Chapter-6

Dspace: Metadata Encoding and Transmission Standards
Chapter – 6

DSpace: Metadata Encoding and Transmission Standards

The preceding chapters have elaborated Open Access Repositories (chapter 4) as well as an overview of Open Source Software Packages (chapter 5). Chapter 6 has slightly dealt with open source software packages including DSpace, which is one of the most popular software packages that have been selected as a model for Open Access Repository in AMU.

6.1 Introduction

DSpace is an open source digital research repository system released in 2002. It is a joint venture of Massachusetts Institute of Technology (MIT) and Hewlett-Packard (HP) Labs, designed to deal with the problem of growing output of scholarly works in an institution. DSpace software is by and large preferred by academic, non-profit, and commercial organizations for building open digital repositories.

Lewis and Yates\(^1\) (2008) define it as a “platform that allows you to capture items in any format – in text, video, audio, and data. It distributes it over the web. It indexes your work, so users can search and retrieve your items. It preserves your digital work over the long term”. According to Pirounakis and Nikolaidou\(^2\) (n.d.) “Consider a case where an institution or university needs a digital repository for research papers and dissertations produced by students and staff. In that case, the most appropriate DL system is DSpace, since it by default represents communities (e.g. university departments) and collections (e.g. papers and dissertations), while workflow management supported is important for item submission by individuals”.

---


DSpace is an ideal tool for institutional repository collections that include resources such as publications, datasets, theses and dissertations, presentations, book, chapters and any material of scholarly importance. Some distinguishing features of DSpace are summarized as under:

- DSpace is an open source technology platform which can be easily customized and its capabilities can be extended.

- Unlike other popular software packages, DSpace provides a good Workflow Management, generates Authority Files and show strength of each collection on website.

- DSpace assists to capture and ingest of materials in any digital medium as well as metadata about the materials.

- DSpace has a well-designed Authentication and Authorization mechanism.

- DSpace has a very good concept of user interface e.g. CSS and Manakin templates. It is worthy to point out here that CSS or Cascading Style Sheets is a style sheet language used for describing the look and formatting of a document written in a mark-up language. It's most common application is to style web pages written in HTML and XHTML language. It can also be applied to any kind of XML document (Ovo international, 2007). Manakin (XMLUI) is a web-based user interface to DSpace that introduces a modular interface layer, enabling an institution to easily customize the interface according to the specific needs of the particular repository, community or collection. (SLCC, n.d.)

- DSpace is considered to be the best suited and entrusted solution for the long term preservation of open access repositories. It also supports backup, refreshing of media and disaster recovery facilities.

- DSpace also sustains persistent identifier i.e. Unique Accession number to each document acquired by the repository.

- DSpace facilitates easy access to the materials, both by listing and searching.
6.2 Brief history

The DSpace project was initiated in July 2000 as part of the Massachusetts Institute of Technology (MIT) and Hewlett-Packard (HP) Labs coalition. The first version of DSpace, openly available to public, was released in November 2002. This version was called DSpace 1.0. As a result of first user group meeting, held in March 2004, DSpace federation was formed by a group of interested institutions. In July 2007 HP & MIT jointly formed the DSpace foundation, a non-profit organization to provide support to the growing community of institutions that use DSpace. The foundation's mission is to lead the collaborative development of open source software to enable permanent access to digital works. (DuraSpace, 2000).

Following versions of DSpace released in reverse chronological order till date are:

- DSpace version 4.0-16th December 2013
- DSpace version 3.2-24th July 2013
- DSpace version 3.1- January 2013
- DSpace version 3.0- 30th November 2012
- DSpace version 1.8- 4th November 2011
- DSpace version 1.7- 17th December 2010
- DSpace version 1.6- 2nd March 2010
- DSpace version 1.5- 25th March 2008
- DSpace version 1.4- 26th July 2006
- DSpace version 1.3- 3rd August 2005
- DSpace version 1.2- 13th August 2004
- DSpace version 1.1- 8th May 2003
- DSpace version 1.0- 8th November 2002

6.3 Technology pertaining to DSpace

While choosing a software package to establish an Institutional Repository, some technological aspects are required to consider the internal structure and set-
up of the Software. Few aspects that are useful for the present study are given below:

6.3.1 Metadata

Metadata is a set of data for every digital object stored in a DL system. It describes the core of that digital object that includes access rights, administrative data, and hierarchical data that helps to signify, how that digital object is related to others in a DL system. This supplementary, illustrative information is known as metadata. In other word metadata is descriptive information about the real data (such as PDF files, word documents, images etc.) stored in DL system. In a DL system three types of Metadata that usually come across are given as under: (Gorton, 2007)

6.3.1.1 Descriptive metadata: Its function is to define the content of the digital object that includes author of the content, title, creation date etc.

6.3.1.2 Administrative metadata: It assists the organization of digital collection with regard to access rights, preservation information, and more technical information about a digital object that includes copyright or publisher information.

6.3.1.3 Structural metadata: it depicts the physical and/or logical structure of digital objects. It expresses the logical boundaries of complex digital resources, also used to explain relationships between an object's constituent parts. Structural metadata is frequently used to make possible the navigation and arrangement of complex digital resources by defining structural characteristics such as pagination, sequence etc. This type of metadata helps collection organization as well as inter-document structural understanding. (University of Illinois, 2012)

Metadata standards play a vital role in the creation of Institutional repositories or digital library system. Some important metadata standards are summarized as below.

6.3.2 Dublin Core (DC)

Dublin core metadata schema is a consequence of an invitational workshop funded by the Online Library Center (OCLC), held at Dublin, Ohio in 1995. Hence
the metadata schema named so; the ‘Dublin’ and the term refers to the location of
the workshop and ‘core’ describes the precision that DC is a collection of
metadata elements that are basic, but flexible. The organization named Dublin
Core Metadata Initiative (DCMI) was made responsible for the growth and
development of interoperable metadata standards. (Kurtz, 2010)

Dublin Core deals with mostly descriptive metadata, but also has some
elements that can have other applications as well. DC draws upon the ideas from
other disciplines too such as Librarianship, Computer Science, and Archival
Preservation (Gorton, 2007). DC contains around fifteen main elements, and seven
additional elements. All of them are optional and repeatable. Two forms of Dublin
core metadata are:

(a) Simple Dublin Core: It provides only basic information, using just the base
metadata elements from the Dublin Core Metadata Element Set.

(b) Qualified Dublin core: It is an extension of Simple Dublin Core developed
through the use of additional elements, element refinements, and encoding
schemes (SAA, 2013). DSpace by default consists of Qualified Dublin core
metadata. Only three fields are required: title, language, and submission date, all
other fields are optional. There are additional fields for document abstracts,
keywords, and technical metadata and rights metadata, among others. (Thakuria,
2008)

6.3.3 Metadata Encoding and Transmission Standards (METS)

METS is another metadata standard that covers all the three types of
metadata including descriptive, administrative, and structural metadata. The
standard is maintained in the Network Development and MARC Standards
Office of the Library of Congress, and is being developed as an initiative of the
Digital Library Federation. METS is exclusively designed for demonstration of
textual and image data in XML Schema. METS documents have different
elements and sections that relate to practically all facets of a digital object and, in
fact, can be employed to represent entire objects including referencing external
metadata in standardized formats. There are around seven major sections in a METS document. (Gorton, 2007)

6.3.4 Extensible Markup Language (XML) and XML Schema

Extensible Markup Language or XML is a markup language used to describe data. Like other markup languages, e.g., HTML, XML describes how other information should be perceived. XML through its hierarchical nature allows the explanation of almost all information in a succinct and logical fashion. It is a constructive, entirely user customizable way to represent information. It is paired with other XML instances to make its technology more powerful. XML schema is one of them, that provides a technique of stating a structure of data within the perspective of XML without specifying the data itself. XML and XML schema are vitally important in numerous aspects of the Internet, web services, and other web based technologies. XML extensively supports interoperability, providing a means of transferring data from one version to another. (Gorton, 2007)

6.3.5 Web services

Web services are normally built upon XML, which has become the de facto standard in information representation. In web services, small pieces of information programmed in unique, definite ways that are broadcasted across the web as a mode of information exchange. In such a methodology, machines send such data, and others listen for incoming requests and provide meaningful responses as required. SOAP (Simple Object Access Protocol) and REST (Representational State Transfer services) are the common forms of web services. Apache Axis, which runs on Apache Tomcat Java servlet container, is the most common research oriented web service platform. In the framework of digital libraries/institutional repositories, web services play an essential role in the communication of dispersed, componentized DL systems. (Gorton, 2007)
Chapter- 6

DSpace: Metadata Encoding

6.4 DSpace System Architecture

DSpace software system consists of three layers that form its three-tiered architecture. The DSpace Foundation described these three layers as below:

6.4.1 Storage layer: It is the lowest layer, responsible for the physical storage of metadata and content. It is made up of Bitstreams and Relational Database management system (RDBMS). All the Bitstreams are stored in this layer as files and location of these files and other metadata information are stored in a Relational Database Management System (RDBMS), in this case PostgreSQL.

6.4.2 Business Logic Layer: It is in the middle of the system. It consists of set of modules that illustrates the inner working of many DSpace objects type such as browsing and searching in DSpace, User related functions, Content management and other works. It also controls DSpace’s public API that allows third party code to interact with DSpace.

6.4.3 Application Layer: It is the top most layers. It controls the backend functionalities of the system, responsible for the provision of services and functionalities at the user end such as display. The Application layer also includes import/export functionality, statistics tools, web based user interface.

![Figure 6.1: DSpace System Architecture](image)

---


173
Chapter 6  

DSpace: Metadata Encoding

6.5 DSpace Directories

The DSpace foundation defines three directories of the DSpace system that are as below:

6.5.1 Source Directory (/DSpace/DSpace-3.1-source/)

All customization generally takes place here:

- config/ - DSpace configurations
- jsp/ - DSpace JSPs
- src/ - DSpace Servlets & Classes

6.5.2 Installed Directory (/DSpace/

This is the place where DSpace is installed and all the configuration files, command line scripts, documentation and webapps of DSpace will be installed to.

6.5.3 Web-application Directory (/DSpace/apache-tomcat-6.0/webapps/)

- It is home to all compiled servlets and JSPs. In other words it is a directory that controls web applications of DSpace software.
- Basic customizations can also take place here (if one decide to do after installation)

It is to be noted here that while using Tomcat server, one has to copy DSpace web applications from [DSpace]/webapps/ to [tomcat]/webapps/

6.6 DSpace Data Model

The digital materials or objects, stored in Dspace are referred as 'items'. They are arranged in a hierarchical order in such a way that related items are grouped and submitted into a collection of related content. The Community is the highest level in the Hierarchy. One can divide Communities and sub-communities into infinite numbers and also create an infinite number of collections into these communities and sub-communities but Collection cannot

---


174
be sub-divided further. Each Item accumulated in a DSpace repository is made up of a bundle of bitstreams, so as many files can be stored in a single digital object as needed. Bass et al. (2002) described a few things that should be clear regarding object model while working on DSpace, as follows:

6.6.1 Community: A Community in DSpace is the entry point into the magnitude of material in the repository. DSpace community corresponds to an organizational entity such as a school, department, laboratory, or research center. Community is followed by sub-community e.g. school of Humanities, Department of Social Sciences, and Science laboratory.

6.6.2 Collection: A DSpace collection groups together a set of items that are related in some dimensions (e.g. Source, Purpose, Existing series, Subject matter, and Research topic).

6.6.3 Items: An item in DSpace is a combination of useful sets of content and metadata that are related in some way. Items correspond to “Archival Atoms” in DSpace. The item includes a working paper, journal articles, presentation, course lecture material, a video clip etc.

6.6.4 Bundle: A group of HTML and image bitstreams making up an HTML document

6.6.5 Bitstream: A bitstream in DSpace is simply a reproducible sequence of bits, with a corresponding bitstream format. Bitstreams typically correspond to content or metadata files that are submitted to DSpace.

6.6.6 Bitstream formats: In DSpace, a bitstream format is a unique and consistent way to refer to a particular file format. DSpace accepts many file formats like PDF, MS-WORD, Postscript, mpeg, jpeg, tiff, gif etc.
6.7 E-People/Groups in DSpace

There are certain E-people or Groups in DSpace who assigned certain roles within the DSpace system by the system Administrator.

6.7.1 Unspecified Users (anybody): These are the end users and are not assigned by the DSpace administrator.

6.7.2 Members: These are the E-persons who wish to subscribe to a collection (one cannot subscribe to communities).

6.7.3 Collection Administrators: In a large digital repository various E-groups are assigned as collection administrators. They can choose the reviewers, metadata editors among members and decide the collection policy. They are different from DSpace administrators, who have the overall responsibility and power, a kind of super user.

6.7.4 Submitters (authors): Those who submit their publications to a collection (they should be members of the DSpace system and have been authorized to submit).
6.7.5 **Reviewers:** These are the authorized members to review submissions. They can either accept or reject submissions. Usually, they are subject specialists.

6.7.6 **Metadata Editors:** are those who certify the metadata. Normally, they are library professionals.

6.8 **DSpace Ingest Process**

DSpace is a suite of software system that stores digital objects in it. It is a turnkey to institutional repositories. Ehtesham (2008) enumerates two main ways in which data enters into the system, as follows:

- The registered users of DSpace, submit items to collections through a web-based User Interface (UI).
- The DSpace Administrator who has a large amount of content to be batch imported may take advantage of the import/export functionalities of the system.

The batch item importer controls the submission of content and allows users to submit content into the software system.

![Ingest process of DSpace](image)

**Figure 6.3: Ingest process of DSpace**

6.9 **DSpace Submission Process**

DSpace is one of the first open source repository systems that effectively combat the troubles that occur during the submission of different types of information to different collections (Smith et. al., 2003). The DSpace submission workflow system is a critical part of the DSpace architecture that
allows for the submission, processing, and final addition of content to the live repository. DSpace’s underlying model includes E-People, users who have registered with the system and have certain authorizations, roles, rights, and privileges that translate the abilities to complete certain tasks within the DSpace system. Gorton (2007) explains the following submission steps that are outlined in the table as below:

**Table- 6.1**

Submission Process in DSpace

<table>
<thead>
<tr>
<th>S.No.</th>
<th>Submission steps</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Describe</td>
<td>Here the metadata about the document by submitted to the system is entered by the user. This metadata includes Author, Title, Subject, Abstract, short description etc.</td>
</tr>
<tr>
<td>2</td>
<td>Upload</td>
<td>The files are uploaded, resided in the local machine related to the metadata submitted previously.</td>
</tr>
<tr>
<td>3</td>
<td>Verify</td>
<td>An outline of all the details of the submission is given including a summary of the entered metadata and the files concerned with the submission.</td>
</tr>
<tr>
<td>4</td>
<td>License</td>
<td>The user is shown and must agree to the license that the system’s administrator has assigned to the submitted contents for his collection.</td>
</tr>
<tr>
<td>5</td>
<td>Complete</td>
<td>The submission process is completed and the item may immediately be added to the collection or undergoes workflow process, set for the collection by the system administrator before final submission.</td>
</tr>
</tbody>
</table>

6.9.1 Workflow Process in DSpace

There are certain workflow steps assigned by the DSpace administrator for the different collections to the different e-peoples or Group of e-peoples. These e-persons can be reviewers, metadata editors, and collection administrators. These steps are shown through the diagram as under:
Tansely (2003) has described the workflow steps in DSpace as below:

Table- 6.2

**Workflow Steps in DSpace Software**

<table>
<thead>
<tr>
<th>Workflow Steps</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step 1</td>
<td>The assigned E-person or group can accept or reject the submitted item for inclusion.</td>
</tr>
<tr>
<td>Sep 2</td>
<td>The assigned e-person or group can edit metadata provided by the user with the submission, but cannot change the submitted files. Also can accept submission for inclusion, or</td>
</tr>
</tbody>
</table>
Steps

Step 3

The assigned e-person can edit metadata provided by the user with the submission, but cannot change the submitted files. Then commit the submission to archive, may not reject the submission.

“When a submission reaches a particular workflow step, members of a Group set up exclusively to deal with that workflow step in that collection, are alerted using email. Upon login into their DSpace account, they will find a new item in their Task Pool. This is visible to everyone within that workflow group, and anyone may take the task to do themselves. When they do this, the task is removed from the Task Pool and placed into the Owned Tasks section, which is only visible to that particular user. The user may then perform the workflow requirements and take whatever action is required, or the task may be returned to the Task Pool”\(^1\).

6.10 Salient features of DSpace

DSpace is a landmark open source digital Library software that captures, stores, indexes, preserves and redistributes the intellectual output of an institution. DSpace is flexible to different community needs as it can be customized or extend its capabilities. There are a number of reasons to opt DSpace software to establish institutional repositories and make available their scholarly digital content searchable and retrievable on the web. DSpace has various extensible features. Some of the features are enumerated below.

6.10.1 Content Support

DSpace accepts all modes of digital content and format as well. The items include articles, preprints, post-prints, chapters of an edited book, conference papers, technical reports, Data sets, Computer programs, images, multimedia

---

publications, Audio files, Video files, learning objects etc. It can support many types of file formats such as text, images, audio, and video.

6.10.2 Search Functionality

Search is an essential component of discovery in DSpace. Users' expectations from Web search engines are quite high, so a goal for DSpace is to supply as many search features as possible. DSpace's indexing and search module has a very simple API which allows for indexing new content, regenerating the index, and performing searches on the entire corpus, a community, or collection (Tansely, 2003). DSpace uses Lucene search engine for its search functionalities. Lucene is a part of Apache Jakarta project. It supports following type of searches

6.10.2.1 Field search: DSpace provides field search such as by Author, title, Subject, year or date.

6.10.2.2 Exact term: The search term can be a single word or whole phrase, e.g. Cataloguing or Cataloguing tools.

6.10.2.3 Wild card: The Two types of searches can be done through wild card search that are:

Single character search by using symbol ‘?’ as in ‘Re?t’ that matches words like ‘Rent’, ‘Rest’ etc.

Multiple character by using symbol ‘*’ as in “Cla*” matches with Classification, Classifier etc.

6.10.2.4 Fuzzy search: A fuzzy searching is a process that retrieves a word even if it is misspelled. A fuzzy search is made by means of a fuzzy matching program, which retrieves a list of results based on probable relevance even though search query words and spellings may not closely match. Fuzzy search is based on the ‘Levenshtein’ distance algorithm, devised by Russian scientist, named Vladimir Levenshtein in 1965. The algorithm is also known as ‘Edit
Distance algorithm' (Prasad & Patel, 2005). The Levenshtein distance algorithm\(^1\) has been used in a number of ways such as:

- Spell checking
- Speech recognition
- DNA analysis
- Plagiarism detection etc.

In DSpace search example of fuzzy search is: Syam can match Shyam or Sankar can match Shankar.

6.10.2.5 Range Search: A range search is one that returns all values between two precise values e.g. There are two types of range searches include Inclusive ranges that return any values that match the two specified values. Exclusive ranges that do not return any values that match the two specified values.

- If the search query is Year: it can be ranged documents published [2005 to 2013], the documents retrieved by the search engine that have been published in between the year ‘2005’ and ‘2013’. On the contrary the query {2005 to 2013} excludes the documents published in the year ‘2005’ and ‘2013’
- Or if the search query is author: [gupta to prasad] Then the system retrieves documents authored by names that fall between ‘gupta’ and ‘prasad’. Whereas, the query ‘author’: {gupta to prasad} excludes gupta and Prasad

6.10.2.6 Proximity search: Lucene supports finding words that are within a definite distance away. To do proximity search use the symbol ‘~’ at the end of a Phrase. For example to search for “Information” and "Retrieval" within 10 words of each other in a document use the search:

"Information Retrieval" ~10

---

6.10.2.7 **Boosting a term:** To boost a term use the caret, "^", symbol with a boost factor (a number) at the end of the term is being searched. The higher the boost factor, the more relevant the term will be. For example, search for

Classification Cataloguing

And if the term "Classification" need to be more relevant, boost it using the `^` symbol along with the boost factor next to the term as below

Classification^4 Cataloguing

This will make documents with the term ‘Classification’ appear more relevant.

By default, the boost factor is 1. Although the boost factor must be positive, it can be less than 1 (e.g. 0.2).

6.10.2.8 **Boolean search:** Boolean ‘AND’, ‘OR’, ‘NOT’ are used for Boolean combinations. Boolean operators should be caps. Alternative symbols can be used for each operator such as

- ‘OR’ is the default conjunction operator. One can use ‘||’ instead of ‘OR’
- For Boolean ‘AND’ the symbol ‘&’ can be used
- For Boolean ‘NOT’ the symbol ‘!’ can be used

6.10.3 **Handle system**

DSpace’s core feature is the creation of persistent identifier for every item, collection and community stored in DSpace which in return provides the benefit of long term preservation. To persevered identifiers, DSpace requires a storage- and location- independent mechanism for creating and maintaining identifiers. DSpace uses the Handle System from CNRI (Corporation for National Research Initiatives) to assign and resolve persistent identifier for each digital object. Handles are URN-compliant identifiers, and the Handle resolver is an open-source system which is used in conjunction with the DSpace system.

6.10.4 **Preservation**

DSpace ensures long term preservation as it uses CNRI’s handle system. DSpace uses Handles primarily as a means of assigning globally unique identifiers to objects. Each site running DSpace needs to obtain a Handle 'prefix'
from CNRI, a guarantee, that if identifiers are created with that prefix; they will not clash with identifiers created elsewhere. Smith et. al. (2003) categorizes two levels of digital preservation in DSpace as below:

- **Bit preservation**: It ensures that a file remains exactly the same over time – not a single bit is changed – while the physical media evolve around it.

- **Functional preservation**: It goes further the file does change over time so that the material continues to be immediately usable in the same way it was originally, while the digital formats (and the physical media) evolve over time. DSpace allows one, to choose three levels of preservation for a given format are as under:

  - **Supported** formats are those which one can functionally preserve using either format migration or emulation techniques such as TIFF, SGML, XML, AIFF, and PDF.

  - **Known** formats are those that one cannot assure to preserve, such as proprietary or binary formats, but the third party migration tools may help to format migration. Examples include Microsoft Word and PowerPoint, Lotus 1-2-3, and WordPerfect.

  - **Unsupported** formats are those that one does not have adequate knowledge, to do any sort of functional preservation. This may include some proprietary formats or a one-of-a-kind software program.

### 6.10.5 DSpace SWORD Interface

DSpace emerges with its own SWORD (Simple Web-service Offering Repository Deposit) Server (the 'sword' webapp), which allows any SWORD client to submit documents to DSpace via e-mail. SWORD is a lightweight protocol for depositing content from one location to another. The SWORD visualization is ‘lowering the barriers to deposit’, chiefly for depositing content (any content) into repositories, but potentially, for depositing into any system which wants to receive content from remote sources.
6.10.6 Checksum Checker

This tool can be programmed to perform a full fixity (checksum) check of all (or some) content files stored in DSpace instance.

6.10.7 Open URL support

DSpace supports the OpenURL protocol from SFX server, in a quite simple fashion. It is worth mentioning here that SFX software is a widely used open URL link server. If the institution has an SFX server, DSpace will display an OpenURL link on every item page, automatically using the Dublin Core metadata. Additionally, DSpace can respond to incoming OpenURL.

6.10.8 Creative Commons Support

DSpace provides support for Creative Commons licenses to be attached to items in the repository. It is worthy to point out here that Creative Commons licenses provide a flexible range of protections and freedoms for authors, artists, and educators. It is one of the several public copyright licenses that authorize the sharing of copyrighted works.

6.10.9 OAI Support

The Open Archives Initiative has developed a protocol for metadata harvesting. The Open Archives Initiative-Protocol for Metadata Harvesting has become the de facto standard for metadata harvesting. This allows sites to programmatically retrieve or 'harvest' the metadata from several sources and offer services using that metadata, such as indexing or linking services. DSpace exposes the Dublin Core metadata for items that are publicly (anonymously) accessible. Additionally, the collection structure is also exposed via the OAI protocol's 'sets' mechanism.

6.10.10 Storage Resource Broker (SRB) Support

DSpace offers two means for storing bitstreams. The first is in the file system on the server. The second is using SRB (Storage Resource Broker). Both are achieved using a simple, lightweight Application Programming Interface (API).
6.10.11 My DSpace

This is the most attracting feature of DSpace system. DSpace offers to each user personalized access to information within the system through their My DSpace page. As appropriate given their role(s) in the system, users can view their:

- Items being assembled
- Submission of pending archive
- Archived items that they submitted
- Review tasks for Items pending archive

DSpace filters the above sections so that only the sections relevant to each user are presented. (Bass et al., 2002)

6.10.12 Batch Metadata editing tool

DSpace provides a batch metadata editing tool. The batch editing tool is able to produce a comma enclosed file in the Comma Separated Value (CSV) file format. It is worthy to mention here that CSV file format is a common structuring strategy for text format files. In CSV files, each line in the file represents a row of data and, within each line of the file, the different data fields are separated from one another using a comma.

The DSpace Foundation describes that the batch editing tool facilitates the user to perform the following:

- Batch editing of metadata (e.g. perform an external spell check)
- Batch additions of metadata (e.g. add an abstract to a set of items, add controlled vocabulary such as LCSH)
- Batch find and replace of metadata values (e.g. correct misspelled surname across several records)
- Mass move items between collections

- Mass deletion, withdrawal, or re-instatement of items
- Enable the batch addition of new items (without bitstreams) via a CSV file
- Re-order the values in a list (e.g. author)

6.10.13 User interfaces of DSpace

DSpace has two user interfaces as described below:

- The JSPUI user interface
- The XMLUI user interface

6.10.13.1 JSPUI Interface

Java Server Pages User Interface (JSPUI) is the standard user interface of DSpace that allows quick creation of fizzy web content. The customization of JSPUI interface is performed using Java Servlets which handle the business logic, and Java Server Pages (JSPs) which produce the HTML pages sent to an end-user. In view of the fact that the JSPs are much closer to HTML than Java codes, changing the look and feel of DSpace is rather simple. The Customization process in JSPUI is broadly divided into two categories:

- **Simple Customization**: It involves very basic changes to the page elements and style sheets to alter the look and feel of DSpace. These are very simple and quick to implement and it needs basic understanding of Cascading Style Sheets (CSS) and Hypertext Mark Up Language (HTML)

- **Advanced Customization**: It requires additional knowledge of programming Languages such as Java, Java Server Pages, etc, as it involves the editing of Java Server Pages to alter the look and feel or customize completely.

6.10.13.2 XMLUI/ Manakin User Interface

Manakin (XMLUI) is an “add on” based on Apache cocoon framework. It is worthy to mention here that “Apache Cocoon is a web development framework built around the concepts of separation of concerns (making sure people can interact and collaborate on a project, without stepping on each other toes) and
component-based web development. Cocoon implements these concepts around the notion of ‘component pipelines’, in which each component on the pipeline specializing on a particular operation. This makes it possible to use a ‘building block’ approach for web solutions, hooking together components into pipelines without any required programming”.

Manakin is a web-based user interface of DSpace that introduces a modular interface layer, enabling an institution to easily customize the interface according to the particular needs of the repository, community or collection. Manakin provides the new way to modify the user interface of DSpace.

The DSpace software divides its intellectual capacity into two: Aspects and Themes. Aspects are different parts of a DSpace website that can be customized according to the particular Digital Library request, such as the login mechanism. On the other hand Themes are the visual characters of these different Aspects that make up DSpace websites. These together provide a more customizable and attractive web-based interface for DSpace. (Gorton, 2007)

The customization of XMLUI interface is much complex type and requires higher knowledge of programming languages. Lewis and Yates (2008) gave four proposed objectives of the customization of XMLUI interface as given below:

- To allow each community and collection represented in DSpace to maintain a distinct look and feel.
- To increase support for internationalization in DSpace
- To separate the business logic from stylistic controls, increasing ease of adaptability.
- To provide an alternative interface to the current JSP based implementations, requiring no changes to the core of DSpace (including the

---

database), while to specify the enabling of both user interfaces to operate simultaneously.

6.11 Conclusion

DSpace software is the leading and widely used software for the generation of institutional repositories/digital libraries. DSpace has the privilege of having maximum number of installations over other software packages. It has various unique qualities which make it more valuable and flexible than other software packages. It has easily customizable and user friendly user interface. DSpace system is regularly developed and remains up to date. Introduction of Manakin system is the current novelty into the system which deals with the customization of user interface. The DSpace system has quite a large and vigorous community of users. Its installation is slightly tough but working with DSpace is extremely simple.
References

Salt Lake Community College. (2013). DSpace evaluation. Retrieved September 20, 2013, from slccarchivesir.wordpress.com/...collections/needs.../DSpace-evaluation-06...


190
Chapter 6  DSpace: Metadata Encoding


