Techniques of Digital Preservation and Curation Process for the Printed Data with the Dspace (Digital Library Software)

Manoj Das[1], Dr. Sangeeta Singh[2], Madhumita Sahoo[3]
[1](Librarian, Cooch Behar Govt. Engineering College, West Bengal), Ph.D. Research Scholar, Dept. of Library and Information Science, Dr. C. V. Raman University, Kargi Road Kota, Bilaspur (C. G.)
[2]Professor, Dept. of Social Science (Library & Information Science), Dr. C. V. Raman University, Kargi Road Kota, Bilaspur (C. G.)
[3]M. Phil Research Scholar, Dept. of Library & Information Science, Vidyasagar University, Midnapore, West Bengal

***Abstract*-There is a lot of research going on in the present age and new research information is also growing, but the present age is so advanced that all those research documents are being converted and preserved easily by converting them to digital but how about all the old, printed documents stored in the library? Demonstrate how to convert digitally and how to store them over a long period of time. This article shows how to convert and preserve printed documents to digital and how to store them for a long time. If steps are not taken to preserve these printed documents, many valuable resources will be wasted, so all institutional libraries will have to take steps now to preserve old, printed documents. This article discusses how to preserve old, printed documents. Finding old documents and converting them from print documents to digital documents. Some tools and software have been used for preservation such as scanner machines, editing software, file converting software, computers, digital library software (Dspace), etc. Old documents need to be digitally preserved, so that researchers of the present age can use those documents in their research. They need to be cursed as soon as they are preserved so that the documents get the right dimensions. Here, digital library software called Dspace is used for curation and preservation, so that documents can be cursed and preserved in a good way. Dspace is open-source digital library software that helps store and preserve documents. Metadata can be edited and cursed through this software while the documented ones are preserved. Preservation means that documents can be stored for a long time and delivered to users at the right time.**

***Index Terms:* Digital Preservation, Curation, Digital Curation, Metadata Editing, Curation Process, Preservation Techniques.**

1. INTRODUCTION

Digital preservation is how to store documents in a particular place that can be used for a long time. If every institution converts all their printed documents into digital documents, preserves them in a specific place and gives them digital access, then many students of the present age will have the opportunity to use those documents.

Digital documents are increasingly increasing in size and additional information is produced by the research project. Growing in complexity and diversity, especially in the fine arts, humanities and cultural issues, documents are slowly being created day by day. Gradually, as the current situation progresses, it is becoming more and more difficult for students to go to a library to study, so all institutions should digitally convert and print all the old, printed documents they have so that all types of students can use them in their studies.

Digital curation is mainly based on the selection of selected documents. Digital Preservation is the process of selecting, storing and creating access insurance for digital information. We argue that metadata and enrichment are critical to long-term access to relevant information in digital resources. What is meant by digital correction here is how to preserve a digital document and give it the right dimension. Check where the preservation documents are going wrong or what they are missing or what kind of metadata needs to be given in case of an accurate document. Many times, it is seen that the documents have been

preserved but they have not been corrected or reviewed. It is seen that there are many misconceptions in the case of documents which are very difficult for the users to understand, so it is very important to do digital correction. A curator determines the accuracy of a preservation document by examining how relevant the metadata of the preservation documents is to the document and re-prescribing if there are any errors.

Digital carving helps maintain research data and preserves and adds value throughout its life cycle. Active management of this data reduces their long-term threats and reduces the risks of digital obsolescence. In addition to the reduction of duplication, efforts to create research data, digital correction strengthens long-term values and makes existing data available for reuse in high-quality new research.

Dspace is a digital library software that allows you to preserve documents very well. Here all these printed documents will be converted to digital, they will be pressed by entering metadata through dspace and metadata will be cursed with the help of software on dspace. Document Preservation is an open source and free software on Dspace that installs it for free.

2. OBJECTIVE OF THE STUDY

The aim of the study is to explore the digital preservation and digital curation techniques and process. The main objectives of the present study are as follows:

- To find out the printed documents in institutional library.
- To show that what is the identification policy of printed documents.
- To show that how to scan the printed documents for transform to digital documents.
- To Find out that what is the techniques of digital preservation.
- To show the process of digital curation through dspace software.
- To show how to maintain the preserved documents; etc.

3. METHODOLOGY OF THE STUDY

This study is techniques of digital preservation and curation process through dspace software. There explore the methods of the process and techniques for digital preservation and digital curation, those are follows:

- Identify the printed documents with the proper way,
- Select the printed documents as per institution guidelines,
- Scan the selected documents one by one as per instruction of institution,
- Edit the scanned documents by the editing tools and save to another new folder according to file name,
- Convert the edited image file to pdf by the pdf converter tool as per institution policy,
- Store the pdf file to dspace software with proper metadata entry as step by step,
- Review and curate the stored documents through the dspace software and finally archived.

4. LITERATURE REVIEW

Literature review is a text of a scholarly paper, which includes the current knowledge including substantive findings, as well as theoretical and methodological contributions to a particular topic.

Doorn & Tjalsma (2007) this paper review of the general introduction into the special issue of Archival Science on ‘‘archiving research data’’. It summarizes the different contributions and gives an overview of the main issues in this special field of archiving. [4] Hedges & Hasan (2007) in this paper view outline an approach to implementing digital curation strategies in data grids based on the iRODS middleware, in particular by exploiting iRODS’ Rule Engine. [10] Shah & Marchionini (2007) this paper review of the digital curation primarily involves selecting, preserving, and insuring access to a repository of digital information. [26] Yakel, (2007) stated the review paper of digital curation is the active involvement of information professionals in the management, including the preservation, of digital data for future use. [28] Chris, (2008) this paper review of the data deluge is beginning to have an effect on libraries and archives. [13] Cunningham, (2008) this paper reviews the digital archiving experiences of the National Archives of Australia (NAA), in light of the key messages outlined in the first half of the paper placing those endeavours in the broader Australasian

context. [3] Hedges, Blanke & Adil (2008) this paper review of an increasingly large quantity of digital data is being produced by research projects, and in addition the data is growing in complexity and diversity. [10] Marketakis, Tzanakis & Tzitzikas (2009) this paper review of the preservation of digital objects is a topic of prominent importance for archives and digital libraries. [18] Patel, (2009) this paper review of RI for crystallography data and its role in the curation, maintenance and management of such data. [21] Gunia & Sandusky (2010) this paper review of the PREMIS (Preservation Metadata: Implementation Strategies) was utilized to specify the requirements for preservation metadata for DataONE, one aspect of DataONE's technology architecture. [9] Lancia, (2010) this paper review of the research has produced a metadata set fitted for the implementation of curation and preservation policies in small research institutions. [15] Gracy & Kahn (2011) this paper surveys research and professional literature on preservation-related topics published in 2009 and 2010, identifies key contributions to the field in periodicals, monographs, and research reports, and provides a guide to the changing landscape of preservation in the digital age. [8] Beaudoin, (2012) this paper investigates the various discussions in the literature surrounding contextual information, and then presents a framework which makes explicit the various dimensions of context which have been identified as useful for digital preservation efforts and offers a way to ensure the capture those aspects of an object's context that are often missed. [1] Gaura & Tripathi (2012) this paper review of the concerted efforts are being made worldwide for preserving e-journals' content for posterity. [7] Iorio, et. al. (2014) this paper review of the Information Packages' (SIP) preservation metadata was encoded in the semantics of the PREMIS standard, which is the implementation metadata set, mapped from the OAIS conceptual model. [12] Molloy, (2014) this paper review of the digital curation awareness and practice of a sample of practitioners from the UK performing arts community. Here, this enquiry is set into the broader context of digital curation and preservation. [20] Perrin, Winkler & Yan (2014) this article examines the problems that can arise years after the transition from a physical to electronic collection and presents documentation solutions that can make ETD preservation and curation more effective. [23] Ray, (2016) this paper review of the development of digital curation education and practice in museums in the USA through: an overview of the development of digital curation. [24] Erkimbaev, (2019) this paper review of role of the curation process in supporting scientific data repositories is studied. [6] Mello & Godoy (2020) this article presents a proposal model for curation of digital objects from an oncology research centre in Santa Catarina State. [19] Sharma, (2020) this paper review of the most of these libraries maintain a unique collection of manuscripts and rare documents of historical importance that need to be preserved for future generations. [27] Patlakas, Musso & Larkham (2021) this paper review of the Digital Curation Model (DCM) for POE data is presented. [22]

5. PRESERVATION TECHNIQUES

Data preservation is an important thing to do in order to follow certain rules. When data is preserved, some steps need to be taken, such as first choosing the topic on which to preserve and then looking for data like that. In order to search for data, you have to select them, and it is very important to use the registration book, otherwise it will not be possible to understand which documents are being taken and which documents are not being taken. It is very important to maintain them. Then scan the selected documents and after scanning re-register in the registrar. After the scan process is complete, they have to be edited, converted to files, entered into metadata and uploaded to the dspace, checked and reviewed, made storage arrangements, etc. So, all the techniques and steps of preservation are discussed in detail below:

5.1. FINDINGS

The first thing to do when preserving documents is to look for the documents or add which documents will be preserved. An institute library contains a variety of documents, some of which are copyrighted and some non-copyrighted. First you need to find out the non-copyrighted documents, then you need to find out which ones are more important than which documents will be useful for the future students.

Fig-1: Documents Finding in an Institute Library

5.2. SELECTION

Documents selection is a meeting with a library committee to decide which study materials to scan. What should be the type of documents that will be selected? An institute library contains various types of documents such as books, journals, theses, magazines, newspapers, etc. So you have to select the documents that are scanned and preserved so that no one has any problem. That is why the selection process has to be done in a good enough way, which if scanned and preserved can be used by the students in the future for their study and research. Documents should be selected keeping all these things in mind.

5.3. ARRANGEMENT

Once the documents are selected, those documents have to be registered in a registration book. Some rules have to be followed to register documents in the registration book, such as serial number, document title, call number, access number, volume, edition, author, page count, etc. If it is done, there will be no record of which copy is being scanned and which is not. So it is very important to maintain a physical register for all this work.

5.4. SCANNING

If you want to scan documents, first you need a scanner machine and some scanning tools with it; you have to procure them and then scan. The scanning machine has to scan the Bookeye 4 and scanning software BCS2. First you have to keep the documents properly in the scan machine and then you have to set up the page well with the scan software. Once the page is set up, you have to scan the documents with the file name. Scan one page at a time so that it is not bad. Once scanned, it will be saved as a file in a folder and each page of these documents will be saved as an image copy.

Fig-2: Scanner Machine with Scanning Document

5.5. EDITING

After scanning the documents, they have to be edited with editing software, such as having editing software BCS2 with which we can edit the whole document on each page. When documents are scanned, each page has some stains, page disorders, etc., so you have to use editing software to clean those pages. With this software, scanned documents can be edited very well. Once edited, the edited documents have to be saved as a separate folder file, so that the documents can be easily marked for later use.

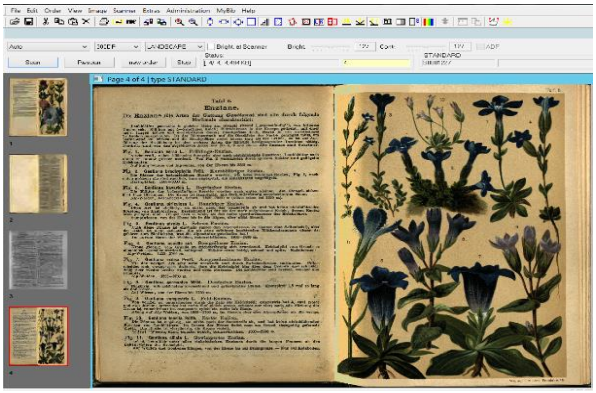


Fig-3: Editing the Document by BCS2 Software

5.6. CONVERT AND RENAME

Scanned documents are saved as image copies so those image files have to be converted to PDF files and documents have to be renamed as titles. Some converting tools are needed to convert scanned documents to PDF files, such as Abby Fine Reader 11, this software converts from image copy to PDF. When converting these documents, the file size of the documents should also be measured. Once the files have been converted to PDF files, they have to be renamed and saved, renaming them as the title of the documents. Once the documents have been renamed,

the scanned documents must be saved as a list in a folder.

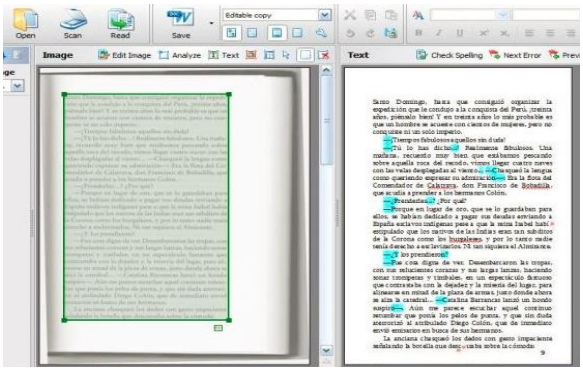


Fig-4: Convert the Document by Abby Fine Reader Software

5.7. MANAGE OF DIGITAL STORAGE

When it comes to preserving documents, the most important thing is digital storage, where documents are metadata entry, uploaded and archived. Creating digital storage requires digital library software such as dspace and computer, which help create digital storage. The digital library software is installed on a computer on the dspace and then a digital storage is created for configuration. If you want to install of Dspace, then you need some software such as - PostgreSQL database software, ApacheAnt, Apache Maven, Apache Tomcat, Java, Dspace, etc. If you install software, you can go and install of Dspace. After all the installation is done, who will configure it, such as what the metadata entry form will look like as documents, what metadata fields to use as per document, etc.

Required software's for create the digital storage by Dspace:

- Computer with Windows or Linux Version OS;
- PostgreSQL (Open-Source Relational Database Management System).
- Apache Ant (Software Tool for Automating Software Build Processes).
- Apache Maven (Build Automation Tool used Primarily for Java Projects).
- Apache Tomcat (Free and Open-Source Implementation of the Java Servlet, JavaServer Pages, Java Expression Language and WebSocket Technologies).
- Open JDK (Free and Open-Source Implementation of the Java Platform, Standard Edition); and
- Dspace (Digital Library Software).

5.8. DOCUMENTS SUBMISSION

Once the metadata entry and the document have been uploaded, it should be checked to see if the metadata entry as a document is correct. Whether all the fields are used as documents, if everything is correct, then the document has to be finalized by making a license agreement with Dspace. When the documents are finalized, the sets are preserved, that is, archived in digital storage. Shown below are some screenshots of how documents are submitted, and license agreement is made.

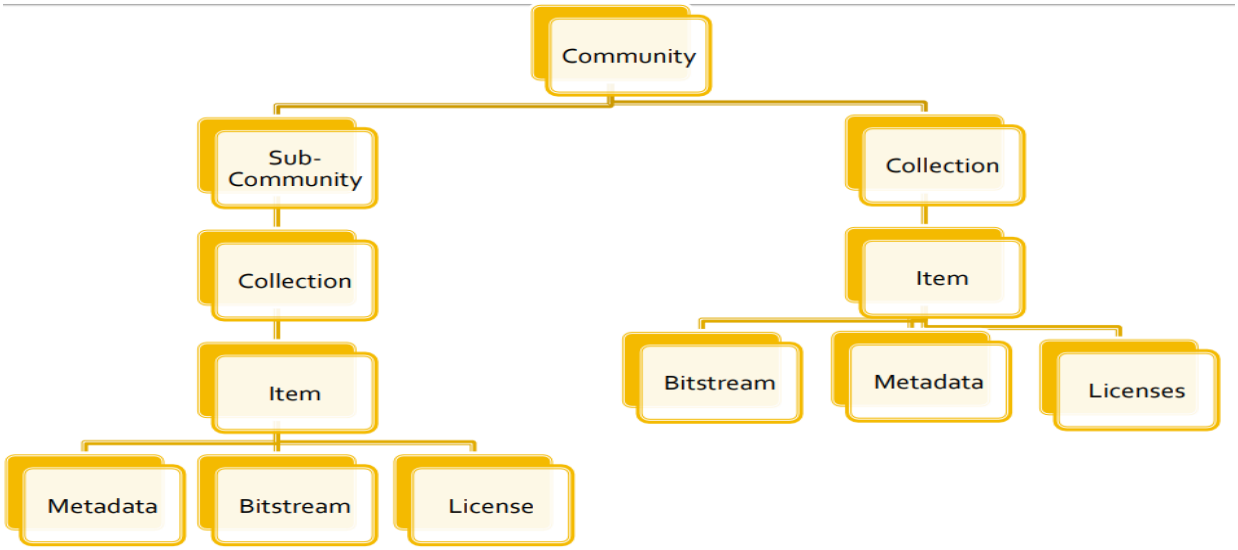


Fig-5: Document Submission Model

5.9. METADATA ENTRY & DOCUMENTS UPLOAD

When uploading documents to digital storage, you need to know what the metadata of the document is, you need to analyse all the metadata in a good way. Documents have to be uploaded by entering metadata through digital library software. Below are some step-by-step screenshots of how to upload documents to Dspace:

Steps: Login the digital storage by admin user id and password >Click on community >Click on collection >Click on the new item submission >Fill up the metadata in step by step

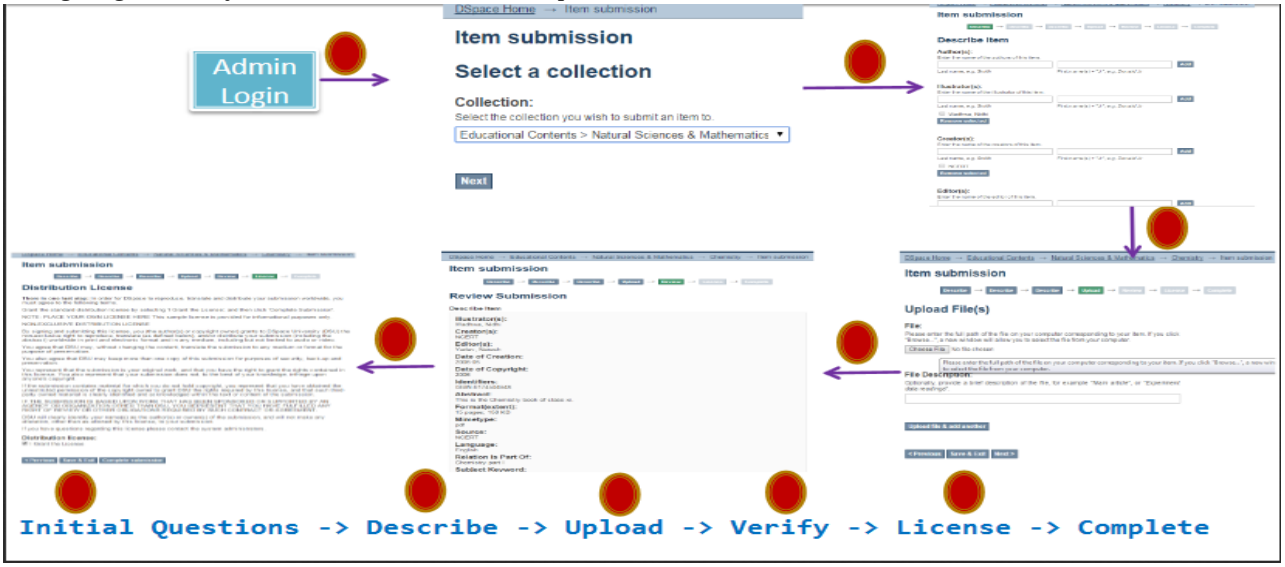


Fig-6: Metadata Entry & Document Upload

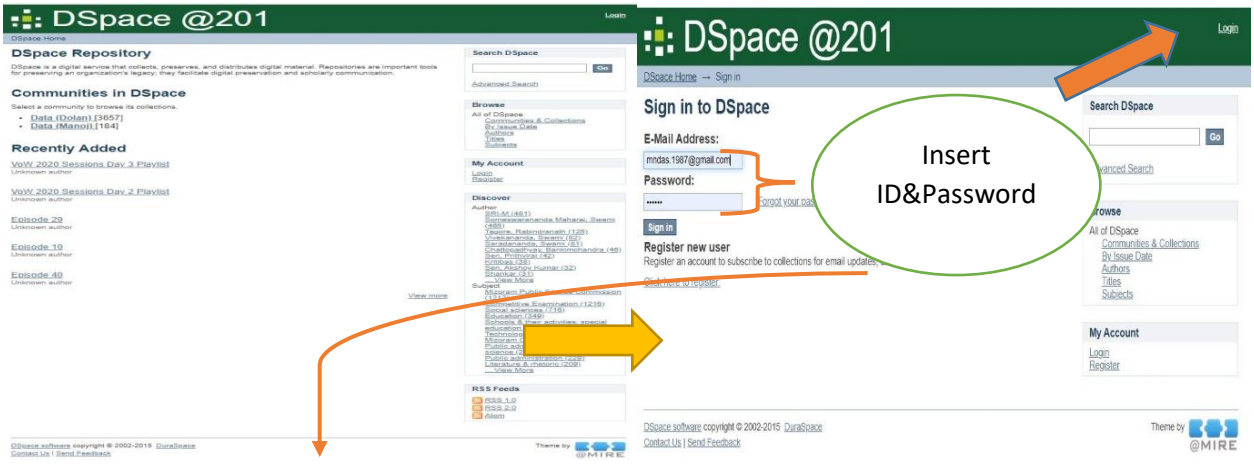
6. CURATION PROCESS

The curation process is that it is often seen that some errors remain even after the documents have been preserved, so that a second person is checked to make sure there are no errors. The second person is given permission to check so that he can check and edit. If the person finds any errors while checking, such as the author's name is misspelled or any other field and value is wrong, to correct those mistakes by editing. Once edited, the document is archived again. So how to check and edit and re-archive is discussed step by step below:

6.1. REVIEW THE SUBMITTED DOCUMENTS

When the documents are uploaded to digital storage, another person is allowed to review whether all the metadata fields and field values are correct:

Steps: Open the Storage > Click on Login >Login by ID & Password > Click on any Title > Click on Show full item record > Review the Metadata with field values.



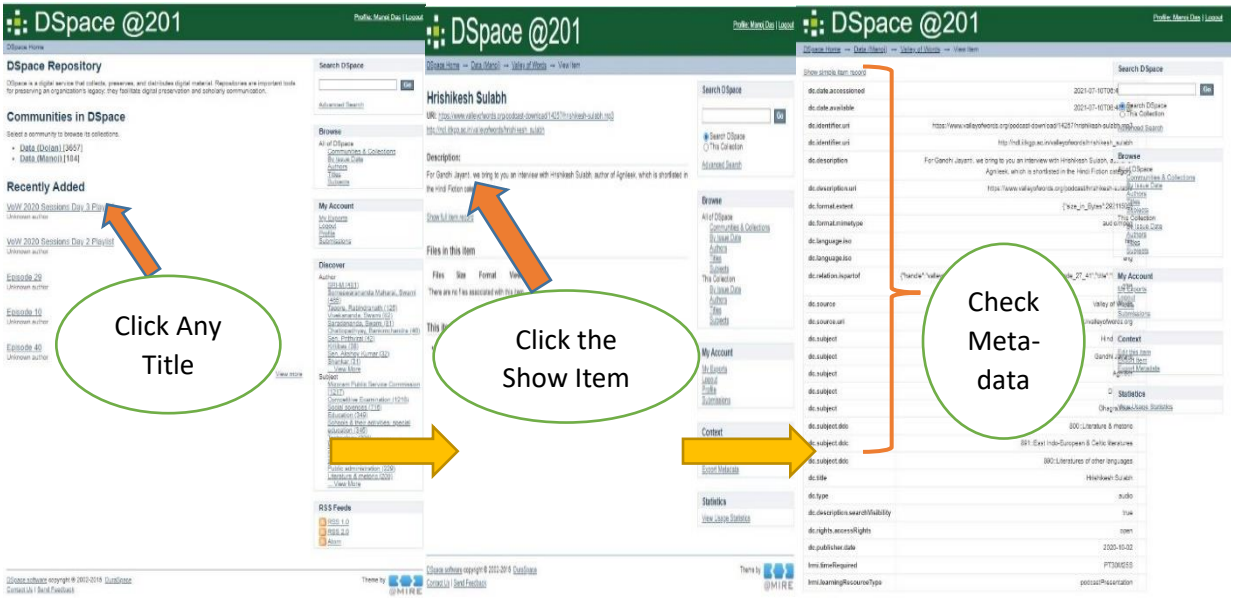
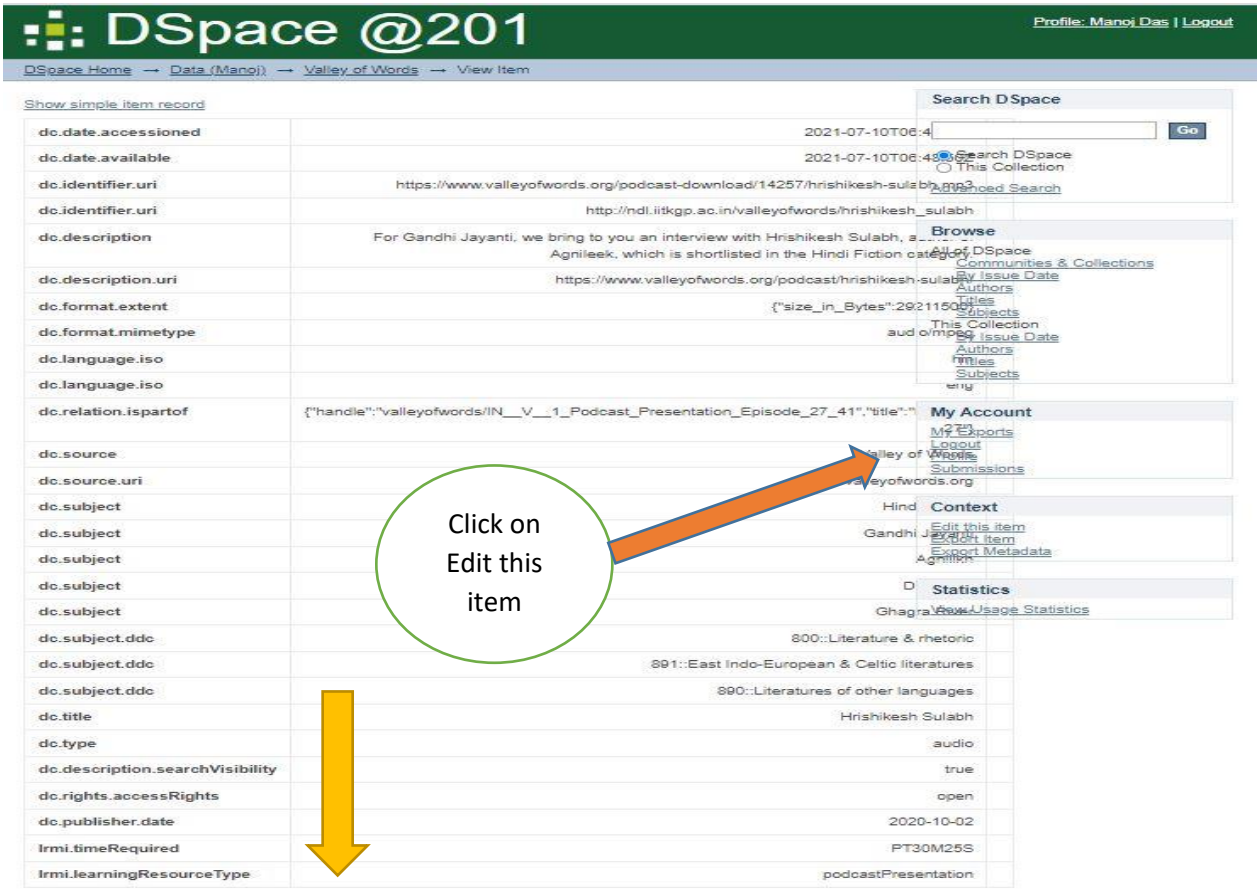


Fig-: Check the Metadata filed with Values

6.2. EDIT METADATA FIELDS WITH VALUES

If any metadata fields and values are wrong, then have to edit the proper metadata with values.

Steps: Click on Edit this item > Click on Item Metadata > Edit the Metadata fields and values



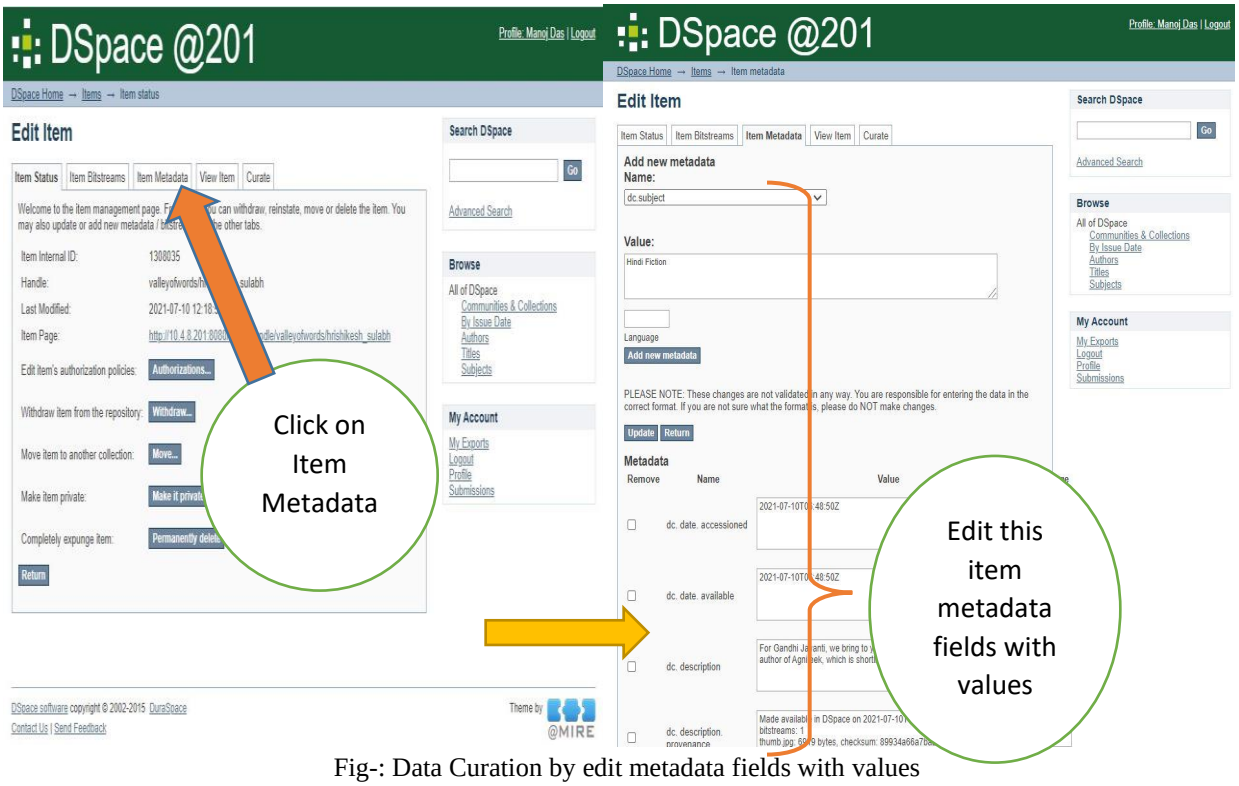


Fig-: Data Curation by edit metadata fields with values

6.3. ADD THE METADATA FIELDS AND VALUES

Steps: Select the Metadata field > Type the Field value > Click on Add new metadata > Click on Update

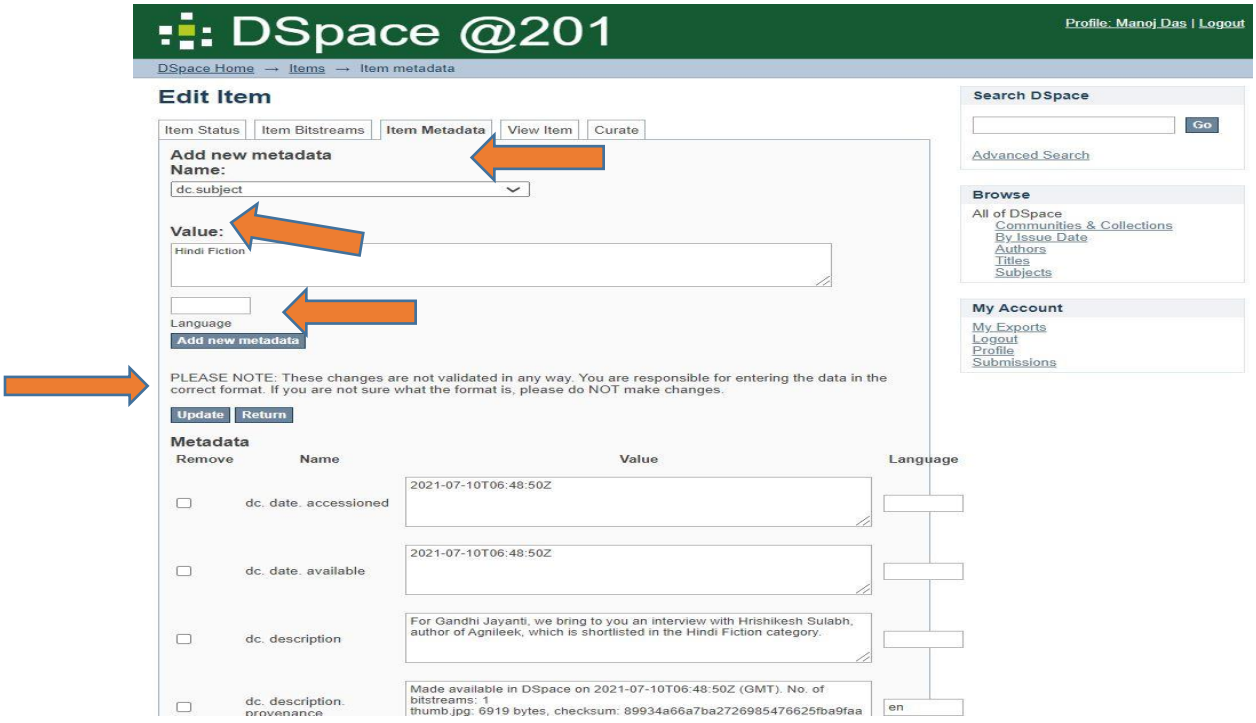


Fig-: Add the metadata fields and values

6.4. ARCHIVED DATA TO DIGITAL STORAGE

Steps: After edited > Click on Return > Archive data to Digital Storage

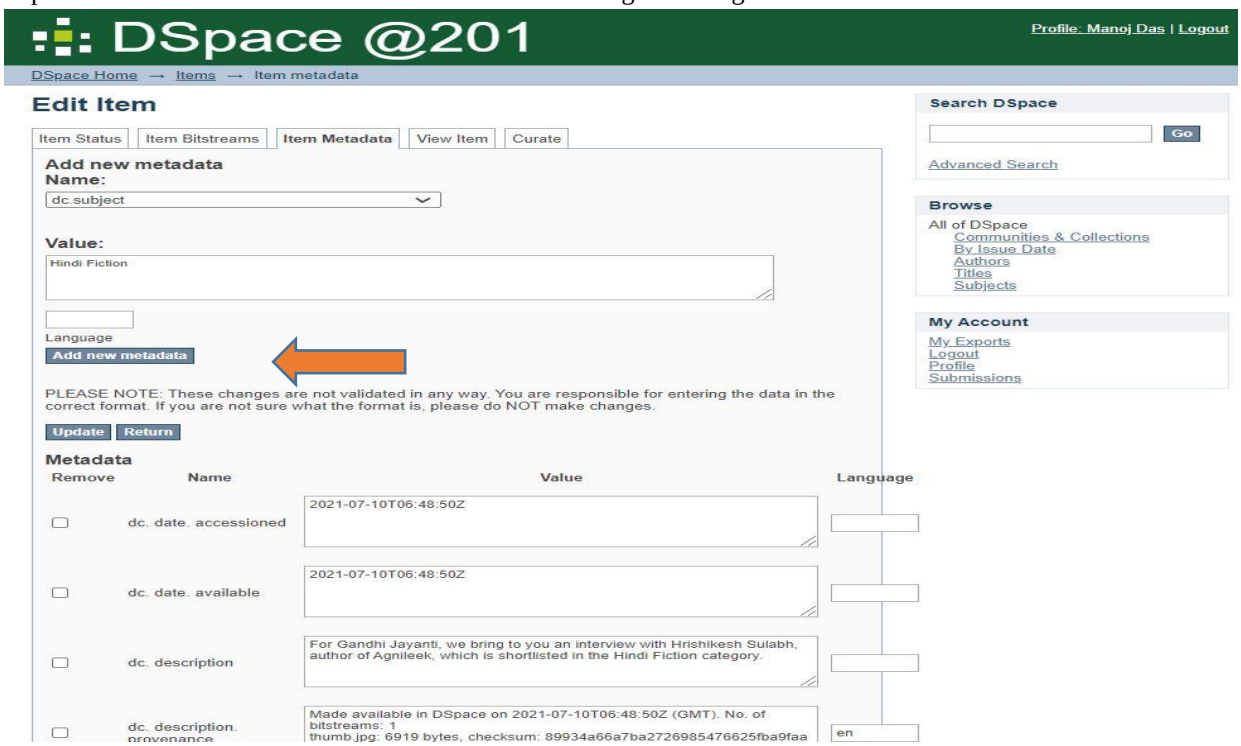


Fig-: Archive this data to digital storage

7. CONCLUSION

Digital Preservation and Correction plays an important role. All the documents that can be used as traditional documents are being used properly in this situation, so all those documents are scanned and converted into digital documents, and they are preserved. So that those digital documents can reach the current generation of students and the students can use those valuable documents for their study and research. If digital preservation is done, then those documents can be stored for a long time and can be used by students for a long time. When digital preservation is done, some importance should be given to it so that all the information based on documents is recorded through metadata, the resolution and cleanliness of the documents is right so that even the most ordinary students can understand it well. And the biggest thing is that if all the traditional and important documents are converted to digital and preserved, then the current generation of students will benefit a lot, because in the current situation students are not able to go to their educational institutions and continue their education. So if digital preservation is done and those digital documents are services online, students will be able to collect the documents and they need at home and continue their studies and research.

REFERENCES

[1] Beaudoin, J. E. (2012). Context and Its Role in the Digital Preservation of Cultural Objects. *D-Lib Magazine*, *18*(11-12), 1-11. doi:10.1045 /november2012-beaudoin1

[2] C., S. (2020). Digital curation practices in institutional repositories in South India: A study. *Global Knowledge, Memory and Communication*, *69*(8-9), 557-578. doi:10.1108/ GKMC-10-2019-0125

[3] Cunningham, A. (2008). Digital Curation/Digital Archiving: A View from the National Archives of Australia. *The American Archivist*, *71*, 530-543. Retrieved May 15, 2008.

[4] Doorn, P., & Tjalsma, H. (2007). Introduction: Archiving research data. *Arch Sci*, *7*, 1-20. doi:10.1007/s10502-007-9054-6

[5] Durante, K., & Hardy, D. (2015). Discovery, Management, and Preservation of Geospatial

Data Using Hydra. *Journal of Map & Geography Libraries, 11*(2), 123-154. doi:10.1080/15420353.2015.1041630

[6] Erkimbaev, A. O., Zitserman, V. Y., Kobzev, G. A., &Kosinov, A. V. (2019). Curation of Digital Scientific Data. *Scientific and Technical Information Processing, 46*(3), 192-203. doi:10.3103/S0147688219030092

[7] Gaur, R. C., & Tripathi, M. (2012). Role of consortia in preservation of e-journals. *Annals of Library and Information Studies, 59*, 204-211. Retrieved September 01, 2012.

[8] Gracy, K. F., & Kahn, M. B. (2009). Preservation in the Digital Age. *LRTS, 56*(1), 25-43.

[9] Gunia, B., & Sandusky, R. J. (2010). Designing Metadata for Long-Term Data Preservation: DataONE Case Study. *ASIST*, 1-2. Retrieved October 22, 2010.

[10] Hedges, M., Blanke, T., & Hasanc, A. (2009). Rule-based curation and preservation of data: A data grid approach using iRODS. *Future Generation Computer Systems, 25*, 446-452. doi:10.1016/j.future.2008.10.003

[11] Hedges, M., Hasan, A., & Blanke, T. (2007). Curation and Preservation of Research Data in an iRODS Data Grid. *Third IEEE International Conference on E-Science and Grid Computing*, 457-464. doi:10.1109/e-Science.2007.8

[12] Iorio, A. D., Schaerf, M., Guercio, M., Ortolani, S., &Bertazzo, M. (2014). A Digital Infrastructure for Trustworthiness: The Sapienza Digital Library Experience. *Verlag Berlin Heidelberg*, 59-69.

[13] Jordan, C., McDonald, R. H., Minor, D., &Kozbial, A. (2008). Cyberinfrastructure Collaboration for Distributed Digital Preservation. *Fourth IEEE International Conference on EScience*, 408-409. doi:10.1109/eScience.2008.163

[14] Júnior, L. S., & Borges, M. M. (2016). Fundamental concepts in digital preservation. *ACM*, 285-290. doi:10.1145/3012430.3012530

[15] Lancia, R. (2010). The DAF DDI Profile, a Metadata Set to Address Digital Curation and Preservation Issues in Cultural Heritage Institutions. *Humanities Advanced Technology and Information Institute*, 291-305.

[16] M, P., & J, W. J. (2015). Framing the Digital Curation Curriculum Conference. *Getting data creators on board with the digital curation agenda. Lessons learned in developing training for researchers* (Vol. 1016).

[17] Malone, J., Brown, A., Lister, A. L., Ison, J., Hull, D., Parkinson, H., & Stevens, R. (2014). The Software Ontology (SWO): A resource for reproducibility in biomedical data analysis, curation and digital preservation. *Journal of Biomedical Semantics, 5*(25), 1-13. Retrieved from http://www.jbiomedsem.com/content/5/1/25

[18] Marketakis, Y., Tzanakis, M., & Tzitzikas, Y. (2009). PreScan: Towards Automating the Preservation of Digital Objects. *ACM*, 404-411. Retrieved October 27, 2009.

[19] Mello, J., & Viera, A. F. (2020). Proposal of Model for Curation Digital Objects of an Oncology Research Center. *Social Informatics and Telecommunications Engineering*, 235-249. doi:10.1007/978-3-030-50072-6_18

[20] Molloy, L. (2014). Digital curation skills in the performing arts – an investigation of practitioner awareness and knowledge of digital object management and preservation. *International Journal of Performance Arts and Digital Media, 10*(1), 7-20. doi:10.1080/14794713.2014.912496

[21] Patel, M., Coles, S., Giaretta, D., Rankin, S., &McIlwrath, B. (2009). The Role of OAIS Representation Information in the Digital Curation of Crystallography Data. *Fifth IEEE International Conference on E-Science*, 132-139. doi:10.1109/e-Science.2009.27

[22] Patlakas, P., Musso, M., & Larkham, P. (2021). A digital curation model for post-occupancy evaluation data. *Architectural Engineering and Design Management*, 1-22. doi:10.1080/17452007.2021.1901219

[23] Perrin, J. M., Winkler, H. M., & Yang, L. (2015). Digital Preservation Challenges with an ETD Collection — A Case Study at Texas Tech University. *The Journal of Academic Librarianship, 41*, 98-104.

[24] Ray, J. (2017). Digital curation in museums. *Library Hi Tech, 35*(1), 32-39. doi:10.1108/LHT-12-2016-0154

[25] Ross, S. (2004). The Role of ERPANET in Supporting Digital Curation and Preservation in Europe. *D-Lib Magazine, 10*(7-8), 1-9. Retrieved August 01, 2004.

[26] Shah, C., & Marchionini, G. (2007). Capturing Relevant Information for Digital Curation. *ACM*, 1-1. Retrieved June 18, 2007.

[27] Sharma, S. (2021). Preservation and digitization in modern and heritage libraries of Jammu Province (J&K): An analytical study. *Annals of Library and Information Studies, 68*, 119-126. Retrieved June 01, 2021.

[28] Yakel, E. (2007). Archives and Manuscripts Digital curation. *OCLC Systems & Services: International Digital Library Perspectives, 23*(4), 335-340. doi:10.1108/10650750710831466