered and overcome, like removing the duplicate copies, assigning metadata, and saving files as PDF/A, budgetary constraints. The study has relevance for other libraries that intend to undertake digitization projects. The paper also describes the Digital Thesis Tracking System which facilitates the research scholars to submit their theses online for evaluation and award of the degree. JNU implemented this online system a few years ago and has reaped rich dividends mostly during this crucial time when the university is closed. It is beneficial for the professionals to implement a system like Digital Thesis Tracking System, facilitating the user community to submit their theses online. The authors recommend implementing such a digital system where the research community can upload their theses, plagiarism report, and other necessary documents. Further, the same can be forwarded to the examiners with ease, and action could be taken. This digital system has the potential to save time and address space and storage concerns of libraries.

*Keywords: digital archiving, digital preservation, grey literature, academic universities, institutional repository, Archives on Contemporary History (ACH), P C Joshi Archive, Dr B R Ambedkar Central Library, Jawaharlal Nehru University (JNU), thesis tracking system*

## 1. Introduction

Grey Literature (GL) is information produced at different government, academics, business, and industry levels in electronic and print formats. It is not controlled by commercial publishing, which means publishing is not the producing body or department's primary activity. There are different types of GL, such as conference proceedings, unpublished literature, Lecture notes, Presentations,

Theses/Dissertations, reports, policy statements, working papers, preprints, technical reports, and other documents as listed on [www.greynet.org/greysourceindex/documenttypes.html](http://www.greynet.org/greysourceindex/documenttypes.html). The other types of GL in the digital milieu are blogs, wikis, presentations, and lectures shared on YouTube, social networking Platforms, posts, or user-generated content on Facebook, Twitter, Instagram, etc. The GL is a vital resource in scholarly communication, research, and development, policymaking in business, industry, and governance. It provides considerable contents of evidence, argument, innovation, and understanding in different subject areas, of science, engineering, health, social sciences, education, arts, and humanities. (Pisa Declaration on Policy development for grey literature resources, 2014)

## **2. Review of Literature**

### **2.1 Importance of GL**

Grey Literature (GL) is produced at different government, business, and higher education and research organizations. The other terms, like grey data and grey information, have also been used with grey literature. The different examples of GL are white papers, technical reports, newsletters working papers as listed on the websites. The grey data refers to user-generated web-based content like tweets, blog posts, Facebook posts, and updates. The grey information alludes to the informally published meeting notes, emails, and so on so forth. It is found across all disciplines and should be accorded equal importance as is given to the white literature. It is meaningful and essential for all the sectors of society (Gul, Shah, Ahmad, Gulzar and Shabir, 2020). In certain subject areas like applied public health, the different health interventions, and their evaluations are primarily available as GL (Adams, et al., 2016). GL is vital for disseminating research findings that may not get published in peer-reviewed journals. The availability of GL minimizes publication bias, where only confirmed, statistically significant results are published. The experiences and firsthand accounts of practitioners, consultancy documents, tacit knowledge do not get covered in the scholarly journals. Thus, GL provides a more balanced view of experiences and evidence. It makes significant contributions to systematic reviews as it provides data and content not available through formal channels of communications.

The Campbell Collaboration, Cochrane Reviews have specified that references of GL must be covered in systematic reviews to ensure their specificity, validity, and reliability of meta-analyses. The researchers adapt and build upon the perspectives and approaches laid down in the GL (Mahood, Van Eerd, and Irvin, 2014; Campbell Policies and Guidelines, 2019). Further, the inclusion of GL in systematic reviews is important when a literary warrant has not developed in the field. The GL available in vernacular languages, unpublished theses, and dissertations should be considered while writing reviews (Hartling et al., 2017). The GL may contain important insights and perspectives which may not get published. Moreover, every research does not get published. There is a big-time gap between research done and its findings reported through formal avenues of information. Further, the rejection rates of scholarly journals are very high. In the wake of these facts, the GL should be preserved for future use and access.

## **2.2 Use of GL in different subjects**

The researchers must ensure search efficiency and maximum coverage of information by consulting both scholarly and GL (Hanneke, and Young, 2017).<sup>7</sup> The need for including GL for conducting multivocal literature reviews in software engineering has been advocated by several scholars (Garousi, Felderer and Mäntylä, 2019). Further, it is considered as a source of knowledge as well as evidence (Rainer, 2019). Kamei (2019) has observed that GL is heavily consulted in the field of software engineering. The systematic literature review does not provide content that is relevant to the practitioners of the field, while GL sources do. The author has elaborated on how to conduct a GL review. Soldani et al. (2018) have searched and analyzed GL available on microservices to identify the technical and operational procedures and problems of microservice-based architectural style. The importance of GL has been well acknowledged in medical sciences and management (Paez, 2017). The information on the potential impact of new and emerging health technologies may assist health care professionals and policymakers before they are deployed in the health care system. The health agencies and companies produce reports, press releases on these technologies. The reports generally include details such as the technology's intended use, available clinical evidence, cost, and current regulatory status. Since new and emerging technologies have not diffused, literature will not be available through traditional bibliographic databases (Farrah and Mierzwinski-Urban, 2019). The conference papers have tremendous importance in research and scholarly communication. Chiware and Backer (2018) have observed that conference papers are cited the most cited sources after the journal articles, books, and online sources. The authors analyzed the citations of theses and dissertations submitted to the Cape Peninsula University of Technology over ten years. Marsolek, et al. (2018) have reported that grey literature is covered in 118 out of 173 databases and 109 out of 115 Institutional repository (IR) studied and analysed to find out the status of GL. These have conference papers, technical reports, theses, and dissertations. The study underlines that the use of standard metadata is important for locating and accessing GL harvested by IRs and databases. The use of standard metadata and controlled vocabulary facilitates easy findability of GL. Since the findability hinges on useful metadata, the Library staff may edit and modify the metadata that authors provide through self-deposit to IRs. Gupta & Sharma (2020) have analyzed 14,547 references of 126 theses submitted to two universities located in Haryana-Kurukshetra university and MDU, Rohtak. Their study also highlighted that out of the 14,547 references, 1344 or 9.23% were for reports followed by government publications, working papers, and websites. Moid, et al. (2020) have highlighted the different types of GL. There are 1,819,319 –theses, synopses, and patents available in the National Digital Library of India. NDLI provided open access to 57.04% of them are available in OA while for 34.62% access is restricted. It is important to mention here that NDLI has other forms of GL as well – audio, video, presentations, applications, and simulations as well.

## **3. Organization and Management of GL**

Bhatt et al. (2020) have studied the organization and management of GL in select social science libraries of Delhi. Their study has highlighted that majority of the libraries, 83.3%, has a separate wing for organizing and displaying the GL. Their bibliographical details are available through OPAC for the users to access and browse. Taylor and Collister (2019) have recommended that altmetrics can be used to find out how the archived content is being used. Altmetrics will help the archivists to engage with the users who are accessing the content. Schopfel and Prost (2020) have analysed scientific papers that mention GL, especially in the field of medical and life sciences. The GL may be retrieved from various sources like Google, Google Scholar, specialized digital libraries, and relevant websites and by interacting with the experts in the field. Schopfel and Rasuli (2019) have debated that Electronic Theses & Dissertations (ETDs) are GL, and at the same time, they are not. ETDs are GL because these are difficult to identify, preserve, and access. They are no longer grey because these can be preserved, archived through IRs, and made accessible to one and all. Dwelling on GL involves all the aspects of LIS-acquisition, organization, availability, findability, preservation, business models, and so on, so forth. The problems associated with GL in actual academic publishing is the lack of appropriate style of citing and lack of archiving for posterity (smart, 2019). But for ETDs, the situation is relatively good as standard infrastructure and institutional control are already in place. There are Information systems in place-IRs, ETD databases that collect ETDs of various universities and facilitate their access. To enhance the impact, dissemination, access, and use of ETDs, the FAIR principles advocated for research data management by Wilkinson (2016) should be adopted.

#### **4. Digital Archiving**

Digital Archives are the repositories where the digital information related to - social, economic, cultural, and intellectual heritage is preserved for long-term accessibility. Digital archiving identifies and moves inactive data out of the current production system and databases into long-term storage systems. The archives preserve digitized or born-digital material to preserve them for the long term because it may have evidential, legal contextual, cultural, and historical or informational value. The archived data is kept and maintained for future reference and use. The digital archiving facilitates storage, preservation, easy access, and use by posterity, promotes paperless organizations, and maintains confidentiality of data. The Memory institutions like museums, library and information centers archive resources of learning and cultural heritage to make them accessible for use in the future. UNESCO has urged its member states to document and archive the implications and consequences of COVID-19 (UNESCO).

The National Archives of Australia, New Zealand, UK, Spain, Switzerland, Germany, USA, India, and Canada maintain records of government departments, military units, customs, patents, history, and culture. The community-based archives document, record, and preserve their histories, heritage, and experiences for future use and reference (Rodrigues, 2016).

Personal Digital Archiving (PDA) is a new concept that has emerged lately. It means when individuals want to capture and archive their digital records; or when libraries or museums want to preserve individuals' digital records. Nevertheless, the individuals and institutions may adopt different

practices in managing and preserving digital archives, based on the knowledge and skillsets possessed by them. PDA involves practices of traditional archiving and personal information management (Condron, 2019).

The born-digital material like digital photographs, digital documents-PDFs, harvested web content, digital manuscript, electronic records of organizations, user-generated content of twitter, Facebook, music, and movies need to be archived and preserved. All stakeholders should collaborate in archiving and preserving born-digital content for future use.

In 2016 the British Library undertook a project to archive born personal digital collections of Carmen Callil, Hanif Kureishi, and Ronald Harwood. The metadata was harvested from digital objects, uploaded to archival software, catalogued, and published through the British Library catalogue. The Library used an FTP server to host PDF/A surrogates of the migrated digital objects and made the content accessible to the users. The users were asked to share their experiences of navigating the online archived born-digital content. The users must comply with copyright and data protection practices while using archives and manuscripts (Jaillant, 2019). The John Ryland Library, Manchester University undertook Carcanet Project (2012-1014) to preserve 215,000 emails and 65,000 attachments generated by Carcanet Press, which is one of the leading poetry publishing houses in the UK. The emails are the communication of the press with poets, critics, editors, translators, and artists. The content has tremendous importance as primary research material (Baker, 2015).

#### **4.1 Institutional Repositories (IRs) for archiving GL**

IRs have great potential in making GL discoverable. The universities worldwide are setting up IRs to preserve GL generated as theses and dissertations. IRs can be developed with the use of open-source software such as Dspace and Eprints. The content available on open platforms can be read, download, copy, disseminate under creative commons license. This software is available free of cost and can be customized according to the requirement of the Institutions. The Registry of Open Access Repositories provides information about the growth and status of repositories throughout the world. It lists more than 4000 OA repositories located worldwide. The Registry of Open Access Repository Mandates and Policies (ROARMAP) highlights the growth of Open Access mandates and policies adopted by the higher educational institutions of the world. The OAIster database is a union catalogue of 50 million records of OA resources. The database was developed by using the Open Archives Protocol for Metadata Harvesting (OAI-PMH).

Shivaram and Biradar (2019) emphasized the kinds of GL to be preserved and policies to be framed for collection development, infrastructure for archiving, preserving, and providing access. They have also observed that the technical reports are used the most by the CSIR researchers, followed by doctoral theses. GL is predominantly archived in European and North American OA repositories. In India, OA repositories archive primarily reviewed literature followed by theses and dissertations.<sup>28</sup> The metadata standards for GL should be developed, which would enhance searching in databases and support interoperability. An increased level of collection development of GL in IRs would facilitate preservation and increase the findability, wider dissemination, and use (Marsolek, et al., 2018). The

successful completion of digital preservation activities depends on the technology, use of appropriate metadata, human resources, organizational strategy, and continuous collaboration with preservation experts (Spence, Mccutcheon and Mahon, 2019). Libraries do not have skilled human resources and financial support to manage and run digital archives in IRs satisfactorily (Anyaku, et al., 2019).

GL is a valuable source of information about the activities of universities and organizations. But there are certain problems which prevent the full utilization of GL- the varied formats and huge size. There is a need to identify sources, collect them to avoid their duplication, and assign unique and persistent identifiers, add appropriate metadata by adhering to the standards. The Pisa Declaration has advocated that for deriving maximum dividends out of GL, all stakeholders should collaborate and work in unison. The governments and organizations should promote open access to GL. The organizations which are engaged in the production of GL should unite and use interoperable standards for description (including open metadata standards), collection and management. The copyright laws and legal deposit should be amended to facilitate libraries in collecting GL and extending their access to the user community. The governments and organizations should provide budgetary provisions for collecting, preserving, and making it accessible for use (Pisa Declaration on Policy development for grey literature resources, 2014).

## **5. Objectives of the study**

The objectives of the paper are as under:

- To provide a general overview on grey literature
- To highlight the digitization & preservation of grey literature at Central Library, JNU.
- To spotlight the Digital Thesis Tracking system of the university which facilitates submission and evaluation of theses electronically.

## **6. Significance of the Study**

The literature published on digital archiving, grey literature, digital preservation, etc. was surveyed for the present study. The case of Dr B R Ambedkar Central Library, JNU, New Delhi, has been presented in the paper. The paper enumerates the two initiatives which the JNU Library undertook, i.e., implementation of the digital tracking system of theses and dissertations and the digitization and preservation of PC Joshi Archives on Contemporary History. The purpose of conducting the study is to sensitize the professionals to digitize the grey literature and rare documents and make them available and accessible to the wider community. The aim of sharing the existence of such system is to sensitize the professionals to have a model which would support the research community in the difficult times when the institutions are closed. These initiatives help library to provide access based services by reaching out to the user community as per their convenience.

The study drives home that these initiatives may be replicated in other libraries to support academic and research activities across the country. These services assume more importance especially in the

challenging times of COVID-19 crisis when library facilities are under suspension. The paper has relevance for other other library professionals as they may initiate these activities at their organizations to enhance services.

## **7. Jawaharlal Nehru University: A short profile**

Jawaharlal Nehru University is ranked number one among the central universities in NIRF rankings 2020 and is next to IISc, Bangalore. The university has consistently maintained the position during the last four years since 2017-2020 in NIRF rankings (NIRF Ranking, 2020). Dr B R Ambedkar Central Library, JNU is a nine floor building and a mezzanine floor. The different collection is housed at each floor. During the crisis period due to COVID 19, library has extended different services to user community by reaching out to them.

Previously, the process of theses submission was online but the students visited central library to complete some formalities. But now due to the challenging conditions, the submission process is completely online.

### **7.1 Digital Archiving and Preservation of Theses and Dissertations**

The Central Library, JNU has a collection of 18,000 + theses and dissertations which were digitized by an external agency following an award of contract through 2 bid step procedure. INFLIBNET Centre extended financial assistance for executing the digitization work. The work was completed within 18 months. The bibliographical details of the theses are available through OPAC, and the full text is available on Intranet; efforts are on to make them available on the Internet. Central Library has already uploaded 4722 theses on the Shodhganga portal.

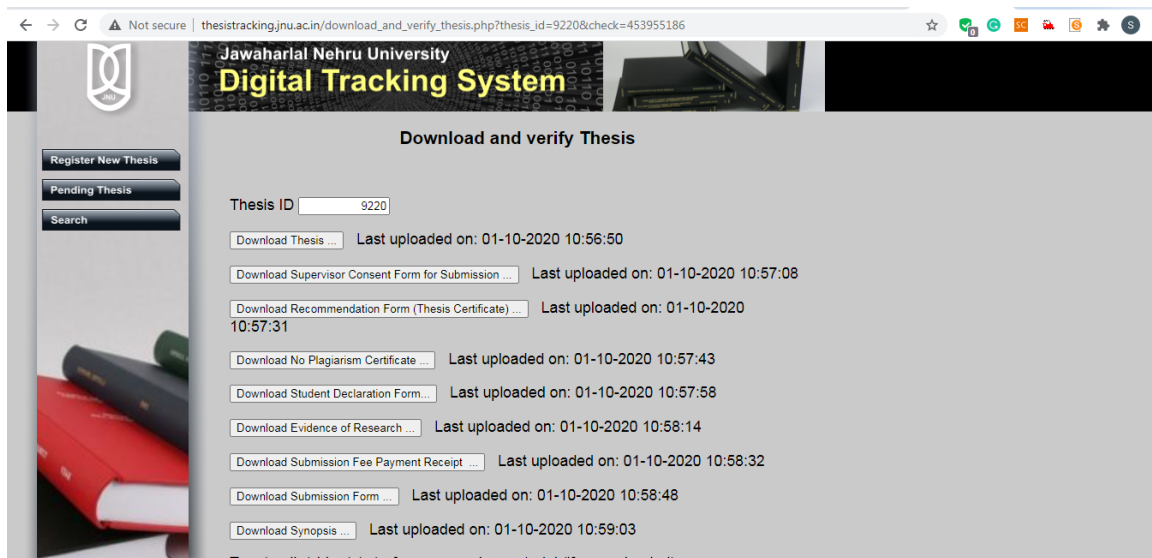
### **Digital Thesis Tracking System**

The Thesis Tracking System is an online platform which is available on Intranet that connects different levels of thesis submission and evaluation. In the manual system, the student had to print 4 copies of the thesis and get them bound, followed by their submission at the school and evaluation branch.

The Fig 1 from the “Thesis Tracking System” depict the clear picture of the system for a better understanding of the workflow:-

Figure-1. Digital Thesis Tracking System, JNU





The Thesis Tracking System is the outcome of the hard efforts of the Central Library and Communication Information Service (CIS), JNU, which ended many challenges faced in the manual system by the Central Library and research scholars. The new system enables the research scholars to register in the Thesis Tracking System from their respective schools and centers of study, and a soft copy of the theses and other mandatory documents are uploaded to the system at the school level. The students are required to upload the soft copies of the theses, certificate/declaration, originality report generated through the anti-plagiarism software, the declaration form which contains the metadata of the file and consent for archiving the e-theses at Shodhganga platform. The embargo period of three years is maintained for uploading the file at the platform. The Central Library scrutinizes the uploaded theses and other related files and verifies the completeness and accuracy of the documents. If the record is found correct, it is forwarded to another level of the system, which is the Evaluation Branch of the university, where the next course of action is performed for evaluation. Else, if the uploaded files and supporting documents are not found in order, the record gets rejected at the university library level and is sent back to the concerned school/centre for corrective measures and re-submission.

The authorized access to "Thesis Tracking System" is allowed to offices at the school/centre level, central Library, and the evaluation branch of the university. The limited access to the authorized officials ensures data accuracy and data security, and thus the desired results of better management and quality services are achieved.

### ***New features of the system***

Recently the university has added certain features to enhance the system. Within 7 days of submission, a system-generated SMS will go to the designated examiner. The examiner has to give his consent within 7 days, failing which the next examiner will be sent SMS for consent. The examiner has to give his feedback report within 6 weeks of giving his consent to the university, for agreeing to evaluate the thesis. The university will hold viva- voce digitally within a month after getting feedback

from the examiner. It is relevant to mention here that around 150 Ph D /M. Phil viva were held virtually during the lockdown period due to the COVID-19 pandemic.

## **7.2 Challenges faced and overcome**

The Digital Thesis Tracking System initiative has been taken up to assist the user community and to enhance the digital collection. During this process several challenges have been faced by the library:

- Students upload sham documents- Article submission is not a mandatory requirement for M.Tech students. So unless the students upload some documents at that tab, it does not allow them to move to the next level. In order to circumvent the system, the students upload sham documents.
- The students upload the complete originality report (OR) which is not required. To overcome this issue the system has put a cap on file size to upload the first two pages of the OR.
- The smooth functioning of the system is possible when the users of the system are trained to handle and understands the requirement of the system. The users need to ensure whether the relevant document has been uploaded, it has been observed by the library that the students are not serious while uploading the content.
- The uploaded documents are not in a proper format or they are not endorsed by the Supervisor or Chairperson.
- The partnership with different sections is must as the teamwork is involved in the successful implementation and use of the system.

## **8. Digitization and Preservation of PC Joshi Archives on Contemporary History, JNU**

The Archives on Contemporary History (ACH) in Jawaharlal Nehru University is involved in the acquisition, preservation, and dissemination of information related to Contemporary History of India, especially the history of the Indian nationalist movement and people's movements in the post-independence India. The Archives on Contemporary History is also known as PC Joshi Archives, named after its collector, P.C. Joshi, whose personal collections have been organized digitized and maintained in the Archives. The Archives was part of School of Social Sciences till 2017 and brought under the administration of University Library from April 2017.

### **8.1 Objectives of Digital Archiving**

Digitization of Archival material is one of the crucial tasks taken up in this project. Most of the documents preserved in PC Joshi Archives are very brittle, and they needed immediate steps for preservation and digitization. The digitization of these records has helped the Archives to preserve the original documents for a longer period. After digitization, the scholars who visit the Archives for reference could refer to digital documents instead of original documents so that the longevity of the original materials can be ensured. The digitization of records is also necessary to provide online access to scholars.

The following steps were followed for digitization:

- a. Selection of documents for digitization
- b. Digitization of documents
- c. Metadata creation
- d. Implementation of a digital archive with technologies and features that are cost-effective, scalable and engaging;
- e. Ingest of digitized material and their corresponding metadata records into the archival platform;
- f. Dissemination via permissions-based secure on-site access for researchers.

## **8.2 Financial grant**

International Institute of Social History, Amsterdam in 2013 and National Archives of India, Ministry of Culture, Government of India in 2014 provided a financial grant of Rs. 3, 00,000 and Rs.7,54,000 respectively for digitization and preservation of archival material.

## **8.3 Selection of documents**

The digitization project started with the selection of materials for digitization. The documents for digitization were selected based on the contents physical condition, significance to the overall collections, user demand and interest.

The task of digitisation and metadata creation was assigned to a private company through a two-step bidding procedure. The university committee decided to pay Rs.2.70 per page for digitization and Rs.8.30 per document for metadata creation. Since the funding agencies did not give any guidelines for digitization, Archives on Contemporary History decided to carry out digitization with the following specifications.

- Image type: TIFF
- DPI: 400 DPI
- Colour
- Bit Depth: 24

The outsourcing agency digitized 7500 documents comprising 3,00,000 pages in the format specified for digitization. Since the image size was huge, it was a challenging task to access the original TIFF images. Hence, the original images were converted to JPEG format for easy access.

The digitization process included metadata preparation, scanning, image correction, with constant quality control.

Figure-2. Book eye scanner deployed for digitisation



The archival material was fragile, delicate, or brittle, and an initial assessment of the material was undertaken to assess its status and condition, figure out and record the appropriate methods of handling before initiating any procedure. The vendor was asked not to break the binding of the document and use only “face-up” (a.k.a. “orbital”) scanners for scanning.

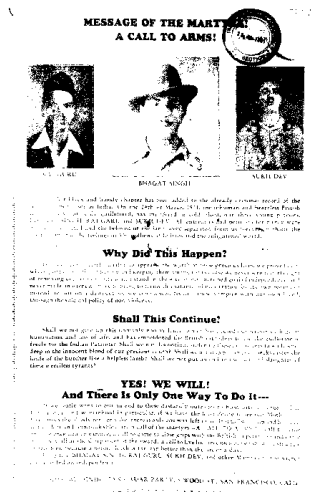
### 8.4 The output of the Digitization and Preservation

After completing the digitization, free online access to all the digitized documents has been provided to the research scholars and interested people through a digital archive interface on the Central Library website. The digital archive of ACH can be accessed at

[http://10.107.16.120/cgi-bin/koha/opac-search.pl?branch\\_group\\_limit=branch%3AARCH](http://10.107.16.120/cgi-bin/koha/opac-search.pl?branch_group_limit=branch%3AARCH)

The documents digitized under this project are useful for researchers, academics, journalists, political activists, and general public. The document covers various themes like the history of contemporary India, especially democratic movements in India and other parts of the world.

Figure-3. Quality of the digitized material



The Figure 3 displays the clarity and accuracy with which the material has been digitized.

### 8.5 Importance of the Archives

The effort of the Archives is going to help the researchers working on the democratic movements in South Asia to have a fresh look at the history of freedom movement and democratic movements in post-independence India. It will throw a light into the lesser known aspects of the above subjects. It will also help the scholars to have a comprehensive understanding of the events related to India's freedom struggle and post-independent India. The documents are also useful to the researchers working on subjects like History, Political Science, Economics, Sociology, International relations and politics, Indian culture.

Figure-4. Central Library Website, JNU

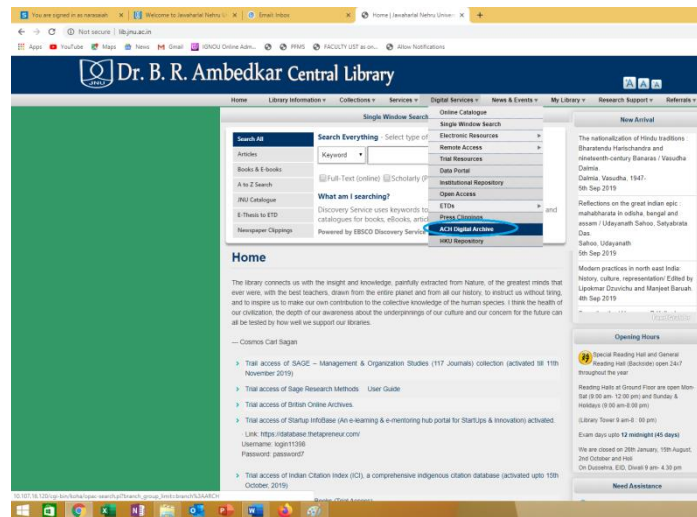


Figure 4 displays that the digitized material is accessible through the library website www.lib.jnu.ac.in.

The digitization of the documents would also help visually challenged students and researchers to access and read the documents with little effort.

### 8.6 Challenges faced in the digitization

ACH had taken up digitization for the first time, and it had to face several challenges in the process of digitization and creation of the digital archive.

Duplication of documents in the archival collection was one of the major problems faced during digitization. The collection in the Archives consists of unpublished documents of the political parties. Some of the documents have only a slight difference between one another. Some of them are drafts presented for approval of the higher-level committees, and some of them are the documents finally approved by the higher-level committees of the political parties. Some of them are approved without any changes, and approved documents with changes have also been included in the collection. Some of them were later published to make them available for the general public. But the titles and creators of both the documents were the same, and the staff engaged in the selection of documents faced a dilemma which one to select for digitization. Only an expert on the subject could decide which document should be selected for digitization. The problem of duplication was overcome with the help

of the staff who were specialists on the subject matter of the documents preserved in the ACH. Proper care was taken to avoid digitization of duplicate copies of the same documents.

Metadata creation was another problem faced by the staff deputed by the private vendor. Since many of the documents digitized were unpublished and manuscript documents, the staff were not able to identify details of the documents. One has to have a thorough knowledge of the subject matter of the documents to prepare metadata for the documents. The permanent staff employed in the archives assigned the metadata for the digitized documents.

## **8.7 Technical issues**

The format decided for the creation of the digital image of the documents was TIFF format colour image with 400 DPI. The size of each image ranged from 10 MB to 100 MB depending on the background colour of the image. It has become very difficult for the staff to view the images as the computers installed in the ACH were old and it took more time to process (open) the images. Later, archives acquired higher capacity computers and converted the Original tiff images to small size JPEG files and PDF files. Now the digital data is stored in TIFF format in the External Hard drives; JPG files and PDF-A files in the external hard disks and Archival quality DVDs.

**8.8 Financial constraints** The common challenge is the budget constraint. Due to the paucity of funds, Archives was not able to digitize all its collections.

## **9. Conclusion**

The visibility of grey literature can be enhanced by archiving the material in institutional repositories, depositing at open-access platforms, or at any of the platforms which are accessible to the researchers. The grey literature is the intellectual output that should be disseminated and shared as well among the researchers for further use, and it certainly increases the prestige of an organization. The paper recommends that all universities digitize the grey literature and deposit it on a public platform for the optimum utilization among the research community. The Government and Non-Government agencies should facilitate the universities through finance and skilled manpower for digitization. Many times Universities could not undertake such activities due to paucity of funds and lack of clarity in copyright laws; there is some hesitation to join the movement of digitization. The authors of this paper strongly recommend designing a blueprint before taking up the activity of digitization. Few standards need to be formulated and followed, such as budget, selection of material, and hardware compatibility and obsolescence parameters should be considered. The university libraries could be surveyed again in a few years to explore the change in practices regarding the digital archiving of grey literature and the development of dedicated theses repositories. In future, studies could be undertaken to evaluate the use of the digital archives and IRs. Previous studies also inferred that digitization in India is a challenging issue because of budgetary constraints and lack of skills sets. There is no legal framework and concrete policies laid down by the government for digital preservation. Though, the Indian

government has taken some initiatives in this regard, but they need to be reformed to fit in the present scenario (Ansari,2010).

## References

1. Pisa Declaration on Policy development for grey literature resources. (2014). *Grey literature network service*. Retrieved from [http://www.greynet.org/images/Pisa\\_Declaration,\\_May\\_2014.pdf](http://www.greynet.org/images/Pisa_Declaration,_May_2014.pdf)
2. Gul, S., Shah, T. A., Ahmad, S., Gulzar, F., & Shabir, T. (2020). Is grey literature really grey or a hidden glory to showcase the sleeping beauty. *Collection and Curation*. Retrieved from <https://doi.org/10.1108/CC-10-2019-0036>
3. Adams, J., Hillier-Brown, F. C., Moore, H. J., Lake, A. A., Araujo-Soares, V., White, M. & Summerbell, C. (2016). Searching and synthesising ‘grey literature’ and ‘grey information’ in public health: critical reflections on three case studies. *Systematic reviews*, 5(1),164.
4. Mahood, Q., Van Eerd, D. & Irvin, E. (2014). Searching for grey literature for systematic reviews: challenges and benefits. *Research synthesis methods*, 5(3), 221-234.
5. Campbell Policies and Guidelines. (2019). Retrieved from <https://onlinelibrary.wiley.com/page/journal/18911803/homepage/author-guidelines>
6. Hartling, L., Featherstone, R., Nuspl, M., Shave, K., Dryden, D. M., & Vandermeer, B. (2017). Grey literature in systematic reviews: a cross-sectional study of the contribution of non-English reports, unpublished studies and dissertations to the results of meta-analyses in child-relevant reviews. *BMC Medical Research Methodology*, 17(1),64.
7. Hanneke, R. & Young, S. K. (2017). Information sources for obesity prevention policy research: a review of systematic reviews. *Systematic reviews*, 6(1), 1-10.
8. Garousi, V., Felderer, M., & Mäntylä, M. V. (2019). Guidelines for including grey literature and conducting multivocal literature reviews in software engineering. *Information and Software Technology*, 106, 101-121.
9. Rainer, A. (2019). Using argumentation theory to analyse software practitioners’ defeasible evidence, inference and belief. *Information and Software Technology*, 87, 62-80.
10. Kamei, F. K. (2019). The use of grey literature review as evidence for practitioners. *ACM SIGSOFT Software Engineering Notes*, 44(3).
11. Soldani, J., Tamburri, D. A. & Van Den Heuvel, W. J. (2018). The pains and gains of microservices: A Systematic grey literature review. *Journal of Systems and Software*, 146, 215-232. DOI: 10.1016/j.jss.2018.09.082.
12. Paez, A. (2017). Gray literature: An important resource in systematic reviews. *Journal of Evidence-Based Medicine*, 10(3), 233-240.
13. Farrah, K. & Mierzwinski-Urban, M. (2019). Almost half of references in reports on new and emerging nondrug health technologies are grey literature. *Journal of the Medical Library Association: JMLA*, 107(1), 43.
14. Chiware, E. R. & Becker, D. (2018). Citation patterns of conference proceedings in master’s and doctoral studies: A case study of information technology and systems. *SAGE Open*,8(2).
15. Marsolek, W., Cooper, K., Farrell, S., & Kelly, J. (2018). The types, frequencies, and findability of disciplinary grey literature within prominent subject databases and academic institutional repositories. *J Librariansh Schol Comm*, 6(1).
16. Gupta, D. K. & Sharma, V. (2020). Citation analysis of grey literature reflected in social science ETDs submitted in two state universities of Haryana, India. *Collection and Curation*.
17. Moid, A., Raza, M. M., & Jahan, K. A. (2020). Study of grey literature in National Digital Library of India: analyses and trends. *Collection and Curation*. Available at <https://doi.org/10.1108/CC-12-2019-004>
18. Bhatt, R. K., Singh, M., & Rana, J. (2020). Organisation and management of grey literature in select social science libraries, *Collection and Curation*. Available at <https://doi.org/10.1108/CC-10-2019-0033>
19. Taylor, A. L. & Collister, L. B. (2019). Informing the Digital Archive with Altmetrics. *Scholarly and Research Communication*, 10(1),12.
20. Schöpfel, J. & Prost, H. (2020). How scientific papers mention grey literature: a scientometric study based on Scopus data. *Collection and Curation*.

21. Schopf, J. & Rasuli, B. (2018). Are electronic theses and dissertations (still) grey literature in the digital age? A FAIR debate, *The Electronic Library*, 36(2), 208-219. Available at <https://doi.org/10.1108/EL-02-2017-0039> (Accessed on 4 August 20)
22. Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., ... & Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship, *Scientific data*, 3(1),1-9.
23. UNESCO. Available at <https://en.unesco.org/news/unesco-launches-new-blog-living-memory-making-most-documentary-heritage-covid-19-decision> (Accessed on 28 July 2020)
24. Rodrigues, A. (2016). Introducing an archival collecting model for the records created by South African Portuguese community organizations. *Archives and Manuscripts*, 44(3), 141-154.
25. Condrón, M. (2019). Identifying individual and institutional motivations in personal digital archiving. *Preservation, Digital Technology & Culture*, 48(1), 28-37.
26. Jaillant, L. (2019). After the digital revolution: working with emails and born-digital records in literary and publishers' archives. *Archives and Manuscripts*, 47(3), 285-304. DOI:10.1080/01576895.2019.1640555
27. Baker, F. (2015). E-mails to an editor: safeguarding the literary correspondence of the twenty-first century at the university of Manchester library. *New Review of Academic Librarianship*, 21(2), 216-224.
28. Shivaram, B. S. & Biradar, B. S. (2019). Grey literature archiving pattern in open access (OA) repositories with special emphasis on Indian OA repositories, *The Electronic Library*, 37(1), 95-107. Available at <https://doi.org/10.1108/EL-05-2018-0100> (Accessed on 28 July 2020)
29. Spence, A., Mccutcheon, V. & Mahon, M. (2019). Case Study: the University of Glasgow's digital preservation journey 2017-2019. *Insights*, 32(10),1-9. DOI: <https://doi.org/10.1629/uksg.461>
30. Anyaoku, E. N., Anthonia, U., Echedom, N. & Baro, E. E. (2019). Digital preservation practices in university libraries: An investigation of institutional repositories in Africa, *Digital Library Perspectives*, 35(1), 41-64. Retrieved from <https://doi.org/10.1108/DLP-10-2017-0041>
31. NIRF Ranking: 2020. Retrieved from <https://www.nirfindia.org/2020/UniversityRanking.html>
32. Ansari, S.M. (2010). Long term digital preservation: Indian Scenario. *Changing Role of Library professionals & libraries in the digital age*.117-125. Retrieved from [https://www.academia.edu/9860706/Long-term\\_Digital\\_Preservation\\_Indian\\_Scenario](https://www.academia.edu/9860706/Long-term_Digital_Preservation_Indian_Scenario)
33. Jawaharlal Nehru University. Retrieved from <https://www.jnu.ac.in>