

# Restructuring and Upgradation of the DSpace-based Institutional Repository of CEPT University Library: A Case Study

Namrata Dobariya<sup>1</sup> and Gayatri Doctor<sup>1</sup>

<sup>1</sup>CEPT University Library, CEPT University, Ahmedabad, Gujarat, India

## Abstract

*This paper aims to present a strategic overview of the restructuring of the CEPT repository structure, maintaining and upgrading its software for continuous access and improved services. The process of restructuring and upgrading CEPT University Library's repository to DSpace 7 involved identifying the need for technical proficiency, establishing a new server, configuring the repository, and migrating existing data while preparing metadata. Furthermore, data in the new storage hierarchy was restructured, and SSL integration with Shibboleth was implemented. The paper highlights the benefits and outcomes of the restructuring process, emphasizing the enhanced accessibility, discoverability, and preservation of scholarly works. The study discusses the challenges faced by the institution's existing repository, the rationale behind the restructuring, and the step-by-step approach taken to upgrade the repository. The case study serves as a valuable reference for other academic institutions seeking to optimize their institutional repositories.*

**Keywords:** CEPT Repository, DSpace, Electronic Theses and Dissertations (ETD), Institutional Repositories

## 1. Introduction

The CEPT University Library, which is named Lilavati Lalbhai Library (LLL) in the year 2017, is considered one of the best libraries in India to get habitat-related resources. The library supports teaching, learning, and research at the University. The library stacked some of the best collections of books, references, oveoversizedoks, rare books, special publications of ISI/ BIS, electronic resources, IS codes, non-book material, faculty publications, maps, etc.

The library's mission is to share information with the community of users, advance knowledge, and facilitate connections. The means by which researchers, scholars, and students collect, access, and disseminate research results continue to change with changes in technology. New ways of discovery and use, new mechanisms of collecting, storing, cataloguing, and sharing information, and new forms of publishing are demands of the time. The digitization of resources is moving one step forward towards this new trend.

The CEPT Library houses many internally generated resources. It includes the final year thesis reports of all postgraduate, undergraduate, and doctoral programs. Students' reports and drawings prepared during the coursework are also part of the library collection. It was decided that all thesis reports would be digitized and made available online initially by developing the institutional repository. The IR has further helped the institution to optimise resource sharing at local, regional, and international levels. This achievement could have not been possible had the library continued only stocking hard copies of theses and dissertations. (Nason B. Bimbe, 2017).

During COVID, students were not on campus and had no access to print copies of thesis reports. The repository was thus created and made available since it was required to provide students with online access to all thesis reports available in the library.

### **1.1 Institutional Repository of CEPT**

Institutional repositories were originally developed to provide a solution for the collection, preservation and dissemination of the research output created at universities and research institutions. The importance of such repositories was quickly recognized since they provided the means for the institutions to showcase their academic work. Using open-source software gave the ability to developers worldwide to build custom features and functionalities for the systems. (Marios Zervas, 2019). The CEPT Repository was implemented with the collection of 8,000+ final-year undergraduate and postgraduate theses and PhD dissertation reports in September 2020. The CEPT repository can be accessed from <https://repository.cept.ac.in>.

The existing CEPT Repository model was designed to be centralized and controlled by CEPT University library staff. Students and Faculty can submit their resources to the library in digital form. The library team oversaw the overall workflow, which included exporting metadata from the Library Management System KOHA, converting data into Dublin Core XML format, and uploading it to the DSpace server.

According to university policy, it was decided that the repository content would be restricted and accessible only to CEPT communities and that control over the shared PDFs would be retained. The major thesis collection in the repository is available to the CEPT internal community only, with their logins, as full-text access to all the repository content. Every item uploaded under the category of Theses and Directed Research Project (DRP) reports is uploaded in pdf. file format, which can be accessed by clicking on the preview pdf. It was achieved by embedding a PDF JS viewer because it establishes the conditions for the viewing experience. The current contents can be searched by title, author, guide, and date of publication.

## **2. Objectives**

The main purpose of this paper is to provide an outline of the strategic repositioning of the CEPT repository structure, including maintaining and upgrading its software for 24x7 access and improved services. Furthermore, digitisation has helped to market the library resources and services to the user community and improve liaison with the institution's faculties and schools, because the library collection and services now to a larger extent meet the changing needs of the users. (Bimbe, 2017).

The CEPT repository was built on DSpace software version 6.3. With the developments in technology, the DSpace community released the latest version of the software with the introduction of a new Angular-based User Interface that brings together the best of JSPUI and XMLUI. Users conveyed several queries by email and in person regarding search issues, including difficulties in rapidly accessing content and multiple clicks to access the desired information. The new release version of DSpace 7, is backed by a brand-new REST API, which opens all data and features to the web, allowing it to integrate with external systems easily.

CEPT University embraces five distinct faculties/departments. Hence, the existing repository collections were structured by the department; however, considering the responses received from the users, repository administrators have decided to include another collection, such as “Faculty Publications”, in the repository. Therefore, it was necessary to restructure them by collection, ensuring that the hierarchical structure remained intact.

Technological advancements and evolving user expectations require repositories to adapt and offer enhanced user experiences. Challenges may include maintaining data integrity during migration, ensuring backward compatibility, and minimizing downtime. On the other hand, opportunities arise from improved user interfaces, expanded search capabilities, better metadata management, and integration with emerging technologies.

### **3. Methodology**

The library carried out an extensive review of the existing DSpace repository to discover its limitations and opportunities for development. The repository’s day-to-day user queries and difficulties were explored. It was discovered that finding relevant content often demanded multiple clicks and steps. From an administrative perspective, the ability to add users in bulk was lacking, and authorization was needed for each item. Consequently, the assessment identified issues with the user interface, search functionality, metadata handling, and overall system efficiency. Nonetheless, DSpace 7 introduces features designed to address these concerns comprehensively. Based on the reviews, the library formulated a comprehensive plan to upgrade and restructure the repository.

#### **3.1 Assessment and Planning**

Several key factors require careful consideration for a successful transition when upgrading any DSpace-based repository. First and foremost, a comprehensive backup of data, configurations, and customizations is essential to prevent any potential data loss. Ensuring compatibility of existing customizations, themes, plugins, and add-ons with the new version is crucial to maintaining functionality. Following the provided upgrade documentation meticulously is paramount, as it outlines the specific steps needed for a smooth transition. Rigorous testing in a controlled environment helps to identify and address potential issues before applying changes to the live repository. Thoughtful planning and testing of data migration, including items, metadata, collections, and user accounts, must be conducted to safeguard data integrity. Ensuring compatibility of third-party tools and dependencies, adjusting configurations, informing users about changes, preparing a rollback plan, providing training, implementing monitoring tools, and seeking support from the DSpace community and technical experts all contribute to a successful upgrade process.

Before upgrading and restructuring, a comprehensive assessment of the existing repository was imperative. The project has been planned across three quarters to achieve the work in the appropriate time frame. During Quarter One, the focus encompassed several tasks: creating a repository backup, establishing an environment conducive to installing DSpace 7 on a new server, tailoring the user interface to align with the university's requisites and policies, and effectively managing metadata.

In Quarter Two, the project transitioned into its second phase, involving the execution of the migration process, the meticulous restructuring of the repository, the upload of metadata accompanied by full-text content, the authorization of content, and a rigorous assessment of the uploaded data to ensure impeccable quality.

In Quarter Three, a new collection, "Faculty Publications," will be added to the repository, and the repository will be made live with the necessary integration of SSL and Shibboleth user access management in due course. The decision to move the repository to a new version during the summer break is driven by most users who will be engaged in internships during that period, leading to minimal repository usage. This strategic timing will minimize disruptions to repository access, facilitating seamless completion without the need for an upgrade during active use.

### **3.2 Technical Implementation**

The technical implementation phase involved selecting the appropriate DSpace version for upgrade, setting up a testing environment, and ensuring compatibility with existing integrations. Migrating content, metadata, and user data requires meticulous planning to prevent data loss and maintain continuity. During this phase, repository administrators collaborated closely with the CEPT IT team to address technical challenges, optimize workflows, and ensure a smooth transition. In collaboration with the CEPT IT Team, a dedicated standalone server was set up, meeting the necessary requirements for the establishment of a DSpace 7 instance.

### **3.3 Prerequisites for Setting up DSpace 7**

The following prerequisites are to be installed on the test server.

- ❖ Java Oracle JDK 11
- ❖ PostgreSQL 12.x or 13.x (with pgcrypto installed)
- ❖ Apache Solr 8x (full-text index/search service)
- ❖ Tomcat 9,
- ❖ Node JS 14 +

It was created on an Ubuntu server 20.X with a higher GUI (Graphical User Interface), with the increased storage volume as required. Then, the new DSpace 7.4 instance was successfully installed with the help of the CEPT IT team.

### 3.4 Configuration

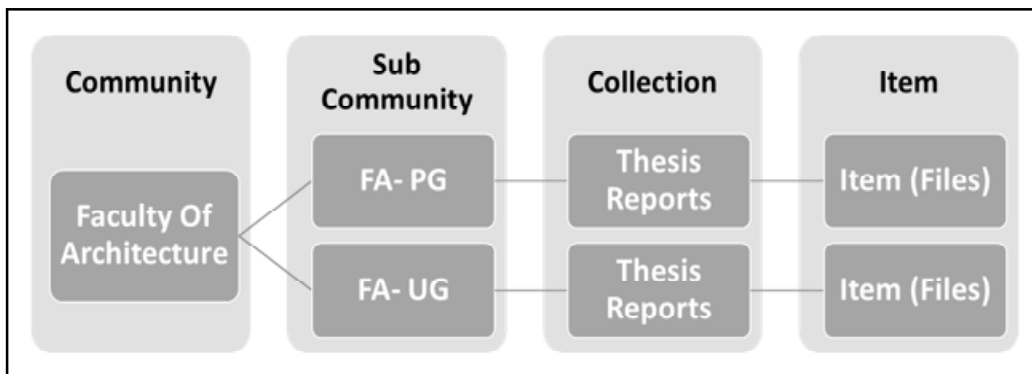
The new repository layout was designed to align with the defined criteria and branding directives of the university. CEPT University has a branding policy to develop and manage any web page, like logos, fonts, colours, and landing page themes, all in accordance with the approved branding guidelines. The repository's home page has also been updated with a disclaimer.

Prior to commencing the migration process, a backup of the existing repository was generated.

The data migration was carried out with the same structure as the existing repository, which included metadata, bitstreams (files), and user data. Following the successful migration, comprehensive guidance was provided to the library staff on how to manage the new DSpace 7 instances. The library staff is quite acquainted with the existing repository version, DSpace 6.3. Hence, minimal training is required to handle the new DSpace instance.

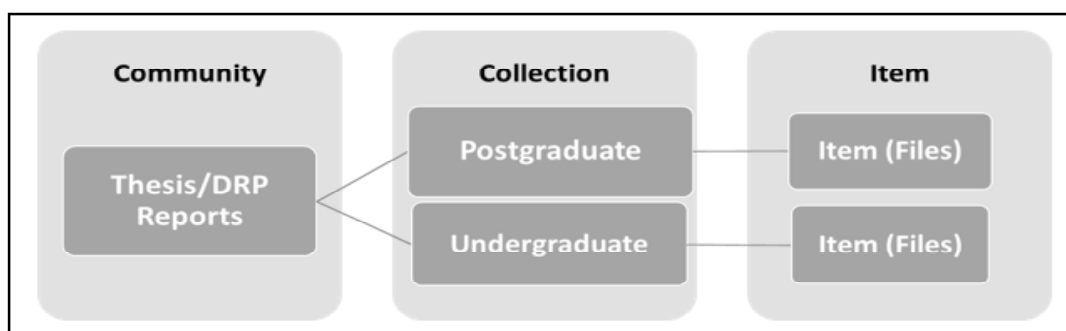
### 3.5 Restructuring of the Repository

At the time of the repository's initial implementation, only thesis reports were required to be added because CEPT University final year thesis reports and PhD dissertations were only available in print form. There was a COVID time to implement it and provide access to the users because thesis reports and dissertations were in high demand by the users. Thus, the repository was structured according to the academic departments (Faculties), as illustrated in Figure 1.



**Figure 1: Existing Repository Structure**

However, after creating an institutional repository, it became apparent that there was a need to develop the repository's extent by incorporating additional collections. This aimed to enrich institutional content and facilitate seamless access for users. This realization prompted the requirement for a restructuring of the existing data. The purpose was to align the data with a new storage hierarchy, resulting in a revised version of the repository's structure, as illustrated in Figure 2.



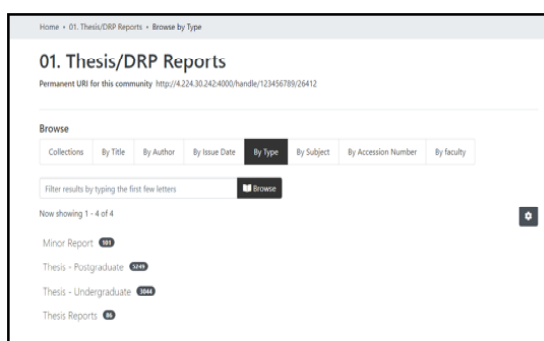
**Figure 2: New Repository Structure**

However, it was challenging to restructure the full-text data, as it was necessary to restructure the entire community. It was challenging to migrate the data with its bitstreams (full-text files) into the new structure as the collection code (Handle URL) changed while moving from one community to another. Hence, it was decided to create a bulk import template and upload all metadata with its bitstreams (files). Thus, metadata was prepared to align with the updated version configuration.

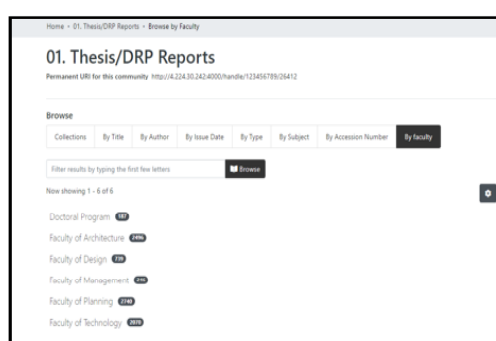
### 3.6 Metadata Management:

The new version incorporates a data filtering feature and has included additional metadata fields featuring drop-down selections. This empowers users to efficiently explore desired data by choosing specific categories. The CEPT repository has different departments and item types to filter the search through; hence, it was added dropdown values in the metadata fields such as departments (Faculty of Architecture, Faculty of Design, Faculty of Planning, etc.) as shown in Figure 3 and item types (Minor reports, Postgraduate/Undergraduate Theses, etc.) as shown in Figure 4. Consequently, users can easily locate the data they are seeking based on their preferences.

**Figure 3: Filter search by Faculty**



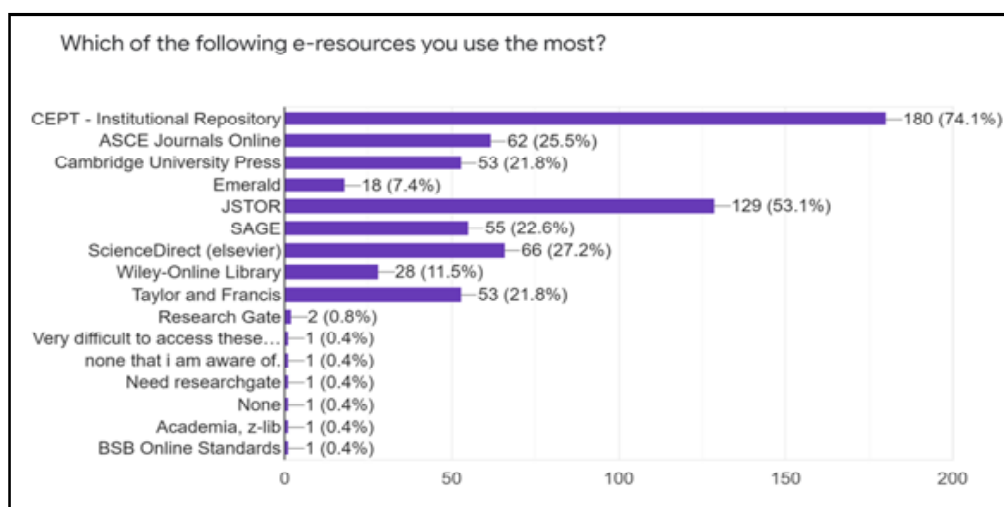
**Figure 4: Filter search by Type**



Over 8000+ files from the existing repository were copied onto a local disk, and metadata was created containing links to the full-text PDF paths for uploading through the bulk upload template. That process consumed a significant amount of time and required careful consideration of factors such as internet speed,

server capacity, and file sizes to ensure a smooth upload. The integrity of the data was verified after uploading to ensure that all metadata, files, and relationships between items were uploaded accurately.

Based on the user’s overall library services and resources feedback survey, it was identified that the repository is the most used resource among the other e-resources as shown in Figure 5. Therefore, it was decided to add other collections to the repository, and it became essential to restructure the existing repository by creating new communities and collections in the new DSpace 7 instance.



**Figure 5: e-Resources Feedback Survey**

Another essential intellectual asset to include in the collection was faculty publications. To increase the visibility of CEPT’s scholarly endeavours, it was decided to include faculty publications in the repository, and the bibliographic information for each faculty research with its source link was uploaded.

Following a comprehensive review of all the uploaded content to the new DSpace instance, the requirement to assign a domain to the instance emerged. This allocation was necessary to facilitate essential configurations such as SSL (Secure Socket Layer) setup to make it live, integration with Shibboleth, and single sign-on (SSO) for user access management. Consequently, the decision was made to assign a subdomain to the new DSpace instance, ensuring that the primary domain “repository.cept.ac.in” remained unaffected by any downtime. This approach enables users to continue accessing resources using the old Uniform Resource Locator (URL) without disruption, and in the meantime, user access will be tested.

### 3.7 User Access Management

The CEPT Library has been using a single sign-on (SSO) authentication technique that allows users to safely log in to different apps and websites with only one set of credentials. The INFLIBNET Center recommended using an INFED - (INFLIBNET Access Management Federation) for user authentication and access control, hence, the Library Committee decided to become a member of INFED.

The Indian Access Management Federation (INFED), the first Federation in India, has adopted Shibboleth, a standard-based open-source software, for authenticating authorized users from colleges and universities and providing them seamless access to e-resources from anywhere, anytime (INFED, 2023).

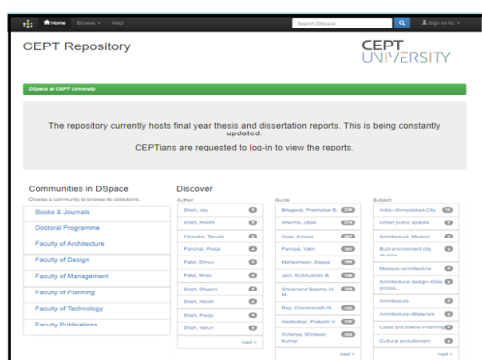
The Shibboleth System allows authorized users to be authenticated using an organization’s internal identity and access management system. Shibboleth does not provide authentication services. Instead, it establishes a set of protocols for the secure exchange of identity information between institutions and service providers, with each participating institution responsible for establishing its own identity provider services for its users. Hence, Shibboleth was used by the CEPT Library to manage authentication with IR (Institutional Repository) services and allow users to use university credentials to log in.

Shibboleth was set up on the Apache server and integrated with the existing DSpace repository version 6.3. However, when attempting to integrate Shibboleth with DSpace 7, compatibility issues arose due to the distinct server configuration. The DSpace 7 instance was configured with a Nginx server. While there was an effort to incorporate Shibboleth with Nginx, considering that DSpace 7 was already set up with data, this endeavour was complicated by the lack of up-to-date documentation, and it became clear that integrating Shibboleth with Nginx would be time-consuming to explore. To streamline the launch of the new DSpace 7 instance, a decision was made to transition DSpace 7 from the Nginx server to the Apache server. This change is currently in the final testing phase.

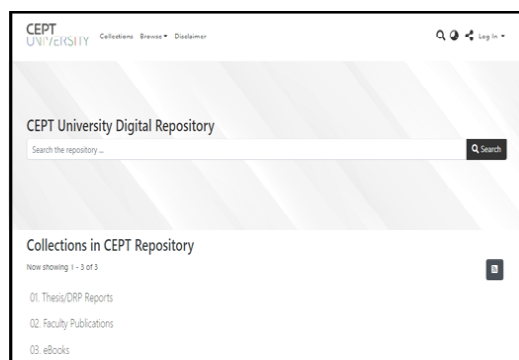
Once everything is set up, tested for user access, and the handle configuration is done, there will be a switch to the new repository software by updating the Domain Name Server (DNS) to point the domain to the new DSpace instance. This would make the new instance live for users.

The following provides illustrations of both versions of the repository. Figure 6 depicts the homepage of the current DSpace 6.3 repository, while Figure 7 showcases the homepage of the newly implemented DSpace 7 instances.

**Figure 6: Existing Repository Instance (DSpace 6.3)**



**Figure 7: New repository Instance (DSpace 7)**





#### **4. Challenges**

To complete this project, it has faced many challenges, such as the time-consuming installation due to a lack of backend software support. The existing repository was set up with the CentOS operating system when DSpace 7 required the Ubuntu Operating system. Hence, it was moved from CentOS to Ubuntu to build an environment for installation.

For the migration of the data, it was initially decided to create a mirror server, which takes a lot of server space and is expensive too. Migration was complicated as it could not move bitstreams (full-text data) with metadata, resulting in only users' data being successfully migrated. So, all content had to be uploaded fresh, along with its corresponding content. To streamline the process, a bulk upload template as a separate configuration was created.

Unfortunately, the server faced issues like the campus server being down for some unforeseen reason while uploading data, causing the deletion of some supportive software. It was necessary to reinstall the essential software components.

Integrating Shibboleth with DSpace 7 became complex due to challenges derived from differences between the Apache and Nginx servers. This led to the necessity of exploring the time-consuming process of integrating Shibboleth with Nginx. Hence, a strategic decision was made to shift the DSpace 7 server configuration from Nginx to Apache.

#### **5. Results**

The original CEPT repository was built on DSpace software version 6.3. With the developments in technology, the DSpace community released the latest version of the software with the introduction of a new Angular-based User Interface that brings together the best of JSPUI and XMLUI. The new release version of DSpace 7 is backed by a brand-new REST API, which opens all data and features to the web, allowing it to integrate with external systems easily.

The CEPT library's restructuring and upgrading efforts resulted in several significant outcomes. Firstly, the user interface will be redesigned to provide a user-friendly experience, making it easier for users to navigate the repository and access scholarly content. Secondly, the search capabilities will be enhanced, allowing for more precise and efficient retrieval of information. Thirdly, metadata management will be streamlined, enabling better organization and discovery of resources within the repository.

Additionally, the library introduced new features and functionalities to enrich the repository's offerings. These included advanced browsing options, citation tracking, and integration with external scholarly platforms. The implementation will also improve the repository's performance, addressing issues related to system speed, stability, and scalability.

The integration of Shibboleth with Nginx or DSpace 7 with Apache is currently in progress. The CEPT repository will be upgraded to the new version once this integration is complete, whether it is with DSpace 7 and Apache or Shibboleth and Nginx. Subsequently, the upgraded repository with all its new features will be officially announced to users. Additionally, the CEPT library intends to retain the existing repository on the server in its current state for a minimum of one month after the switch to the new version. This time will be used to test the new repository version, ensure user satisfaction, and address any potential concerns. Once users are satisfied with the new DSpace 7 version and encounter no problems, the plan is to remove the old DSpace version and only use the repository with the DSpace 7 version.

## 6. Conclusion

This paper presents a comprehensive case study detailing the experience of restructuring and upgrading the CEPT University Library's DSpace-based Institutional Repository (IR). It starts by outlining the university's mission and the role of the library. The repository's origin during the pandemic, the challenges faced, and the decision to upgrade to DSpace 7 are discussed. The paper's objectives centre on sharing the repository's journey, addressing usability concerns, and implementing a better version. The assessment, planning, and technical phases are highlighted, demonstrating collaboration and the handling of challenges.

The updated repository has a new structure, better metadata management, and innovative features. The ongoing work involving Shibboleth integration and the upgrade strategy are emphasized. The conclusion focuses on the project's significant outcomes, such as enhanced user experience and resource management. This case study is an invaluable resource for academic institutions wishing to improve their repositories, bringing out the need for preparation, teamwork, and adaptation. The success of CEPT in this venture will serve as a model for optimizing digital resources and scholarly interaction.

## References

- Alexandra Carlile Butterfield, Q. G. (2022). Expanding your institutional repository: Librarians working with faculty. 48. doi:<https://doi.org/10.1016/j.acalib.2022.102628>
- Bimbe, N. L. (2017). Challenges in Reinvigorating and Upgrading DSpace-based Institutional Repositories: A University of Zambia (UNZA) Library Case Study. (483). Retrieved from <https://opendocs.ids.ac.uk/opendocs/handle/20.500.12413/12858>
- Samson Soong, U. L. (2019). How to Make Your OA Repository Work Really Well. Retrieved from [https://www.eifl.net/system/files/resources/201907/repositories\\_checklist\\_july\\_2019\\_\\_0.pdf](https://www.eifl.net/system/files/resources/201907/repositories_checklist_july_2019__0.pdf).
- Formanek, M. (2021). Solving SEO Issues in DSpace-based Digital Repositories A Case Study and Assessment of Worldwide Repositories. 40(1). doi:<https://doi.org/10.6017/ital.v40i1.12529>
- INFED. (2023). <https://infed.inflibnet.ac.in/index.php>. Retrieved from <https://infed.inflibnet.ac.in/index.php>.

<https://wiki.lyrasis.org/display/DSDOC7x/Upgrading+DSpace>. (2023). Retrieved June 2023, from <https://wiki.lyrasis.org/display/DSDOC7x>.

<https://www.dspace.com/en/pub/home/magazine.cfm>. (2023). Retrieved from <https://www.dspace.com/en/pub/home.cfm>.

Marios Zervas, A. K. (2019). Next generation Institutional Repositories: The case of the CUT Institutional Repository KTISIS. *Procedia Computer Science*, 146, 84-93. doi:10.1016/j.procs.2019.01.083

Mondal, D. (2021). Growth and Development of National Digital Library of India since y of India since.

Nason B. Bimbe, S. L. (2017). Challenges in Reinvigorating and Upgrading DSpace-based Institutional Repositories: A University of Zambia (UNZA) Library Case Study. *Institute of Development Studies*, 2017(483), 15. Retrieved from <http://dspace.unza.zm/handle/123456789/7709>

Tejaswini Joshi, N. D. (2022). *Journey from Digitization to Implementation of CEPT Thesis Repository*. Overseas Press India Private Limited (pp. 371-384). New Delhi: KIIT Deemed to be University. Retrieved May 2022