

Subject: **Library & Information Science**

LYCEUM INDIA

Title: **Exploring The Adoption of Digital Curation in Institutional Repositories: A Study on Universities of North-East India**

Journal of **Social Sciences**

Author(s): **Parimita Bezbaruah**
Ph.D. Scholar, Department of Library and Information Science
Gauhati University, Assam

Gedetchat Islary
Ph.D. Scholar, Department of Library and Information Science,
Gauhati University, Assam

Abstract: Digital curation is an integral part of the digital humanities, and the future of the digital humanities will increasingly depend on the digital resources being made available through institutional repositories and other platforms. The digital humanities depend on the digital form of the information, so it is really important to curate them carefully so that our future generations can easily use them and get maximum benefit. This study aims to gather data regarding the digital curation practices followed in the institutional repositories (IRs) of North-East India and also highlights the tools used to curate the digital content by the IRs. It also tries to understand the potential of IR software to curate digital materials. The online survey was conducted among 13 universities that had implemented IRs in their libraries, and the response rate was 76.92%. The findings show that most of the IR in North-East India followed digital curation practices, though not fully, at least to some extent.

Keywords: Digital Humanities, Digital Curation, Institutional Repository, Digital Library

1. Introduction

The 21st century is about digital data and technology, and users consistently depend on digital information. When digital tools and technology are used to visualize humanities disciplines, it is called digital humanities. It is an interconnection between information technology and the humanities. The digitization, analysis, and presentation of humanities content are covered in digital humanities (Sriram, 2017). Traditional humanities, in addition to technology, create digital humanities, where emerging methodologies and information technology play a fundamental role in processing and creating data. Digital humanities help users access a large amount of historical data over the web; various studies can be done on the progress of language; comparisons of maps can be done based on ancient places and present-day locations; and more. It's an attempt to curate data and present it in front of users so that they can access it not only today but also in the future.

Just like the traditional humanities depend on the existing stock of paper-based information, the digital humanities depend on the digital form of the information, so it is really important to curate them carefully so that our future generations can easily use them and get maximum benefit. Digital curation can be considered a process in which the management of valid information can be used today as well as in the far future. Also, it can help maximize the importance of research based on the digital humanities. Both digital humanities and digital curation play an important role in the scholarly world (Poole, 2017). In the digital humanities, digital curation actively does different tasks, including translating or migrating data into new formats, linking data sets or adding contextual information for works on varieties of products such as scholarly editions, thematic research collections, text corpora, marked-up text, earlier annotated or evaluated data, finding bibliographies,

and many more (Flanders & Munoz, 2012). Digital curation helps digital humanities get durable access, easy finding, recovery, and reuse, and make the best use of the curated content by making data flexible and functional. It also motivates and helps users do efficient and better research (Poole, 2017).

In this present world, digital materials are tremendously used in the day-to-day lives of common people; hence, they need to be carefully handled to increase their durability and maintain their authenticity in the process of communication. In doing so, all the fake data, news, and misinformation must be eliminated. All the output of research is not meant for preservation. Firstly, the degree of importance should be identified among the research outputs. Digital curation is the most practical way to meet this challenge and keep digital resources authentic and reusable for future users. It deals with actively maintaining digital research data and other digital materials over their entire life cycle and over time for current and future generations of users (Shajitha, 2020). The phrase 'digital curation' came into existence in a seminar called "Digital Curation: Digital Archives, Libraries, and E-Science," which was held in London in 2001 (Beagrie, 2006). In the seminar, various communities came together to discuss the urgent challenges of improving the long-term management of and preservation of access to digital information. There are different steps in digital curation: first, finding the required content and then selecting the best; then, it should be edited to add value to the data and arranged and organized for easy access. It is further followed by adding the author's and publisher's names and sharing to make this data available to others. The sharing of curated works is followed by engaging with the users by library and information science (LIS) professionals and also monitoring them to keep track of the development. In this way, digital curation added value to the research data; in other words, it can be said that the data is valuable if it is curated.

An institutional repository (IR) is a pearl of an institution that can work as a knowledge hub for its users. It provides digital access to institutional information, data, records, and scholarly publications. IR's functioning is an information management system that executes the different tasks of preserving, capturing, and getting access to serve the scholarly outputs of the academic members of the institution (Buragohain & Kumar, 2021). The main aim of the IR is to serve their institutional users and crack the control of toll-access publishers so that they curate research data, which also saves time for the scholar by removing unnecessary intellectual efforts. As IRs contain many types of data, such as research papers, presentations, theses, question papers, audio/video files, and case studies, it's really important to curate these digital data; otherwise, with time, they become obsolete. To curate this data and maintain its relevancy in the future, IRs must take the significance of digital curation seriously.

2. Literature Review

To determine the digital curation practices in IRs in South India, Shajitha (2020) survey the IR managers of 23 South Indian IRs and found that the majority of IRs in South India used Eprints software (60%) and DSpace (30%) to manage their content. Only a small percentage of IR students were found to practice digital curation techniques in this study. Poole⁷ investigated how digital curation could increase the return on investment for data projects, particularly in North America and the UK. Poole (2016) described the conceptual foundation of digital curation and its significance in enhancing the value of scholarly and other types of data for stakeholders. (Conceituais,2016) distinguished between digital preservation and curation; to do this, the author concentrated on the development of the concepts of preservation before addressing the idea of scientific data curation to create a framework for comparison between the two academic disciplines. After comparing the two ideas, the author emphasizes the reason why curation is growing in the presence of preservation. According to Higgins (2011), the iterative workshop and agenda-setting process led to the

emergence of the field of “digital curation,” and in the UK, the emphasis on long-term management of digital content eventually shifted from passive preservation to active curation. Digital curation, from the perspective of Poole (2017), aids in maximizing the potential of digital humanities. He examines the current digital curation perform and needs of scholars working in contemporary digital humanities, differentiating how closely these practices and needs align with the literature on digital curation. The author concludes by highlighting that the value of digital humanities and curation is growing daily, though there are still difficulties with terminology translation between the two fields, and that soft skills are also essential.

It is shown that most of the studies have been conducted in the US and UK to analyze digital curation, but in India, only a few articles talk about this topic. In India, a few studies have been done about the awareness and practice of university libraries in India to address digital curation, especially in north-east India. Digital curation activities in IRs are yet to be discussed. The present study tried to provide a clear vision of digital curation practices in IRs at universities in North-East India.

3. Statement of the Problem

The term “digital curation” refers to the essential actions required to protect and maintain digital research data and other digital materials, ensuring their longevity and accessibility for future generations. In North-East India’s universities, the current state of digital curation practices in institutional repositories remains mostly unexplored. To understand the present landscape, potential areas for improvement can be identified, allowing for the development of effective strategies to enhance digital curation practices and ensure the long-term preservation and accessibility of valuable digital resources in North-East India.

4. Objective

These are the three main objectives of this study:

- To identify the IR’s present status in state and central university libraries in North-East India
- To find out the digital curation practices in university’s libraries of North-East India
- To categorize the tools used to curate the digital content by the IRs in North-East India universities.

5. Scope and Limitation

The present study only focuses on universities in North-East India that are under 12B and 2F of the University Grants Commission (UGC). The researchers can locate 35 such universities, which include 11 central and 24 state universities. Interestingly, North-East India has 8 states, i.e., Arunachal Pradesh, Assam, Manipur, Meghalaya, Mizoram, Nagaland, Sikkim, and Tripura, Most of the universities, i.e., 20 are located in Assam only.

6. Methodology

The UGC website provides a list of 35 central and state universities (Table 1), and after analyzing their website, it was found that out of these 35 universities, only 13 universities have their IRs (Figure 1). All these 13 universities (37.14%) were mailed to get a response through the online questionnaire prepared based on the study by Shajitha (2020). A follow-up is carried out through email and telephone calls.

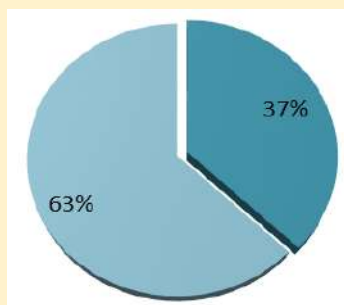


Figure 1: Universities of North-East India have IRs

Out of the 13 universities, 10 universities (76.92%) provided response (Table 2). Sikkim University, Manipur University and Nagaland University, these three universities did not respond to the survey. The data from the online questionnaire were enriched with data collected by the researchers by visiting the website of the respective IR.

Table 1: State and Central Universities in North-East India 10

| Sl. No. | Name of the Universities | Place | Types | IRs |
|---------|---|---------|--------------------|-----|
| 1 | Assam Agricultural University | Assam | State University | |
| 2 | Assam Rajiv Gandhi University of Co-operative Management | Assam | State University | |
| 3 | Assam Science & Technology University | Assam | State University | |
| 4 | Assam University | Assam | Central University | Yes |
| 5 | Assam Women's University | Assam | State University | |
| 6 | Bhattadev University | Assam | State University | |
| 7 | Birangana Sati Sadhani Rajyik Vishwavidyalaya | Assam | State University | |
| 8 | Bodoland University | Assam | State University | |
| 9 | Cotton College State University | Assam | State University | Yes |
| 10 | Dibrugarh University | Assam | State University | |
| 11 | Gauhati University | Assam | State University | Yes |
| 12 | Krishna Kanta Handiqui State Open University | Assam | State University | Yes |
| 13 | Kumar Bhaskar Varma Sanskrit & Ancient Studies University | Assam | State University | |
| 14 | Madhabdev University | Assam | State University | |
| 15 | Majuli University of Culture | Assam | State University | |
| 16 | National Law University and Judicial Academy | Assam | State University | Yes |
| 17 | Rabindranath Tagore University | Assam | State University | |
| 18 | Sri Sri Aniruddhdeva Sports University | Assam | State University | |
| 19 | Srimanta Sankaradeva University of Health Sciences | Assam | State University | |
| 20 | Tezpur University | Assam | Central University | Yes |
| 21 | Central Agricultural University | Manipur | Central University | |
| 22 | Dhanamanjuri University | Manipur | State University | |
| 23 | Manipur Technical University | Manipur | State University | |
| 24 | Manipur University | Manipur | Central University | Yes |

| | | | | |
|----|----------------------------------|-------------------|--------------------|-----|
| 25 | Manipur University of Culture | Manipur | State University | |
| 26 | National Sports University | Manipur | Central University | |
| 27 | North Eastern Hill University | Meghalaya | Central University | Yes |
| 28 | Mizoram University | Mizoram | Central University | Yes |
| 29 | Maharaja Bir Bikram University | Tripura | State University | |
| 30 | Tripura University | Tripura | Central University | Yes |
| 31 | Nagaland University | Nagaland | Central University | Yes |
| 32 | Khangchendzonga State University | Sikkim | State University | |
| 33 | Sikkim National Law University | Sikkim | State University | |
| 34 | Sikkim University | Sikkim | Central University | Yes |
| 35 | Rajiv Gandhi University | Arunachal Pradesh | Central University | Yes |

7. Software for Institutional Repositories

As per the study, most of the IRs in universities used DSpace software (90%), excluding Cotton University, which uses Eprints (10%). This clearly shows the inclination to only specific software, i.e., DSpace, instead of people opting for a diversifying list of software especially in North-East India. In terms of the version of the software, it was found that none of the universities uses DSpace 7.5, which is the latest version available in the market. In terms of the development of the IR, 5 universities (5s) developed their IRs on their own, whereas 3 universities took the help of the Information and Library Network (INFLIBNET) Centre, 1 university each took the help of the National Digital Library of India and the commercial vendor available in the market.

Table 2: Basic Information on the IRs of Universities in North-East India

| Sl. No. | Name of the Universities | Website of the Repository | Software Used | Software Version | Developer |
|---------|--|---|---------------|------------------|-------------------|
| 1 | Assam University | http://idr.aus.ac.in/jspui | DSpace | 6.3 | Developed by Own |
| 2 | Cotton College State University | http://cotton.informaticsglobal.com/ | Eprints | 3.3.15 | Commercial Vendor |
| 3 | Gauhati University | http://10.10.114.8:8080/jspui | DSpace | 5.3 | Developed by Own |
| 4 | Krishna Kanta Handiqui State Open University | http://dlkhsou.inflibnet.ac.in | DSpace | 6.1 | INFLIBNET Centre |

| | | | | | |
|----|--|---|--------|-----|-----------------------------------|
| 5 | National Law University and Judicial Academy | http://www.dlnluassam.ndl.iitkgp.ac.in/ | DSpace | 5.2 | National Digital Library of India |
| 6 | Tezpur University | http://etdr.tezu.ernet.in/ | DSpace | 6.2 | Developed by Own |
| 7 | North Eastern Hill University | http://dspace.nehu.ac.in/ | DSpace | 6.3 | Developed by Own |
| 8 | Mizoram University | http://mzuir.inflibnet.ac.in | DSpace | 6.2 | INFLIBNET Centre |
| 9 | Tripura University | http://172.16.32.12:8080/js pui/ | DSpace | NA | Developed by Own |
| 10 | Rajiv Gandhi University | http://rguir.inflibnet.ac.in | DSpace | 6.3 | INFLIBNET Centre |

8. Metadata Schema and Others

All universities preferred Dublin Core metadata for their metadata schema. Other metadata standards like Metadata Encoding and Transmission Standard (METS), Metadata Object Description Schema (MODS), Text Encoding Initiative (TEI), Anglo-American Cataloguing Rules (AACR2), etc. are not used by any IRs. This is clearly an indication of the dominance of Dublin Core in the metadata.

There are 2 IRs in North-East India that have more than 10 thousand collections. All other IRs is still in the development phase and has only a minimal collection.

The development of the IR in North-East India started with North Eastern Hill University in 2010. It was followed by a huge growth in 2017 and all other IRs was established before and after 2017.

Table 3: Metadata Schema Used in Institutional Repositories of North East India

| Sl. No. | Name of the Universities | Metadata Schema | Total Collection | Year of Establishment |
|---------|--|-----------------|------------------|-----------------------|
| 1 | Assam University | Dublin Core | 1300 | 2016 |
| 2 | Cotton College State University | Dublin Core | 2584 | 2017 |
| 3 | Gauhati University | Dublin Core | 330 | 2017 |
| 4 | Krishna Kanta Handiqui State Open University | Dublin Core | 546 | 2017 |
| 5 | National Law University and Judicial Academy | Dublin Core | 215 | 2019 |
| 6 | Tezpur University | Dublin Core | 1319 | 2016 |
| 7 | North Eastern Hill University | Dublin Core | 11565 | 2010 |
| 8 | Mizoram University | Dublin Core | 1153 | 2017 |
| 9 | Tripura University | Dublin Core | 282 | 2021 |
| 10 | Rajiv Gandhi University | Dublin Core | 16053 | 2020 |

9. Controlled Vocabulary

In terms of the controlled vocabulary, a few IRs used the Sears List of Subject Headings (SLSH), which is followed by the Library of Congress Classification (LCC) and Library of Congress Subject Headings (LCSH). However, most of the IR still preferred to use their vocabulary, which is partially based on the user's search history in the library databases; for example, in Assam, users mostly used the "Biya" word instead of "Marriage", and hence there is an inclination towards this end.

Table 4: Controlled Vocabulary Used in Institutional Repositories of North East India

| Sl. No. | Controlled Vocabulary | Frequency N=10 | Percentage (%) |
|---------|-----------------------|-------------------|-------------------|
| 1 | In-house built | 6 | 50 |
| 2 | SLSH | 3 | 25 |
| 3 | LCC | 2 | 17 |
| 4 | LCSH | 1 | 8 |
| 5 | Other | 0 | 0 |

10. Persistent Identifier

A persistent identifier (PID) can be defined as a long-lasting reference to a digital object like a document, file, web page, or other object that can be accessed over the internet. It is not only constant but also functional; when it is put in the web browser, it has taken to the recognised source. For example; for a person, an orchid id could be his PID, and for a digital document Digital Object Identifier (DOI) could be its PID. It was noticed that 50% of IRs used the Uniform Resource Identifier (URI), followed by Handel (30%). DOI and Persistent Uniform Resource Locator (PURL) are the less-used PIDs (10% each) to curate the digital material.

Table 5: Persistent Identifier and its Adoption in IRs of North East India

| Sl. No. | Persistent Identifier | Percentage (%) |
|---------|-----------------------|----------------|
| 1 | URI | 50 |
| 2 | Handel | 30 |
| 3 | DOI | 10 |
| 4 | PURLS | 10 |
| 5 | ARKS | 0 |

11. Applications for Managing Digital Content

A majority of the IRs used Apache Solr (90%) to manage the digital content, which is an open-source search system and an integral part of DSpace itself. However, even if it is not integral, as in the case of Cotton University, which used Eprints software, they still used Apache Solr as an application tool. Only a single university uses Simple Text Editor in their IR and File Information Tool Set, Almetrics and Openrefine were not being used at all to curate their digital data.

Table 6: Application Tools Used in IRs of North East India

| Sl. No. | Application Tool | Percentage (%) |
|---------|---------------------------|----------------|
| 1 | Apache Solr | 90 |
| 2 | Simple Text Editor | 10 |
| 3 | File Information Tool Set | 0 |
| 4 | Almetrics | 0 |
| 5 | Openrefine | 0 |
| 6 | No Tools | 0 |

12. Conclusion

Digital curation is still considered a new concept in a developing country like India, where people are trying to adapt library services. Almost all the IR's used Dublin Core as the metadata schema. People from LIS backgrounds, though familiar with ready-mades-controlled vocabulary like SLSH, LCC, and LCSH, however, still they feel inclined towards in-house built vocabulary for the IRs. The users' search is influenced by their environment, place, regional language, etc. and it may be a reason for this inclination. The most used PID is URI, and in terms of application and tools to make the content it is Apache Solr. The present study highlights digital curation practice in IRs in universities in North-East India and explores the attention given by IRs to digital curation activities. It also tries to understand the potential of IR software to curate digital materials.

References:

1. Beagrie, N. (2008). Digital curation for science, digital libraries, and individuals. *International Journal of Digital Curation*, 1, 3–16. <https://doi.org/10.2218/ijdc.v1i1.2>
2. Buragohain, D., & Kumar, A. (2021). An Analytical Study of Managing Institutional Repositories by University Libraries in Assam. *Library Philosophy and Practice*. <https://digitalcommons.unl.edu/cgi/viewcontent.cgi?article=11676&context=libphilprac>
3. C, S. (2020). Digital curation practices in institutional repositories in South India: a study. *Global Knowledge Memory and Communication*, 69(8/9), 557–578. <https://doi.org/10.1108/gkmc-10-2019-0125>
4. Higgins, S. (2011). Digital Curation: the emergence of a new discipline. *International Journal of Digital Curation*, 6(2), 78–88. <https://doi.org/10.2218/ijdc.v6i2.191>
5. *Persistent Identifier*. (n.d.). Wikipedia. Retrieved May 9, 2023, from https://en.wikipedia.org/wiki/Persistent_identifier/
6. Poole, A. H. (2016). The conceptual landscape of digital curation. *Journal of Documentation*, 72(5), 961–986. <https://doi.org/10.1108/jd-10-2015-0123>
7. Poole, A. H. (2017). "A greatly unexplored area": Digital curation and innovation in digital humanities. *Journal of the Association for Information Science and Technology*, 68(7), 1772–1781. <https://doi.org/10.1002/asi.23743>

8. Santos, T. N. C., & De Albuquerque Siebra, S. (2016). Curadoria digital e preservação digital. *RDBCI Revista Digital De Biblioteconomia E Ciência Da Informação*, 14(3). <https://doi.org/10.20396/rdbci.v14i3.8646336>
9. Sriram, Ramya. "The Rise of Digital Humanities: Five Interesting Applications." *The Kolabtree Blog*, 25 July 2017, <https://www.kolabtree.com/blog/digital-humanities/>.
10. *State Universities - University Grants Commission*. (n.d.). Retrieved May 9, 2023, from <https://www.ugc.gov.in/stateuniversity.aspx>.

Citation in APA Style:

Parimita Bezbaruah, & Gedetchat Islyary. (2025). Exploring The Adoption of Digital Curation in Institutional Repositories: A Study on Universities of North-East India. *Lyceum India Journal of Social Sciences*, 1(6), 5–13. <https://doi.org/10.5281/zenodo.15077404>

