Design and Development of institutional Repository Using D-Space Digital Library Software at Agricultural Engineering College and Research Institute, Kumulur,TNAU-A Study

Dr.M.C.Subangi[1], Dr.P.Sujatha[2], Mr.M.Mohan[3]

[1]Deputy Librarian, Agricultural Engineering College & Research Institute, (TNAU), Kumulur, Tiruchirappalli,

[2]Associate Professor (Maths), Dr. M. S. Swaminathan Agricultural College and Research Institute, (TNAU), Eachangkottai, Thanjavur.

[3]Library Assistant, Agricultural Engineering College & Research Institute, (TNAU), Kumulur, Tiruchirappalli,

…………………….………..………..…..*************………….…..……………….……………

Abstract: The case study focuses on the design and development of an institutional repository at the Agricultural Engineering College and Research Institute, Kumulur, Tamil Nadu Agricultural University, using the DSpace digital library software. The current work discusses the University Research Project (URP) carried out at the Agricultural Engineering College and Research Institute, Kumulur, TNAU, India, which involved the establishment of an institutional repository. The initiative aims to expand the Institute's research output's exposure and accessibility, which will improve researcher collaboration and information sharing. The paper outlines the steps involved in the development process, including selecting the appropriate hardware and software, configuring the system, creating metadata schemas, and establishing access controls. The study also discusses the challenges encountered during the development process and the solutions that were implemented to overcome them. The study highlights the benefits of using DSpace to preserve and disseminate digital content, such as research articles and theses, and provides full-text search and retrieval capabilities. They also highlights the benefits of using DSpace for institutional repositories, including the ability to accommodate different types of digital content and support interoperability with other systems. Overall, this case study provides insights into the successful implementation of DSpace for institutional repositories and also highlights the problems faced during customization of metadata, interface and workflow as per the requirements of the institution. The study can serve as a useful reference for other institutions planning similar projects.

Keywords: DSpace, Institutional Repository, Open Access, Scholarly Communications, Digital Preservation, Digital Library.

1. INTRODUCTION

An institutional repository (IR) is a digital collection of an institution's intellectual output, including research articles, theses and dissertations, conference proceedings, technical reports, and other scholarly materials. The purpose of an institutional repository is to provide open access to these materials, making them available to the public and promoting their use and dissemination. IRs are typically managed and maintained by academic institutions, research organizations, or libraries, and can be used by researchers, students and faculty. Institutional repositories serve several purposes, including long-term preservation of an institution's intellectual output, making it accessible to future generations. IRs increase the visibility and impact of an institution's research output by providing open access to materials that might otherwise be hidden behind paywalls or other access barriers. IRs ensures compliance with many funding agencies and institutions which require researchers to make their work open access.

2. ABOUT THE AGRICULTURAL ENGINEERING COLLEGE AND RESEARCH INSTITUTE, KUMULUR, TRICHY,TNAU

The Agricultural Engineering College and Research Institute was established as a constituent educational institution of the prestigious Tamil Nadu Agricultural University in the academic year of 1972 at Coimbatore. The college was shifted to Kumulur in the year 1992. There are roughly 20 professors working, including teaching assistants in the core areas of Farm Machinery and power engineering, Soil and water conservation engineering, Basic engineering and Applied Sciences, Processing and Food engineering, Renewable energy engineering, and other general subjects Computer Science, Mathematical Science, etc..

The Institute publishes its research and various activities in a variety of Journals, Newsletters, Newspapers, Magazines, Book Chapters etc. So far the publications are archived and accessed through the Library. The present University Research project (URP) is implemented to unlock the potential of these published content in digital platforms.

3. SOFTWARE USED TO IMPLEMENT THE PROJECT (DSPACE)

DSpace is open-source digital repository software that allows organizations to preserve, store, manage, and disseminate digital content, such as research articles, theses, and datasets. It was initially developed by MIT Libraries and Hewlett-Packard (HP) Labs in 2002 and is currently maintained by DuraSpace, a not-for-profit organization.

DSpace is designed to accommodate various types of digital content, such as research papers, theses, datasets, audio and video files, and more. It allows for easy submission of materials, organization and management of content, as well as dissemination through a user-friendly interface. One of the key benefits of DSpace is its ability to preserve digital assets over time. It has a built-in preservation system that ensures the long-term storage and accessibility of digital content, even as file formats and technology change.

In addition to its preservation capabilities, DSpace also offers robust search and discovery functionalities, making it easy for users to find and access the content they need. It is highly customizable, allowing institutions to tailor the interface to their needs and branding. It also supports interoperability with other systems through various protocols and standards, such as OAI-PMH and SWORD. DSpace has been adopted by a wide range of organizations, including academic institutions, libraries, museums, and government agencies, to manage their digital content. It continues to be actively developed and maintained by a community of developers and users around the world.

4. GOALS OF THE STUDY

- An IR's primary goal is to empower the organisation or institution to create its digital content and make it available to users online. However, the following are the project's specific goals:
- To create a repository for institutional use at Agricultural Engineering College and Research Institute using DSpace digital library software for preserving and disseminating the institution's research output.
- To identify the hardware and software requirements needed for setting up the institutional repository using DSpace digital library software.
- To configure the DSpace digital library software to meet the needs of the Agricultural Engineering College and Research Institute.
- To create metadata schemas for the institutional repository to ensure consistency and interoperability with other systems.
- To establish access controls to ensure the security and integrity of the institutional repository.

- To evaluate the challenges encountered during the development process and propose solutions for overcoming them.
- To assess the benefits of the institutional repository, including increased visibility and access to research output, enhanced collaboration, and knowledge sharing among researchers.
- To provide a valuable resource for other institutions considering the implementation of their own institutional repositories using DSpace digital library software.

5. METHODOLOGY

The methodology of creating and maintaining the institutional repository typically involved several stages, including:

Planning: We defined the objectives of the repository, and identified the types of materials that will be included. This involved identifying stakeholders, assessing institutional needs, and establishing policies and guidelines for the repository.

Collection Development: We identified the materials that will be included in the repository, such as research articles, theses, dissertations, conference papers, technical reports, and other scholarly materials.

Scanning and digitization: Digitization is an essential process for creating and maintaining a digital repository. These processes involve the conversion of physical materials, such as books, papers, photographs, and other documents, into digital format.

Quality control: Quality control measures should be in place to ensure that the digital images are of high quality and meet the repository's standards.

Metadata Creation: The metadata provides information about the materials in the repository, such as author, title, abstract, keywords, and publication date. Creating high-quality metadata is essential for making the materials discoverable and accessible.

Submission and Ingestion: The materials to be included in the repository are submitted by repository staff, and then ingested into the repository. This involves converting materials to the appropriate digital format and attaching the metadata.

Access and Discovery: User-friendly interface that makes it easy for users to search, browse, and access the materials. This includes providing advanced search functionality, faceted browsing, and support for multiple languages.

Preservation: The materials in the repository must be preserved for the long-term to ensure their accessibility and usability. This involves creating backup copies, implementing a disaster recovery plan, and migrating materials to new formats as necessary.

6. LITERATURE REVIEW

The functions of DSpace were enumerated by Tansley et al. (2003) as a data model, metadata, e-people, authorisation, ingestion, workflow, CNRI Handle system, search and browsing, Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH), subscription, and Web user interface. Since its early success, DSpace has expanded into a global company. Additional system functions have been added to DSpace by its users based on those initial system functions. For instance, the Texas Digital Library team developed a Web 2.0-based interface for a map collection (Maslov, Mikeal, Weimer, and Leggett, 2009) and introduced Manakin for specialised user interfaces (Philips, Green, Maslov, Mikeal, and Leggett, 2007). They also added a customised workflow management system and Open Archives Initiative Object Reuse and Exchange (OAI-ORE).

In order to enable more effective search operations among DSpace members and their collections, semantics is one emerging growth area in DSpace functions (Kruk and McDaniel, 2009; Usman and Khan, 2012; Cherukodan, Kumar, and Kabir, 2013). In addition, Cherukodan, Kumar, and Kabir (2013) used Google Analytics to assess how digital objects were distributed and how people were using a DSpace-implemented academic DL.

7. REQUIREMENTS FOR STRUCTURAL LAYOUT

Basic IT infrastructure was needed for the research project, including computers, scanners, printers, digital library software, operating systems,

and other utility software. The list of such IT infrastructure that was purchased and utilized for the project is provided in this section.

7.1 Hardware prerequisites

The P-IV configuration of two Personal Computers is sufficient to carry out digitising operations. A separate, powerful server that is operational around-the-clock is needed for an institutional repository. For data backup, more disc space is needed. The graphic data should be able to be stored in enough RAM. There are multifunctional scanners available. The hardware and software in Tables 1 and 2 were utilised to construct the Digital Repository.

Table 1. Hardware requirements

Sl. No	Item	Configuration	Quantity
1.	Desktop server : Dell T30 Server	Intel XEON E3 1225 V5 3.3 ghz (4 core), 16GB DDR4 RAM (94 dim slot), 1TB HDD (4 bay), DVD, Raid 0&1 (software) DOS, 3 Years Warranty, 18.5 Monitor, Dell keyboard and mouse.	01
2.	Desktop Computer: Intel Pentium Dual Core G3250 (Lenovo) with Preloaded Windows Operating System.	Intel Pentium G3250 Processor, Dual core 3.2Ghz, Memory 4GB DDR3 @ 1600MHz upgradeable up to 16GB, 3 PCI/PCI Express Slots,1Gbps Network card, 4 USB Ports, 2PS 2 Port, 18.5 inch TFT /LCD TCO Certified Colour Monitor, 500 GB HDD, DVD Writer, Membrane Keyboard, Optical Mouse, 3 Bays with preloaded Windows 8 Professional with Media and License. Antivirus with 1 Year License.	02
3	Multifunctional Printer	Print,Copy & Scan, Up to 20ppm Black (normal,A4) Hi-Speed USB2.0,Built-in Wi-FI 802.11b/g/n.	01

7.2 Software Prerequisites

Several Open Source Software packages are currently available to construct and develop Institutional Repositories (IRs) in general and Digital Libraries in particular. One of the key breakthroughs that have contributed to the rapid expansion of IRs is the OSS. For example, Greenstone Digital Library Software (GSDL), EPrints, DSpace, and others are some of the OSS now accessible to IRs. After comparing the OSS's comparative features, it was ultimately chosen to use DSpace to create an institutional repository. These are the key characteristics of this software:

D-Space is an open-source digital repository software that offers a range of features to help organizations preserve, store, manage, and disseminate digital content. Some of the key features of DSpace include:

❖ Customizable user interface: DSpace allows organizations to create a customizable user interface to suit their needs and branding.

❖ Multiple language support: DSpace supports multiple languages, making it accessible to users worldwide.

❖ Metadata management: DSpace enables organizations to create and manage metadata for digital content, ensuring consistency and interoperability with other systems.

❖ Full-text search and retrieval: DSpace allows users to search and retrieve digital content by keyword, author, date, and other criteria.

❖ Access control: DSpace provides robust access control features to ensure the security and integrity of digital content.

❖ Interoperability: DSpace supports interoperability with other systems through various protocols and standards, such as OAI-PMH and SWORD.

❖ Versatile content types: DSpace is a flexible platform for various types of digital collections since it can handle numerous types of digital content, including text, photos, audio, and video.

❖ Preservation support: DSpace provides support for long-term preservation of digital content through features such as file format migration and metadata preservation.

❖ Reporting and analytics: DSpace provides reporting and analytics features, allowing organizations to track usage, downloads, and other metrics.

Overall, DSpace is a comprehensive digital repository software that offers a range of features to help organizations manage and make their digital content accessible to a wider audience.

Table 2. Software Prerequisites

Sl.No	Item	Configuration
1	Name	DSpace
2	Technical Aspects	Written in Java, tested under Linux, Windows/ Mac OSX
3	Database	PostgreSQL / ORACLE
3	License:	BSD
4	Release version	6.0, https://duraspace.org/dspace/download/
5	Documentation	https://wiki.duraspace.org/display/DSDOC/
6	Other prerequisite software	Java 7 or 8, Apache Maven, Apache Ant, Relational Database – PostgreSQL or Oracle, Servlet 3.0 container (Tomcat 7+ or similar).

7.3 Manpower Assistance

A crucial challenge and one of the most essential conditions for constructing an institutional repository is the availability of technically competent labour to assist such projects. Although extensive IT knowledge is not required, the main requirements are that they have the essential set of skills for DSpace-related tasks such as software installation, document scanning and digitalization, file uploading, and other tasks.
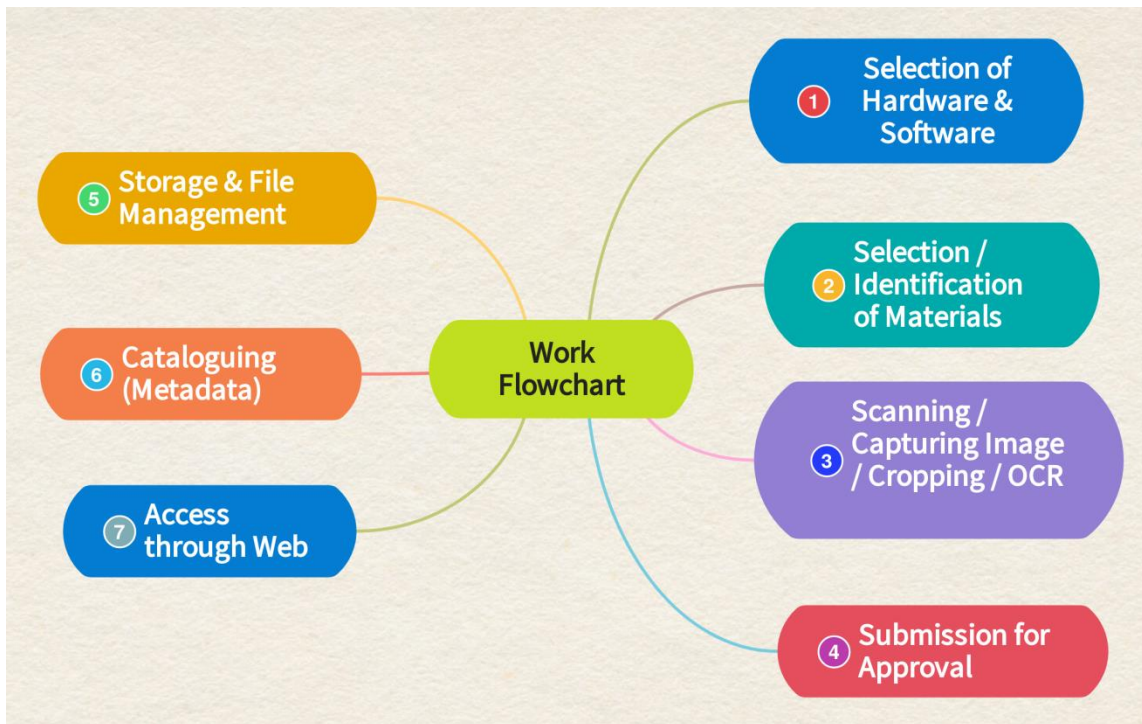
Figure 1. Work flowchart

8. DOCUMENTS IN THE DIGITAL REPOSITORY

The following trend can be seen in the bibliographic formats, for example articles in journals, conference proceedings, books, and magazines, according to the examination of all 2556 entries. The distribution of publications among the four specified bibliographic forms has been tallied and reported in Table 3 of the data

.

Table 3. Types of documents

Document Type	No. of Records	Cumulative Records	%	Cumulative %
Articles in Journals	842	842	32.94	32.94
Articles in Books	358	1200	14.01	46.95
Articles in News Paper	542	1742	21.21	68.15
Articles in Thesis/Project	623	2365	24.37	92.53
Articles in Newsletter	120	2485	4.69	97.22
Articles in Question Papers	71	2556	2.78	100
Total	**2556**		**100.00**	

9.   DISEMBRANGE OF DOCUMENTS IN THE DIGITAL REPOSITORY BY YEAR

Data has been gathered from all of the university's faculties and students and uploaded to the digital repository in order to determine the research output of the university by year. The research articles which are put down in digital repository have been analyzed year-wise and presented in Table 4. Examining the records' publication dates is also crucial. For the purpose of this study, a total of 2556 contributions from the years 1971 to 2022 have been documented.

Table 4 shows the production of the articles under examination by year. According to the aforementioned figures, between 1971 and 2022, the number of publications written by University academic members and students increased steadily. Although there have been 2556 articles published since 1971, this analysis shows that 31.34% of the overall research output—or 17.06 and 14.28 percent—were published in the years 2021–2022, 2019–2020, and correspondingly. Additionally, it has been noted that just 42.68% of research publications published between 1971-1972 and 2011-2012 were written by teachers and students. 57.32% of research articles, or a sizable contribution, were released from 2013–2014 through 2021–2022.

Table 4. Disembrange of Documents in the Digital Repository by Year

Year	No. of Records	Cumulative Records	%	Cumulative %
1971-1972	6	6	0.23	0.23
1973-1974	2	8	0.08	0.31
1975-1976	1	9	0.04	0.35
1977-1978	4	13	0.16	0.51
1979-1980	3	16	0.12	0.63
1981-1982	1	17	0.04	0.67
1983-1984	13	30	0.51	1.17
1985-1986	52	82	2.03	3.21
1987-1988	32	114	1.25	4.46
1989-1990	69	183	2.70	7.16
1991-1992	63	246	2.46	9.62
1993-1994	29	275	1.13	10.76
1995-1996	18	293	0.70	11.46
1997-1998	45	338	1.76	13.22
1999-2000	60	398	2.35	15.57
2001-2002	83	481	3.25	18.82
2003-2004	124	605	4.85	23.67
2005-2006	96	701	3.76	27.43
2007-2008	182	883	7.12	34.55
2009-2010	96	979	3.76	38.30
2011-2012	112	1091	4.38	42.68

2013-2014	210	1301	8.22	50.90
2015-2016	230	1531	9.00	59.90
2017-2018	224	1755	8.76	68.66
2019-2020	365	2120	14.28	82.94
2021-2022	436	2556	17.06	100.00
Total	**2556**		**100.00**	

10. HOW TO INSTALL A DIGITAL REPOSITORY USING DSPACE: PROCEDURE

10.1 Installation of DSpace

DSpace has been installed in one of the computers for the project. DSpace requires Java 8 or higher, Apache Maven, and a database management system such as PostgreSQL or Oracle as prerequisite software before installing DSpace. Latest stable version of DSpace was downloaded from the official DSpace website and installed successfully. New database was created and configured to work with DSpace.

Overall, installing DSpace requires some technical expertise, but there are many resources available, including documentation, forums, and user groups.

10.2 Digitization

Digitization includes acquiring, converting, storing and retaining information in standardized and organized manner with technological support. The documents are being scanned in the resolution of 300 to 600 dpi (dot per inch) depending on the physical condition of the documents. The next process, which is an important task, involves rectifying the errors and making text files error free with the help of OCR process. Scanned images are then being stored in the computer in PDF (for text) and JPEG (images) formats. These files need to be processed in order to be accessible via a network. Sometimes, JPEG and PDF file formats must be compressed.

All faculty members and students at the Agricultural Engineering College and Research Institute, Kumulur, TNAU, contributed a total of 2556 research articles. The identical documents have been photographed, converted to PDF format, and later uploaded to the Dspace software, one of the most popular institutional repository development open source programmes.

10.3 Submission of Document in DSpace

Any document submission must go through a seven-step processing process. Fundamental descriptions of the contents and Dublin Core metadata components resembling author, title, publisher, etc. are included in the first three steps. The fourth stage entails uploading the previously saved digital documents, such as PDF and image files, from the computers. The fifth step involves checking the data that was submitted in the prior steps and making any necessary adjustments. The license agreement is presented in the sixth phase, and the submission procedure is confirmed in the last step. After completion, the submitted item goes through additional formalities like review, editing, or approval in accordance with the established policies. DSpace also has a remote publishing feature that, when activated, allows authorised individuals to submit their works straight from their desktops.

The submission workflow in DSpace can be personalized to suit the requirements of individual repositories. Some repositories may have additional steps, such as content editing or peer review, while others may have fewer steps. The key objective of the workflow is to ensure that the deposited content meets the repository's policies and guidelines and is easily accessible to users.

10.4 Home Page of the AEC&RI Kumulur, TNAU Digital Repository

After installing the Dspace programme, the AEC&RI Kumulur Trichy Digital Repository homepage has been created using the customization feature available in the Dspace software. In Figure 2.
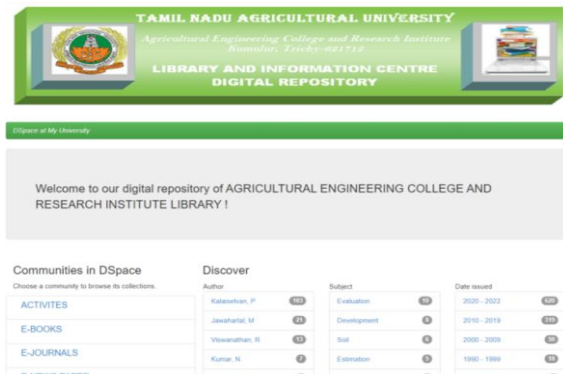

Figure 2. Home page of the Agricultural Engineering College and Research Institute Digital Repository

10.5 Searching and Browsing

Another crucial stage in the digitization of information resources is searching and browsing.

DSpace offers various search and browsing features to help users find the digital content they are looking for. Here are some of the main features for searching and browsing:

Keyword search: Users can search for content by entering keywords in the search box, and DSpace will return results that match the search terms. DSpace also supports advanced search queries using Boolean operators, fielded search, and facets. Researchers attempted to use the Communities and Collections page to search and view the data that had already been uploaded.

Browse by collections: DSpace allows users to browse the repository by collections, which are organized by topic, department, or other criteria. Users can select a collection to view all the items in that collection.

Browse by communities: DSpace also allows users to browse the repository by communities, which are groups of collections that are organized by department, research group, or other criteria. The screen shot of the same is as given in Figure3.
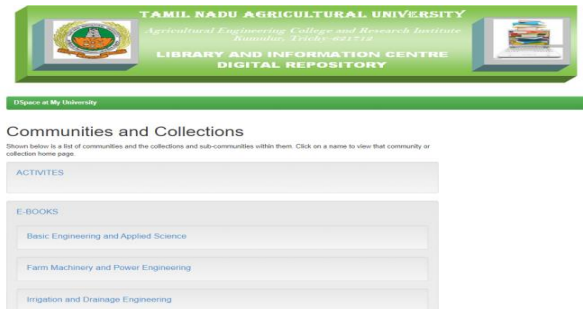

Figure 3. Searching and browsing (communities and collections)

Metadata search: DSpace enables users to search for content based on specific metadata fields, such as author, title, date, and subject.

Browse by Author: Users typically search for literature by title, author, theme, etc. Figure 4 shows a screenshot of this page's author name browsing section.
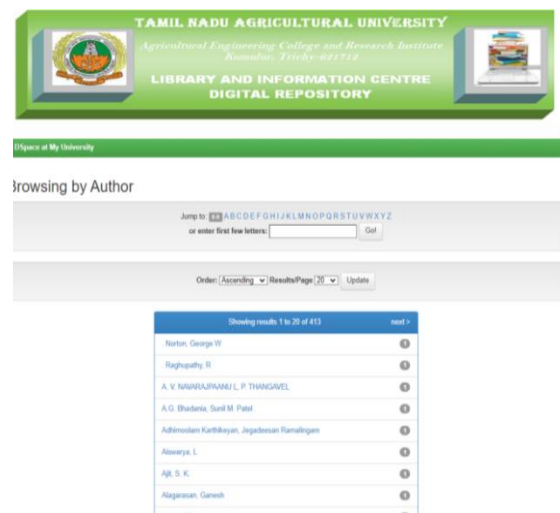

Figure 4. Browsing by author

Browse by Title: Another crucial component of access for users is the title. In light of this, the title index is also produced to make searching by title easier. Additionally, clicking on the selected title will bring up the article's metadata. Figures 5 and 6 show a screenshot of the title and metadata browsing for the selected item.

Figure 5. Browsing by title

Figure 6. Metadata of the chosen document

Browsing by Subject: The user's most popular search choice is the subject. Because of this, users of this digital library can also browse articles by subject while performing a search. The identical is shown in Figure 7.

Figure 7. Browsing by subject

11. VIEW AND OPEN THE ARTICLE'S FULL TEXT

Users will receive the Metadata of chosen articles after reading and searching the literature by author, title, subject, etc., however this information is frequently insufficient. Users want to read the articles in their entirety. In these situations, the user must select View to access the option that is shown on the author, title, and subject search page. Figure 8 shows a screenshot of the display and opening of the full-text article.

Figure 8. View and open of full text article

12. CONCLUSION / PROJECT OUTCOME

Finally, it can be said that the implementation of an institutional repository using the DSpace digital library software at the Agricultural Engineering College and Research Institute, Kumulur, Tamil Nadu Agricultural University, has been successfully completed. The project aimed to

create a digital repository to preserve and disseminate the research output of the institute, facilitate access to research materials, and promote collaboration among researchers.

The DSpace software was chosen due to its flexibility, scalability, and robust features. The implementation process involved several stages, including software installation, customization, content creation, and user testing. The project team also developed a submission workflow and metadata schema to ensure the quality of the deposited content.

The institutional repository has been launched and is now accessible to the institute's stakeholders, including researchers, faculty and students. The repository contains a growing collection of research output, including articles, Newsletters, reports, and conference proceedings, among others. The repository is also integrated with the institute's website, making it easy for users to access the repository's content.

The DSpace deployment at AEC & RI, Kumulur, federation plans, and sustainability concerns. The scope of Institutional Repositories includes separate servers, computers and other accessories. The Faculty and staff need to be trained for uploading of documents and submitting metadata into the repository.

References

[1] Dilek Karahoca, Ibrahim Dulda, Adem Karahoca, Ahmet Yijcel, Batuhan Gulluoglu, Evren Arifoglu 2010, "Interactive e-content development for vocational and technical education" , Procedia-Social and Behavioral Sciences, ELSEVIER Vol 2. Issue 2, pages 5842-5949

[2] Pilar Orero, Irene Tor-Carroggio, "User Requirements When Designing Learning e-Content: Interaction for All", Part of the Human–Computer Interaction book series.

[3] Xiaoquig Gu, Bian Wu & Xiaojuan Xu, 2015 "Design, development, and learning in e-Textbooks: what we learned and where we are going", Journal of computers in education Vol 2, pages. 25-41.

[4] Meysun Hamdi and Thair Hamtini, 2017 "Designing an Effective e-Content Development Framework for the Enhancement of Learning Programming", International Journal of Emerging Technologies in Learning Vol 1, No. 04

[5] J. Muthukumari, N.Ramakrishnan, 2017, "Development and validation of E-content on history subject(The beginning of modern age) of IX standard students", International Journal of research Granthaalayah, Vol. 5, Issue 9

[6] Baldev D.Joshi, R.K. Parikh, 2017 "E-content development: Prospects and challenges", IJARIIE, Vol-2, Issue – 1