Chapter 5

DSpace Digital Library Software Evaluation http://www.dspace.org

Introduction

DSpace is an open source digital library software developed jointly by MIT libraries and HP labs. DSpace provides tools for management of digital assets, and is commonly used for building institutional repositories[1]. It was basically designed to manage, host, preserve and enable distribution of the scholarly output of MIT's faculty.

DSpace helps to create, index and retrieve various types of digital contents which include research articles, grey literature, theses, cultural materials, 3D digital scans of objects, photographs, films, audio/videos, scientific datasets, institutional records, educational materials and other forms of content. DSpace was established to capture, preserve and communicate the intellectual output of an institution's faculty and researchers through central place[2].

The collection in DSpace is organised into communities, collections and items. The communities in DSpace include a high-level organizational structure whose only purpose is to divide collections into related groups. Each community contains one or more collections, which are containers for related items. An item is a deposited object of any type: a published article, an image, audio, or video file, notes, a presentation, etc.

DSpace is specially designed for digital preservation support for all the documents that are added into the repository in a simple fashion[3].

The first version of DSpace was released during November 2002. For the present study installation of 1.4.2 version was carried out and selected for evaluation study. This version was released on 11th May 2007. The latest version of DSpace is 1.5.2 which was released during March 2008. The following sections lists different observations after having installation of version 1.4.2.

5.1 Content Acquisition

5.1.1 **Which document types can be added in the software (books, reports, journal articles, lecture notes, technical reports, thesis, images, audio/video files, data set files etc.?)**

DSpace supports adding all types of digital documents ranging from books, reports, journal articles, lecture notes, technical reports, thesis, images, audio/video files to data sets.

5.1.2 **Does the software have capability to define the domain/scope of Digital Library for whom it is intended for?**

DSpace was especially designed for creating institutional repository of any organization.

5.1.3 **If yes, what level of users it is intended for?**

The software is intended for the faculty members of any organization.

5.1.4 **Which digital document file formats does the system support?**

DSpace by default supports to upload the following file formats in DSpace repository: Adobe PDF, AIFF, audio/basic, BMP, CSS, FMP3, GIF, HTML, image/png, JPEG, Latex, MARC, Mathematica, Microsoft Excel, Microsoft Powerpoint, Microsoft Project, Microsoft visio, Microsoft Word, MPEG, MPEG Audio, PhotoCD, Photoshop, PostScript, Real Audio, RTF, SGML, TeX, Tex dvi, Text, TIFF, Video QuickTime, Wav, Word Perfect, XML

5.1.5 **Does the system carry out duplicate checking while uploading the data? What happens if the same digital document is submitted multiple times?**

No, System doesnot carry out duplicate checking. If the same digital document is submitted multiple times it gets added into the repository.

5.1.6 **How does the system allow to upload the digital objects, whether normal digital object or compressed digital objects?**

Software supports to upload digital objects with normal size as well as compressed digital objects.

5.1.7 **Does the software support uploading digital objects from existing URL?**

No, System doesnot support uploading digital objects from existing URL, the document has to be available on local disk.

5.1.8 **Does the system support distributed/community based acquisition?**

Yes software has facility to acquire material from network through the DSpace users. A DSpace user can upload documents through his/her login area from anywhere on Internet.

5.1.9 **Does the software support metadata/content import? In which file format data is required for importing?**

Yes, DSpace supports content/metadata import feature. If data has to be imported from other DSpace repository then it requires collection id, eperson, source file all in XML format. If data is available in Dublin Core format DSpace supports to import metadata/digital objects.

5.1.10 **Does the system support importing of bulk metadata/contents of digital objects?** Yes

5.1.11 **Does the software support to import automatic metadata for the digital objects that are added into the repository?** No

5.1.12 **Does the software support metadata/content export? In which file format data is exported?**

The software supports content/metadata export of the collections or items which are available in the repository. The data which gets exported from DSpace is available through 'dsrun' command. The software exports data into dublin core format. Only metadata details are exported and full text files of all digital objects can be exported/restored with pg_dump command. DSpace also supports exporting data into METS format. The contents of the METS are in mets.xml file format. Generally data is exported in XML file format from DSpace.

5.1.13 **Does the system support export of bulk metadata/contents of digital objects?**

Yes, DSpace supports to export bulk digital objects.

5.1.14 **Does system allow adding past versions of the digital document files?** No

5.1.15 **Does the software support its own accession number for each document that gets added into the repository?**

Yes, software supports its own accession number and it is called as internal ID in DSpace which is unique for each live DSpace repository.

5.1.16 **It is necessary to know whether software supports weedout policy?**

DSpace supports weedout policy which is called as withdraw an item from the live repository. Withdrawing does not completely remove the item from the archive; it simply removes it from all public view (search and browse indices and so forth). DSpace does have a function to fully delete an item ('expunge') from the repository.

5.1.17 **Does the system support to delete items from the collection/repository, move objects from one collection to other collection, cross-list objects across different collections?**

Yes, DSpace supports to delete items from the collection, but does not support to move objects from one collection to other but instead allows to map one item across different collections.

5.1.18 **Which persistent identifier scheme does the system support?**

DSpace supports CNRI Handle system identifier for all documents that are added into the DSpace repository. Presently handles are assigned to communities, collections, and items in DSpace.

5.1.19 **Does the system's persistent URL method handle an object's change in location and state in future with same identifier number?** Yes

5.2 Content Management

5.2.1 **Does the system have proper workflow in place which handles different submission processes?**

Yes, DSpace has proper workflow in place where it allows to create different submission roles for different groups/members within the repository.

5.2.2 **If the system supports to have proper workflow for all actions into the repository then whether submission supports following options:**

5.2.2.1 **System allows to create overall system administrator** Yes

5.2.2.2 **System allows to submit documents by users via web** Yes

5.2.2.3 **System allows to set up submission rules** Yes

5.2.2.4 **System can define accept/reject policy for all documents that gets added into the repository** Yes

5.2.2.5 **Allows multiple collections within same installation of the system** Yes

5.2.2.6 **Home page for each collection/community** Yes

5.2.2.7 **Allows to set different policies for different collections** Yes

5.2.2.8 **Allows to set different authorization policies for different collection/items added into the repository** Yes

5.2.2.9 **Allow to edit metadata submitted by users** Yes

5.2.2.10 **Sends email notification to users/submitters/authors** Yes

5.2.2.11 **Sends email notification to metadata reviewers** Yes

5.2.2.12 **Sends email notification to reviewer** Yes

5.2.2.13 **Sends email notification to administrators** Yes

5.2.2.14 **Allow users to review completed content**

Yes, if user is authorized to review the content.

5.2.2.15 **Allow users to review uncompleted content**

Yes, if user is authorized to review the content.

5.2.2.16 **Allow content administrator to review submissions** Yes

5.2.2.17 **View pending content administration tasks** Yes

5.2.3 **Does the system support to upload documents into the repository through Librarian's interface as well?**

No, DSpace does not have any separate support of Librarians interface. All functions in DSpace are carried out through Web User Interface.

5.2.4 **Does the software provide easy way for adding/editing of records for Librarian as well as for the remote user? Are insertion and deletion of records easy?**

DSpace data entry interface is same for all. It has customized templates for adding/ editing records. It is easy to add/delete/modify records in DSpace.

5.2.5 **Are the supported file formats well documented in the system? Can new file formats be added or removed by the end user ?**

DSpace supported file formats are well documented in the repository.By default DSpace has defined set of file formats where MIT takes responsibility of providing file format preservation support in future for the defined formats.

5.2.6 **Does the repository software have any capabilities, such as access control lists, Internet address filters, etc., that limit who is allowed to submit items in the software?**

Yes, DSpace supports to have access control based on login and password.

5.2.7 **Does the repository software maintain audit logs that identify by whom and when documents are submitted into the repository?** Yes, DSpace keeps track of users logged into the system as well as keeps track of the IP and login name of the user and actions carried out by each user. All this information is kept in dspace.log file.

5.2.8 **Does the system support knowledge organization systems such as ontologies, thesauri, classification systems and taxonomies for users while adding documents via web in the repository?** Yes, DSpace supports controlled vocabulary support while adding documents.

5.2.9 **Does system supports digital objects to be grouped into organized structure (i.e. linear/hierarchical/graphical structure)**

Yes, DSpace groups all documents added into repository into communities/ collections/items.

5.2.10 **Does the system support to generate authority files?** DSpace supports to generate authority files of Authors and Subjects.

5.2.11 **Does system allow a digital object to be a member of multiple collections i. e. content mapping?** Yes

5.2.12 **Does system allow users to modify (add/delete/update) metadata/complete document added within the system?**

Yes, if the user is authorized to edit items then only system allow users to modify documents added into the repository.

5.2.13 **Does the system support to show strength of each collection?** Yes

5.3 Metadata Submission and Support

5.3.1 **Does the software support to add descriptive metadata (such as author, title, subject, publisher, copyright, year of publication etc.) for each digital object that gets added into the repository?**

DSpace supports to add descriptive metadata for all digital objects that are added into the repository. DSpace by default supports the following metadata fields such as Author, Title, Other Title, Date of Issue, Publisher, Citation, Series/Report No., Identifier, Type, Language, Subject Keywords, Abstract, Sponsors, Description fields[4].

5.3.2 **Does the software support to add administrative metadata (such as technical aspects of digital documents, source of information (when and how it was created), rights management, what is file size, in what file format digital document is in, what is needed to view digital document etc.) for each digital object that gets added into the repository?**

This includes preservation metadata, provenance and authorization policy data. Most of this data is held within DSpace's relational database schema. Provenance metadata is stored in Dublin Core records.

5.3.3 **Does the software support to add structural metadata (such as description, owner, data type, date deposited, version number, data of last revision, how compound objects are put together, how pages are ordered to form chapters) for each digital object that gets added into the repository?**

Structural metadata in DSpace is currently fairly basic level support.

5.3.4 **Does the software support adding/editing metadata fields?** Yes

5.3.5 **Are help messages given in each metadata field while entering data?** Yes

5.3.6 **Does the metadata entry field has one field mandatory? What are the mandatory fields in each software?**

Yes, 'Title' and 'Year of publication' are mandatory fields in DSpace software.

5.3.7 **Does the software support Unicode character set for metadata entry?**

DSpace supports UTF 8

5.3.8 **Does the software allow to create and manage a variety of different metadata schema's to meet different needs across the subject disciplines ?** No

5.3.9 **Can the metadata fields be customized?**

Yes, DSpace has inputforms.xml file which can be used to customize metadata fields as per the end users requirements.

5.3.10 **What type of Metadata Schemas are supported by the software as a default?**

DSpace currently uses a qualified version of the Dublin Core schema based on the Dublin Core Libraries Working Group Application Profile (LAP). DSpace uses the LAP as a starting point for its application of Dublin Core, borrowing most of the qualifiers from it and adapting others to fit. Some qualifiers were also added to suit DSpace needs.

5.3.11 **Does the software have default metadata entry templates?** Yes

5.3.12 **Are the metadata formats documented?** Yes

5.3.13 **Does the software support different interface for metadata entry and whether customized data gets added into the system?**

Yes, It is possible to create different interfaces for metadata entry and it can be customized to get data added into the system but it needs more knowledge about DSpace coding.

5.3.14 **How does the software verify that submitted objects and metadata are correct?**

While submitting document finally DSpace allows to verify all the metadata, and digital object before moving it into the repository.

5.3.15 **Are there automated checks of the metadata, such as to verify that a date entered into a field really is a date string?**

No, there are no automated checks available in DSpace current version.

5.3.16 **Does the repository software have any means to verify that objects or metadata have not been tampered with, such as checksums or digitally signed checksums?** No

5.3.17 **Does the software support real time updating and indexing of accepted contents?** Yes

5.3.18 **Does the software allow for metadata extensibility and complexity?** Yes

5.3.19 **Does the software support metadata versioning? What happens to old versions of metadata or content objects when they are replaced by new versions or modified in some way?** No

5.3.20 **Does the repository software support provenance metadata? What is included in the provenance?**

DSpace supports provenance metadata and in provenance the following information is added:

Name and email address of the submitter, Uploading date and time, Number of bitstreams uploaded, Total number of bytes of each bitstream, & Checksum data of uploaded bitstream.

5.3.21 **Does the software support metadata crosswalk?** Yes

5.3.22 **Does the software support 'thesaurus' building? Thesaurus helps in finding broader, narrower or related terms. Does it specify which thesaurus standard is used? Which subject heading DDC, DC or LC is used for thesaurus building?**

DSpace current version supports controlled vocabulary support. By default the software supports

nsi - nsi.xml - The Norwegian Science Index

srsc - srsc.xml - Swedish Research Subject Categories

The software has capability to add any controlled vocabulary as per the specification recommended.

5.3.23 **Does software has ontology support?** No

5.3.24 **How does the repository verify file types, for example, if a JPEG images is submitted to the repository, does the software verify that it really is a valid JPEG, or does it just blindly take the submitters word?**

Software verifies the listed file formats suppported by the software and then only allow to upload a particular file into the repository. If the users submits file which is not supported by DSpace, DSpace will upload that file into the repository but will not consider that file for long digital preservation.

5.3.25 **While exporting metadata whether software supports checksums?** No

5.3.26 **It is necessary to know whether the software support RAP protocol?**

Yes DSpace architecture has its roots in Kahn and Wilensky's Framework for Distributed Digital Object Services.

5.4 Classification

5.4.1 **Does the system allow digital objects to be grouped into classification/subjects?**

Yes DSpace supports to group digital objects as per keywords.

5.4.2 **It is necessary to know whether system supports creating of class numbers?**

No, DSpace doesnot support creating of class numbers.

5.4.3 **Which classification system does the software support?** No.

5.4.4 **It is also necessary to know whether system supports hierarchical browsing of collections through class number.**

No, DSpace doesnot support browsing based on class numbers.

5.5 Information Search & Retrieval

5.5.1 Browsing

5.5.1.1 **Author/Creator/Contributor** Yes

5.5.1.2 **Title of the document/Title of the book/Title of the article etc.** Yes

5.5.1.3 **Issue Date/Date of Publication** Yes

5.5.1.4 **Collection** Yes

5.5.1.5 **Communities** Yes

5.5.1.6 **Subject browsing** Yes

5.5.1.7 **Publisher wise browsing** No

5.5.1.8 **Table of Contents browsing** No

5.5.1.9 **Multi-Dimensional Browsing** No

5.5.2 Searching

5.5.2.1 **Searching content via different metadata fields such as author, title, subject, publisher etc.** Yes

5.5.2.2 **Full text searching** Yes

5.5.2.3 **Boolean (AND, OR, NOT) searching** Yes

5.5.2.4 **Basic search** Yes

5.5.2.5 **Advanced search** Yes

5.5.2.6 **Does the system allow to search the database putting boolean operators such as !, |, &, $\sim$, +, −, etc. as well as 'must contain', 'should contain' 'may contain', 'not contain', as operators in place of AND, OR, and AND NOT Boolean operators respectively.** DSpace supports boolean operators in search string.

5.5.2.7 **Truncation/Wild card searching** DSpace supports searching via right truncation, middle truncation, but does not support left truncation.It supports searching via '$*$*' or '$\sim$'

5.5.2.8 **Exact words/phrases searching** Yes

5.5.2.9 **Proximity searching** Yes

5.5.2.10 **Stemming** Yes

5.5.2.11 **Fuzzy searching** Yes

5.5.2.12 **Phonetic searching** No

5.5.2.13 **Case sensitive or case insensitive while searching** No

5.5.2.14 **System should support term weighting** No

5.5.2.15 **System should support search history option , reuse of query, query save option.** No

5.5.2.16 **Boosting the term** Yes

5.5.2.17 **Range searching** No

5.5.2.18 **Use of thesaurus or permuted index for searching** DSpace supports thesaurus support for searching.

5.5.2.19 **Expand search** No

5.5.2.20 **Lateral searching** No

5.5.2.21 **Search support for special collections** No

5.5.2.22 **Searching via table of content and classification numbers** No

5.5.2.23 **Ability to browse subject/author authority files** No

5.5.2.24 **PIC variant searching (plurals, international spelling variants, and compound words)** No

5.5.2.25 **Support multilingual search and retrieval** Yes

5.5.2.26 **Allow refinement of searches with qualifiers** No

5.5.2.27 **Ability to group: prior to search, user can delimit pending results listings, and group results by item characteristics (type, format)** No

5.5.2.28 **Spell checker support** No

5.5.2.29 **Refine searches** No

5.5.3 Sorting

5.5.3.1 **Author** Yes

5.5.3.2 **Title** Yes

5.5.3.3 **Issue Date** Yes

5.5.3.4 **Relevance** No

5.5.4 **Which search engine toolkit is used for searching the database?**

DSpace uses Lucene search engine toolkit for searching the database.

5.5.5 **How is the recall and precision ratio while searching the documents?**

Since the system supports searching via metadata fields and full text searching the recall is high via word searching and precision is high if the metadata fields are used for searching the database.

5.5.6 **How is the speed of the retrieval of the documents?**

The speed of the retrieval is very fast though it doesnot show on the screen how many documents are retrieved within how many seconds.

5.5.7 **Does the system provide help for the users while searching digital documents?** Yes

5.5.8 **How the results are displayed to the users? Whether users can download contents after searching? How are results displayed such as metadata, filetype, file size?**

While displaying results, software display issue data, author and title for all. Users can either view or download the contents after searching. Users can also see full item view or short description of each searched document.

5.5.9 **While viewing the results does the system provide access control?**

Yes, DSpace allow to restrict access at different levels such as item level, collection level and community level.

5.5.10 **Does system support streaming audio/video option? Or it allow only to download documents? Whether browser can support any other document viewing? Does the end user need to have plugins and software for viewing particular documents?**

DSpace doesnot yet support streaming video option. For the audio/video files which are uploaded into the DSpace repository are first downloaded into the local disk space of end user and then user can see the file or it is possible to configure any open source streaming server such as Darwin to configure with DSpace and allow end user to directly view multimedia files. End user needs to have a plugin and necessary audio/video software installed to view particular file formats. For e.g. DSpace doesnot support viewing of .dat files in that case if the file is retrieved the software will ask for downloading the file. But if file is in mp3 format and DSpace supports uploading mp3 file formats and it will automatically load that file on the browser.

5.5.11 **Does the system support advanced features such as my binders, virtual bag, shelving etc. ?**

DSpace supports 'My DSpace' feature which allow users to receive daily e-mail alerts of new items added into different collections. Users may subscribe to as many collections as they wish.

5.5.12 **Whether system supports unified search (Unified search is the support for finding items that are scattered among a distributed collection of information sources or services, typically involving sending queries to a number of servers and then merging the results to present in an integrated, consistent, coordinated format)?**

No, DSpace does not support unified search.

5.5.13 **Whether system supports federated search (Federated search allow users to search multiple digital libraries with a single query from a single user interface. The user enters a search query in the portal interface's search box and the query is sent to every other digital library software)?**

No, DSpace does not support federated search.

5.6 Access Control, Privacy and Management

5.6.1 **How users are managed in digital library software?** DSpace creates e-persons for all the members who register themselves through web browser. It is called as My DSpace. For each member DSpace holds email address, first and last name, list of collections for which the e-person wishes to be notified of new items, password details etc.

5.6.2 **Does the system support to add/edit/delete user profiles?** Yes

5.6.3 **Does the system support to define different roles for different users?** Yes

5.6.4 **Does the system keep detailed information of each user that registers in the repository such as name, email, phone number, address etc?**

No,System only keeps details of the name, password and email address.

5.6.5 **How the permissions are given to users who have registered into the repository ?** Each user who has registered into the repository have different permissions for different roles.

5.6.6 **Does the system have provision to identify active and inactive members and take necessary action against inactive members?** System does not keep track of existing and inactive members.

5.6.7 **User Roles**

5.6.7.1 **Does system allow to create different groups for different users** Yes

5.6.7.2 **Does system allow to create different groups of users** Yes

5.6.7.3 **Does system allow definition of different user groups** Yes

5.6.7.4 **Does system limit access by role** Yes

5.6.7.5 **Does system allow collection to be customized for each role/group** Yes

5.6.8 **Does the system allow access limits at different levels?**

(a) File/object level Yes

(b) Collection level Yes

(c) User/Group level Yes

5.6.9 **How passwords are administered in the system?** All passwords are encrypted into the system and are stored in the database.

5.6.10 **Does the system assign passwords for each user?** No, User can define their own passwords after receiving an email from DSpace repository.

5.6.11 **Does the system allow users to select passwords?** Yes, Individual user's can select their own passwords.

5.6.12 **Does the system have mechanism to retrieve forgotten password?** Yes, DSpace will send an email to the end user for retrieving forgotten passwords.

5.6.13 **How is the level of access for Inside/Outside users (such as outside users can access only metadata and abstract level access and internal members can access full texts as well as metadata)?**

DSpace supports to provide different action levels for different groups as well as for different communities, collections and items. By default all items added into the repository have a policy of 'anonymous' READ which means everyone can access all items in the repository. But software supports to define different access permissions at different levels.

5.6.14 **Does the system have any mechanism to provide access to the digital library via**

(a) IP source address filtering No

(b) Proxy filtering No

(c) Credential-based access Yes

5.6.15 **Are access denials flagged in any special manner by the software? Does the software keep access logs and how accessible are the logs to either human or machine processing and interpretation?**

No, DSpace does not deny access control in a special manner only it will ask for the user name and password if a particular community, collection or item is not visible to end user and DSpace will ask to provide user name and password.

5.6.16 **Does the system support access to the collection for a specific period?**

No, DSpace can not define access to a particular collection or item for a specific period.

5.7 Authentication and Authorization

5.7.1 **Does the system support to provide setting different authorization policies?** Yes, DSpace default template can define different delegation of authorities such as submitter, accept reject step, accept/reject/edit metadata step, edit metadata step, collection administrators, item template, collection's authorizations.

5.7.2 **System should support restricting access to contents through password authentication, IP filtering, and proxy filtering.**

Authorization system in DSpace is based on associating actions with objects and the lists of EPeople who can perform them.

5.7.3 **System should be able to support access based on browsing/viewing metadata, certain records, full text item or collections through different authorizations with different privileges for different communities or individuals.** Yes

5.7.4 **System should also be able to allow different roles for different members for different items/collections.** Yes

5.7.5 **Does the system support to have a super user/admin user who is overall manager/ administrator of the software?** Yes

5.7.6 **Does the system use authorization to support different roles within the system (for eg submit user, metadata editor, reviewer, approver, overall administrator etc.)** Yes

5.7.7 **Does the system use an external authentication mechanism (eg. Lightweight Directory Access Protocol).**

Yes, DSpace can be configured with LDAP.

5.7.8 **Does the system display only those pages that are pertinent to each user for which he/she is able to see?** Yes

5.7.9 **Does the system support the feature that groups are assigned to one or more roles?** Yes

5.7.10 **Does the system support the feature that users are assigned to one or more groups or collections?** Yes

5.7.11 **Does the system support the following roles:**

(a) Content Developer Role: must be able to create/add/edit/delete metadata and content in defined (not ALL) collections in addition Yes

(b) Administrator Role: must be able to add/delete/create users and groups Yes

(c) Contributor Role: must be able to submit metadata and content, in addition to general users role Yes

(d) General User Yes

(e) Submitter Yes

(f) Metadata Reviewer Yes

(g) Max Role:must be able to do all activities within the system Yes

5.7.12 **Does the system use authorization to verify users with campus security systems (NetID)**

If LDAP is configured it is possible to verify users with campus security systems (netID)

5.8 Interoperability

5.8.1 **It is necessary to know whether the software supports Z39.50 protocol and OAI-PMH protocol?**

DSpace supports OAI-PMH protocol for metadata harvesting.Since every item that is added into the DSpace repository have one Dublin Core descriptive metadata record, DSpace allows basic metadata interoperability across all of the items in DSpace.

5.8.2 **Which harvesting software is used by the Digital Library Software for metadata harvesting?**

DSpace uses PKP Harvester software.

5.8.3 **SRU/SRW Support** DSpace supports SRU/SRW service.

5.8.4 **Does the software have openURL support?** DSpace supports openURL protocol from SFX.

5.9 Ease of Deployment of each software

5.9.1 Software Installation

5.9.1.1 **Packaging and installation steps**

The installation process of DSpace is added in Appendix B of this document.

5.9.1.2 **Automatic installation script**

There is no automatic installation script available for DSpace. One has to follow different installation steps.

5.9.1.3 **Disk space used**

DSpace requires around 500 MB disk space to install DSpace along with other software.

5.9.1.4 **Time required to install each software**

Since DSpace doesnot have automatic script to install it takes few hours around 3 to 6 hours to successfully install DSpace if an end user has proper knowledge of Linux Operating System and database configuration as well as configuration of Java Servlet Server.

5.9.1.5 **Documentation help available and useful for installation**

DSpace has documentation page as well as wiki page on installation but its not detailed step by step.

5.9.1.6 **Environments needed to set for installation of each software**

JAVA, Tomcat and Ant environments are required to be set before installation.

5.9.1.7 **Operating systems (Linux/Windows/Solaris/MacOS/GNU/AIX) on which software can be installed**

DSpace works on FreeBSD, Debian, Ubuntu, Mandriva, OSX, Mac OS X, Red Hat, Gentoo, SLES9 as well as on windows (but with limited support).

5.9.1.8 Programming languages used

DSpace uses JAVA and server side JAVA technologies including JAVA servlets, JSP, Taglets, Filters, Java Bean Activation Framework and Java mail.

5.9.1.9 Database supported at backend

It uses PostGreSQL database as a default backend database but can be customized to use Oracle.

5.9.1.10 Web Server used

Apache Web Server

5.9.1.11 Java servlet engine needed

Apache Tomcat or Jakarta Tomcat

5.9.2 System Support/Maintenance

5.9.2.1 Documentation/manuals

Documentation is available on DSpace web site http://www.dspace.org

5.9.2.2 Mailing lists/discussion forums

http://sourceforge.net/mail/?group_id=19984 DSpace has general mailing list, developers mailing list and technical mailing lists all these lists are available from sourceforge site.

5.9.2.3 Wiki pages

DSpace wiki page has been widely used by the DSpace community and it can be accessed from http://wiki.dspace.org

5.9.2.4 Help desk support

DSpace has help support from http://libraries.mit.edu/dspace-mit/ info/dspace-help.html

5.9.2.5 **Ease of system administration (ability to configure for different uses)**

System administrator can configure different tasks easily from web user interface of DSpace

5.9.3 Hardware/Software Requirements

5.9.3.1 **What are the minimum hardware requirements for the repository software?**

Atleast P4 machine, with 256 MB RAM and 40 GB hard disk.

5.9.3.2 **How actively is the software maintained by its owners or contributors?**

Since DSpace has been widely used all over the world it is actively maintained by its owners and contributors while bringing new versions often.

5.9.3.3 **Does the software have procedures in place to monitor and receive notifications when hardware technology changes are needed?**

DSpace cannot provide any notification when hardware technology changes are needed.

5.9.3.4 **Does the software creator's have a process to stay current with the latest operating system security fixes?**

No information about this feature is found anywhere.

5.9.4 Security

5.9.4.1 **Does the system have security methods such as**

(a) Data encryption Yes

(b) Digital signatures No

5.9.4.2 **Server security (Does software has secure database connection?)**

Yes the database in DSpace is stored in postgresql where postgres has username and password for DSpace database.

5.9.4.3 **Ability to restrict access repository at item level (For eg to view metadata but not content)** Yes

5.9.5 System Level Security

5.9.5.1 **Fixity - such as checksums or digests**

DSpace supports md5 checksum for all objects that are uploaded into the repository

5.9.5.2 **Management of user accounts and rights to specified locations within the DL** No

5.9.5.3 **Logging and auditing events**

The details of each users logging is saved in dspace.cfg file.

5.9.5.4 **Does the software support any security if OS does not have any firewall implemented?**

In DSpace all the passwords are encrypted. There are no other access restrictions which can be implemented in DSpace.

5.9.6 General features related to technical aspect

5.9.6.1 **Does the software have any back end maintenance?** No

5.9.6.2 **Does the software have RSS support?** Yes

5.9.6.3 **Does the software have easy mechanisms to upgrade the software from old versions to new versions?**

An upgrade from version 1.3.2 to version 1.4.2 was carried out but it did not work successfully.

5.9.6.4 **Does the software support migration of data from one repository software to another repository software?** Yes

5.9.6.5 **Does the system support any ranking algorithms for the documents added in the repository?** No

5.9.6.6 **Are there any known issues/bugs in the system?** No

5.9.6.7 **Whether a novice user can easily publish content ?** Yes

5.9.6.8 **How often user needs to use database schema for making any change?**

End user cannot get access to database schema unless he has login and password of posgtgreSQL database. Only system administrator can access database schema.

5.9.6.9 **Whether user can reorganize database?**

If the user has knowledge of PostgreSQL and DSpace software he can reorganize database.

5.9.6.10 **Does system support automated content acquisition, harvesting and automatic metadata generation, automatic subject indexing/classification?** No

5.9.6.11 **Does the system support to store metadata records separately from the actual contents?**

Yes, metadata is stored in PostgreSQL database and actual files are stored in 'asset-store' directory where DSpace is installed.

5.9.6.12 **Does the software support storing varied file formats?**

By default DSpace has given a list of supported file formats. But if the end user make changes in the format registry of their installations he can define any type of format which DSpace can upload but then it becomes end users responsibility to migrate those file formats to the current file formats.

5.9.6.13 **Does the system keep the original file's name, size and created date?** Yes

5.9.6.14 **While using the software if there are any errors occur does the system allow to submit the error report?**

Yes, end user can submit the error report to the system administrator by sending an email.

5.9.6.15 **How many developers are working for the software?**

There are 14 committers and several development teams around the world who are working for the software as mentioned on sourceforge.net site.

5.9.6.16 **How many users are using the software/strength of community?**

The number of installations of DSpace as on 7th March 2008 are 307 installations in 51 countries with 2.330.639 Documents.

5.9.6.17 **Does the software provide its history on the web page for evidence of sustainability and vitality?**

Yes, software has its wiki page as well as version history page on its web site. Since the project is funded by MIT and HP the software the sustainability issue of the software is not a major concern.

5.9.6.18 **Total number of downloads of the software?**

As on 7th March 2008, the total number of downloads of version 1.4.2 for linux were 3926.

5.9.6.19 **Is the software really under Open Source License terms and conditions?**

Yes, The DSpace system is freely available as open-source software under the terms of the BSD distribution license.

5.9.6.20 **Are there any costs involved for using OSS-DL?** No

5.9.6.21 **Does the repository software have a clear policy regarding software upgrades?** Yes

5.9.6.22 **Whether the software has a long-term leadership and consulting services?**

Yes through DSpace Foundation.

5.9.6.23 **After releasing the software whether it was tested and whether it has integrated new features or not?**

Yes, software beta testing is done regularly.

5.9.6.24 **How is the mailing list used by the users?**

Mailing lists of dspace whether general, technical or developer mailing lists are used by DSpace users very actively. Table 5.1 shows number of queries submitted by users on DSpace general mailing list. The table shows data as observed on 7th March 2008.

5.9.6.25 **Whether software has conducted any workshops, training sessions on local, regional, national and international levels?**

DSpace user meeting is taking place every year since 2004 as well as all over the world there are workshops held on DSpace. In India DRTC has taken a leading role in marketing DSPace at various organizations and have conducted several workshops on DSpace. DRTC has also developed a dlrg page i.e. http://drtc.isibang.ac.in/mailman/listinfo/dlrg. for DSpace support for the users to submit their queries on DSpace, among other discussion regarding DL research.

Table 5.1: DSpace User's Mailing List Data

Year	Number of requests sent on DSpace User's mailing list
2003	89
2004	378
2005	336
2006	463
2007	552
2008, March 7th	107

5.9.6.26 **Whether software is built according to open standards?** Yes

5.9.7 Architecture of the Software

5.9.7.1 **How is the architecture and design of the software whether it is simple or complex?**

The DSpace architecture is three-layer architecture, including storage, business, and application layers, each with a documented API to allow for future customization and enhancement. The storage layer is implemented using the file system, as managed by PostgreSQL database tables. The business layer is where the DSpace-specific functionality resides, including the workflow, content management, administration, and search and browse modules. The application layer covers the interfaces to the system: the web UI and batch loader, in particular, but also the OAI support and Handle server for resolving persistent identifiers to DSpace items. This architecture was designed during 2001[5].

5.9.7.2 **Does the architecture supports separation between different local parts and put into different machines (e.g. centralised/distributed database, relational/object-oriented database management system, different components of the directory to distribute to different machines) and the transport model (protocols for communication between the system and the user interface or between system components).**

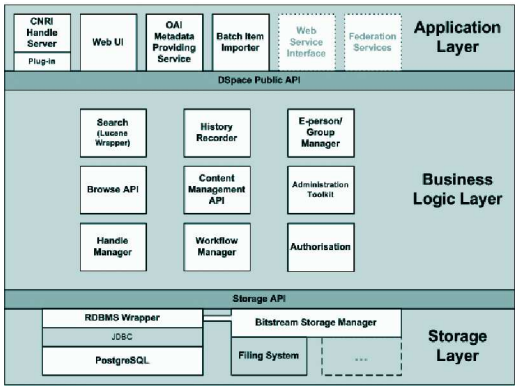

Figure 5.1: DSpace Architecture

Yes, software supports separation between different local parts such as database can be on some machine and web server and actual files on other machine.

5.9.7.3 **Does system support grid architecture? How grids can help to disaster recovery of data?**

Yes

5.9.8 **Scalability: It is necessary to know whether software is scalable in terms of storage and retrieval?**

Current version of DSpace can scale to store up to 3 Tera Bytes of data. DSpace is architected with scalability in mind. The assetstore is decoupled from the application services such that the two can be scaled independently.[6] The scalability factor is considered for the next development of DSpace i.e. Ver. 2.0[7].

5.9.9 **Extensibility: Extensible, to allow new elements (collections, data types, services, etc.) to be easily added to the digital library.**

DSpace can be extended with new services, data types and formats. It is easy to add new features in DSpace.

5.9.10 **Storage: Does the system support data compression, for eg, in case of multimedia collections the size of the files (for eg video/audio files) would be very large hence system should support data compression/decompression techniques. Does the system support to store compress file formats?**

No, DSpace doesnot support any compression and decompression technique.

5.9.11 **Backup/Restore facility**

5.9.11.1 **Does software support backup function, which contains metadata associated with access controls, customized files, as well as full texts of the documents added in the repository?**

No, The software doesnot have any backup function. End users have to take back up of bitstream and metadata fields on their own.

5.9.11.2 **Does the software explicitly require any particular backup strategy, or does it just rely on system-level backup plans, like periodic disk backups to tape?**

Yes, DSpace rely on system level back up i.e. all digital documents are stored in /home/dspace/assetstore directory and pgsql database. Hence end user has to take back up of pgsql and assetstore directory.

5.9.11.3 **In the event of a disaster, what recovery plans are supported by the software?**

No information was found on this aspect.

5.10 User Friendly Interface

5.10.1 **Whether end user can easily customize the user interface 'look and feel'? Whether the end user can change the header, theme, footer, overall web interface of the system?**

Yes, the end user can customize the user interface as per end user's needs. DSpace users can change the complete look of the user interface such as header, footer, theme etc.

5.10.2 **Can user interfaces be localized in their language?**

Yes, DSpace supports to localize user interface in any language.

5.10.3 **How is user interface for the experienced or non experienced users?**

It is easy for end user to submit the document in DSpace ones a user becomes member of the repository.

5.10.4 **How is the user interface designed for the use of Librarian or for the use of users?**

User interface is same for the library staff and for the end user.

5.10.5 **Whether system provides web based interface for all its functionalities?**

Yes all functions in DSpace are carried out through web interface.

5.10.6 **Does the system provide online help?** Yes

5.10.7 **Does the system provide multilingual access support?** Yes

5.10.8 **Does the system allow users to submit feedback on system problems?** Yes

5.10.9 **Does the system notify user's current state of the task?** No

5.11 Usability

5.11.1 **What user skill levels are expected to use the web interface?**

It is easy to use the web interface of DSpace.

5.11.2 **Is the user able to navigate and find his/her ways to information quickly and easily?**

Yes

5.11.3 **How are the help features provided with the software? Does help feature allow a user to feel comfortable within a digital library? Does the help feature provide general as well as technical answers to the problems?**

The help feature provides only general information.

5.11.4 **Does the user need to know which player or viewer is necessary to view multimedia or other digital objects from the system?**

No, DSpace automatically finds out appropriate plugin and open the document for the end user.

5.11.5 **Does the user interface provide facility to give user opinion on the web as well as does the software support user to fill up online user feedback form?** No

5.11.6 **Does the system support to show usage statistics such as usage patters and which material is used more by which group, how often, when, for what reasons/decisions etc:**

i. Usage patterns Yes
ii. Use of materials Yes
iii. Usage statistics Yes
iv. Who uses what, when Yes
v. For what reasons/decisions No

DSpace supports to generate different reports such as :

- Number of item views
- Number of collection visits
- Number of community visits
- Number of OAI Requests
- Customisable summary of archive contents
- Broken-down list of item viewings
- A full break-down of all system activity
- User logins
- Most popular searches

The results of statistical analysis can be presented on a by-month and an in-total report, and are available via the user interface. The reports can also either be made public or restricted to administrator access only.

5.11.7 Transaction Log Analysis

5.11.7.1 **Does the system keep error log files?**

DSpace keeps all log data in log directory where dspace.log file records every action that is carried out in the repository.

5.11.7.2 **Does the system keep user logs?**

Yes, DSpace store all data about users logged into the system in dspace.log file.

5.11.7.3 **Does the system keeps track of query's submitted to the system? Does it store query log?**

No, DSpace doesnot store query log.

5.12 Copyright/Policy Issues

5.12.1 **Does the repository software have any means to manage, store, or enforce contract agreements for all the digital documents that are added into the repository?**

Yes, the software has a facility to keep track of all documents that are added into the repository along with their copyright details.

5.12.2 **Does the software have any capabilities, such as access control lists, Internet address filters, etc., that can be used to enforce copyright or access restrictions? How**

granular are these access controls? Can different restrictions be applied to different objects in the repository?

Yes, DSpace supports to control access based on access control list. If the copyright doesnot allow any document to be openly accessible DSpace has the facility to restrict access at item/collection/community level.

5.12.3 **If repository ingests digital content with unclear ownership/rights, does it have policies addressing liability and challenges to those rights.**

Yes, DSpace supports to record rights policies for every document that is added into the repository.

5.13 Advanced Features

5.13.1 **Whether software has long term leadership and consulting services?**

Yes, DSpace has long term leadership and consulting services. Number of installations of DSpace are increasing day by day. Since DSpace is jointly developed by MIT and HP it will be supported in future.

5.13.2 **Does the system provide citation data?** No

5.13.3 **Does the software have defined road map for the future?**

Yes DSpace 2.x roadmap is ready and development is going on as per the road map.

5.13.4 **Does software support indexing of all documents that are added in the repository in a distributed way, without central control?** No

5.13.5 **Virtual Collection Support: Does software supports to generate virtual collections /special collections across several content providers?** No

5.13.6 **Load Balancing: Does software has any mention of load balancing?**

No mention about load balancing was found.

5.13.7 **Visualization: Does the software support visualization feature?** No

5.13.8 **Personalization: Does the system support Personalization feature?**

DSpace supports “My DSpace” feature where end user can subscribe to different collections.

5.13.9 **Translation Service: Does the system provide translation service?** No

5.13.10 **Community Services: Does the software supports to provide community services such as allowing members of the digital library community to exchange ideas, make announcements, write reviews etc.** No

5.13.11 **Page Ranking: Does the software supports page ranking such as supported by commercial digital libraries.** No

5.14 Digital Preservation

5.14.1 **Does the software support any digital preservation strategy and if yes, does it explicitly support any particular preservation strategy, such as described by PREMIS like, bit-level preservation, format migration, format normalization, emulation, or restrictions on submission formats?**

DSpace supports bit preservation, where a digital file is carefully preserved exactly as it was created without the slightest change. While submitting documents in DSpace system keeps track of known bitstream formats and their support level. The repository provides a list of supported file formats. Supported formats include those that are documented standards (e.g., TIFF, AIFF, XML) or have published specifications (e.g., PDF, RIFF).[8].

The other two categories of support for MIT's DSpace are "known" and "unsupported". "Known" formats are those that are common enough to be familiar and usually quite popular, but which are proprietary in that there are no published specifications on which to base functional preservation. "Unsupported" formats are those that are either unknown to the Libraries or are extremely rare (e.g., a compiled program, a commercial CAD/CAM file, etc.).

If the software supports a particular file format it shows the list of supported file formats else, if format is not supported by DSpace it will give a message "DSpace could not identify the format of this file. Please describe the format file in the input box below the list".

5.14.2 **Whether system preserves the file's original identities such as its name, size and created date?** Yes

5.14.3 **Does the system have any data integrity check for a file added into the repository?**

Yes the system has built in data integrity check by using MD5 (a message digest algorithm for security applications) to ensure the correctness of each file.

5.14.4 **Does the software have quality control measures to ensure integrity, and persistent documentation identification for migration purposes?**

No information about this was found anywhere.

5.14.5 **Does the repository software preserve pre-existing persistent identifiers for submitted packages or objects?**

No, DSpace does not have any feature to preserve pre-existing persistent identifiers for submitted digital objects.

5.14.6 **Does the software support preservation of metadata? Preservation metadata is information that supports and documents the long-term preservation of digital materials. Maintaining a history of a digital object with metadata is a key part of digital preservation strategy. It addresses the following attributes of an archived digital object:**

- provenance —documenting the history of the object.
  The provenance information in DSpace consists of information such as submitted by, email of the submitter, date, total number of bitstreams submitted, date of making document available on DSpace web site, actual title of the file format uploaded, total bytes and checksum data. This data is maintained in DC metadata.
- authenticity —validating that the digital object is in fact what it should be, and has not been altered. No
- preservation activity —documenting the actions taken to preserve the digital object. No
- technical environment —describing the technical requirements, such as hardware and software, needed to render and use the digital objects. No
- rights management —recording any binding intellectual property rights that may limit the repository's ability to preserve and disseminate the digital object over time. Yes

5.14.7 **How does the software manage compound objects (where multiple file formats of same object are linked together)?**

Submitters are allowed to submit multiple file formats of same document at one time.

5.14.8 **Does it have mechanism to keep licensing conditions for individual images/objects in the repository?** Yes

5.14.9 **Does the software support to track and manage copyrights and restrictions on use as required by contract or license?** Yes

5.14.10 **Where does the repository software store actual digital files and the metadata?**

Actual digital files are stored in assesstore directory of DSpace installation and metadata is stored in postgersql database.

5.14.11 **Can the repository software automatically validate checksums on a periodic basis?** No

5.14.12 **Are the checksums cryptographically signed to prevent tampering?** Yes

5.14.13 **Does the repository software have a well documented process by which a submission is ingested into the repository for storage? For example, how and when does the repository software generate fixity data, such as checksums?**

The software generates fixity data after submitting a file into the repository be the submitter.

5.14.14 **How does the repository software verify that archival objects and metadata are correct? Are there automated checks of the metadata, such as to verify that a date entered into a field really is a date string?** No

5.14.15 **Does the repository software provide audit logs of all events that have occurred in the life cycle of a package?**

Yes, DSpace create dspace.log file which keeps track of all transactions that are carried out in DSpace as well as Tomcat server has its own log which records server access log details.

5.14.16 **If repository ingests digital content with unclear ownership/rights, does software has policies addressing liability and challenges to those rights?** Yes

5.14.17 **Does the software have mechanisms to change its preservation plans as a result of its monitoring activities?** No

5.14.18 **Does the system have any mechanism in determining when objects in digital archives should migrate to new hardware and software?** No

5.14.19 **Can the software support scheduled events such that a human can be notified on a preset schedule to manually check for format obsolescence?** No

5.14.20 **Can new file formats be added or removed?**

It is possible to add new file formats in DSpace but they may not be supported by DSpace from preservation point of view.

5.14.21 **Does the software have ability to handle variety of file formats and does it also supports file format versioning?** DSpace version 1.4.2 does not support file format versioning.

5.14.22 **Does the software have mechanisms in place for monitoring and notification when format (or other representation information) obsolescence is near/or are no longer viable?** No

5.14.23 **Can the software monitor any standard format registries in order to ascertain format obsolescence?** No

5.14.24 **What standards does the repository software use to describe file formats and does the software record representation information? Does it use Internet MIME Types?**

DSpace uses MIME type standard for all the description of all file formats which are supported by DSpace. Software doesnot record any representation information.

5.14.25 **Does it use any format registries? If yes, which format registry is used such as PRONOM or GDFR or DCC to represent format information?** No

5.14.26 **Does the system support automatic format registration, For unknown formats does the system sends any message to the submitter requesting for additional information?**

System does not have any way for automatic format registration. While submitting any document into DSpace repository, the software gives list of DSpace supported file formats and if the particular file format is not in list it provides option for the submitter to write details about the file format which is submitted by the submitter. System does not send any message to the submitter for additional information.

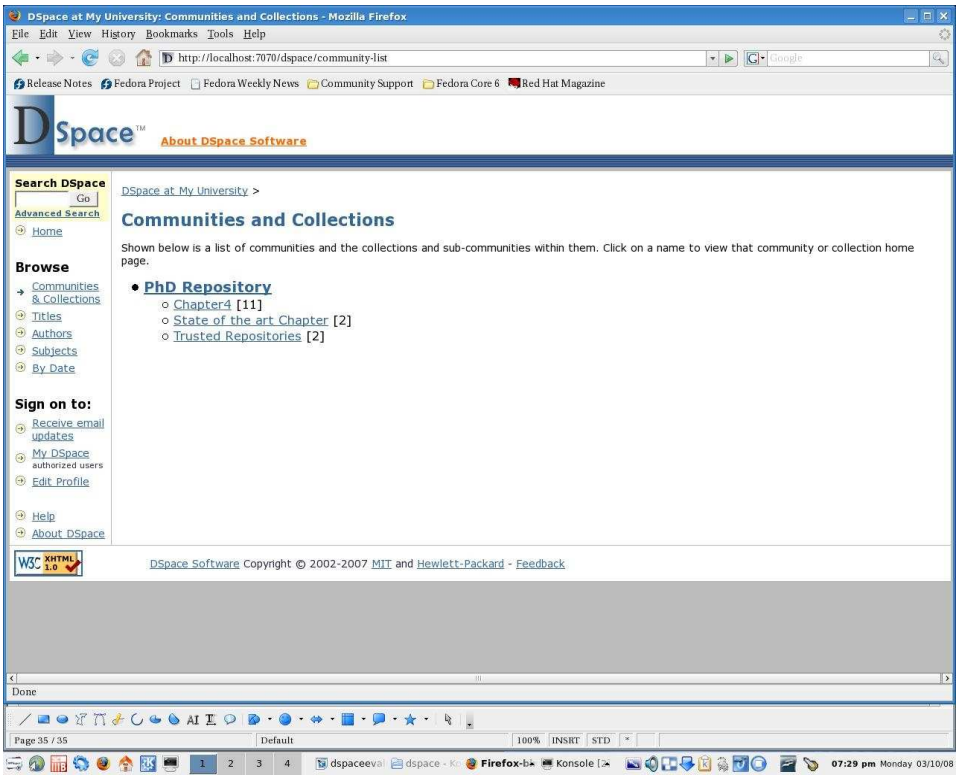

Figure 5.2: DSpace Installation Screen

Bibliography

[1] http://en.wikipedia.org/wiki/dspace. (Browsed on 7th March 2008).

[2] Smith, M., Bass, M., McClellan, G., Tansley, R., Barton, M., Branschofsky, M., Stuve, D., Walker, J. H.(2003). DSpace: An Open Source Dynamic Digital Repository. *D-Lib Magazine, 9(1)*, http://www.dlib.org/dlib/january03/smith/01smith.html. (Browsed on 7th November 2006).

[3] Tansley, Robert, Bass, Mich & Smith, MacKenzie (2004). DSpace as an open archival information system : current status and future directions. *Research and Advanced Technology for Digital Libraries, 2769*, Lecture Notes in Computer Science, p. 446-460.

[4] Tansley, Robert [et al.](2003). The DSpace institutional digital repository system : current functionality. *IEEE*, p.87-97.

[5] Smith, M., ibid.

[6] Baron, M. R., Walker, Julie H., (2002). MIT libraries' DSpace business plan project final report to the Andrew W. Mellon Foundation, 33 p. http://wiki.dspace.org/static_files/4/42/dspace_governance_report_03142006.pdf. (Browsed on 10th June 2007).

[7] http://cadair.aber.ac.uk/dspace/bitstream/2160/316/2/DSpace+Foundation.ppt. (Browsed on 17th December 2007).

[8] Smith, M., ibid.