

Chapter 3

Evaluation Criteria Defined for Evaluating Open Source Digital Library Software (OSS-DL)

3.1 Introduction

In 1999[1] Lakos used the phrase "culture of assessment" in libraries to develop evaluation strategies regarding library services and resources available in libraries. It is important for libraries to continuously assess services they provide and understand user requirements. Many evaluation studies are carried out in libraries while assessing the library services provided to the library users. Culture of assessment is an ongoing activity which makes continued improvement in meeting both library and customer needs.

During 1994, major initiatives took place in building digital libraries and lot of funding was made available nationally and internationally to preserve world's cultural heritage in digital form. Many approaches are being used for building digital libraries and many types of collections and media are included by processing in different ways and these collections are made available to end users.

Since now the digital collections are growing, it is necessary to evaluate various services offered by these digital libraries. In digital environment, assessment can be carried out at various levels such as by understanding requirements of the users and libraries as well as assessment can be carried out by assessing the software which is used for creating digital libraries.

Today large number of researchers all over the world are trying to research on various aspects of digital libraries such as architectures of digital libraries, information access in digital libraries, multimedia/ multilingual retrieval in digital libraries, user interfaces/ visualization, digital preservation, database management systems in digital libraries, scalability etc. Well established researchers in Europe and USA are exploring how next generation digital libraries would be.

While allotting millions of dollars for building digital libraries, fraction of amount is spent on research on any evaluation aspect of digital libraries. Evaluation of digital libraries is still in its infancy[2]. Researchers are still investigating who should evaluate, when to evaluate, what to evaluate, how to evaluate and why to evaluate digital libraries. Parallely, libraries have started making use of commercial software or open source software for building digital libraries. A trend has been established recently, to build digital libraries using open source software. Use of Open Source Software came in libraries during 1998[3] and large number of library professionals started making use of Open Source Software.

Open Source Software are available for all types of applications that are carried out in libraries. Under Open Source license many applications including applications for document delivery, Z39.50, MARC record readers and writers, integrated library system, systems to read and write bibliographies, and digital library software are being made available.

There are many commercial digital library software available in the market for creating digital libraries such as CONTENTdm, Digital Commons, Digitools, Hyperion, Vital, Open Repository, Bepress, as well as open source software available for creating digital libraries. These software provide number of features and services to the end users. In this environment it is necessary to understand from users point of view, whether a particular software satisfy user's requirements?

Assessment of the software which is used to create digital library can be carried out extensively to know whether the software is performing its all the functions that are required by users. How the software performs and what are the issues that needs to be considered while deciding to use any software for building digital libraries. Librarian have to develop evaluation activities and strategies in digital environment for assessing various parts of the digital library software.

Today there are more than a dozen software available on Internet under Open Source Li-

cense terms and conditions for building digital libraries. The main purpose behind using Open Source Software is no cost is involved in using the software. With the use of Open Source Software libraries can manage to make their contents available on web without investing any money. There are many such open source based digital library efforts, projects and implementations, from all over the world.

Hence to know the current trend and to understand how future digital libraries would be, it is necessary to understand how these Open Source Digital Library Software function (OSS-DL), what features are supported by each software, what are their future plans, whether software satisfy minimum requirements of the users, whether software satisfy minimum functions of a digital library, whether software satisfy minimum standard support, what are the unique features of each software, how is the installation of each software etc.

Each software will support specific activities in specific contexts and hence they need to be evaluated to determine how useful, usable and effective they are as compared to others[4]. Many such questions needs to be answered. To understand these questions and to get answers, it is necessary to evaluate them on the defined evaluation criteria.

The present chapter hence deals with the basics of Evaluation as it is the basic purpose of the present study and lists detailed list of evaluation criteria which are defined for evaluating each software covering broad and narrow aspect of all the functions that are generally carried out by the digital library software.

3.2 What is Evaluation?

Evaluation is defined as "the systematic process of determining the merit, value and worth of something"[5] It is a general term that includes various aspects of performance measurement and assessment. Activities include laboratory experiments, regional, national and international surveys or quasi-experiments, time series analysis, online monitoring of user-system interactions, observation of use, and other forms of data collection[6]. Evaluation has many connotations ranging from highly focused and well-defined product testing to the highest form of cognitive reflection.

A system is evaluated to ascertain the level of its performance or its value. Evaluation is executed according to a specific procedure. Evaluation study is an objective study based

upon observations not opinion. ISO 14598 distinguishes four activities during the process of evaluation: analysis, specification, design and execution.

Digital libraries similarly can be judged by their effectiveness (how well does a system or any of its parts perform the roles or tasks for which it was designed?) and efficiency (at what cost?)[7] Evaluation can be performed at different levels, involving different objectives and related criteria.

3.3 Types of Evaluation

There are different types of evaluation depending on the object to be evaluated. Four types of evaluation are important with respect to Digital Libraries[8]:

3.3.1 Formative evaluation

Formative evaluation is a method of judging the worth of a program while the program activities are forming or happening. Formative evaluation begins at the initial stages of a development project to establish baselines on current operations, set goals, and determine desired outcomes.

3.3.2 Summative evaluation

Summative evaluation is a method of judging the worth of a program at the end of the program activities. The focus here is on the outcome. It helps to determine if the intended goals of the program were met or not.

3.3.3 Iterative evaluation

Iterative evaluation takes place throughout a project, beginning in the earliest design and development stages. Interim stages of design are assessed in comparison to design goals and desired outcomes, and the results inform the next stages of design. Iterative approaches encourage designers to set measurable goals at the beginning of a project and provide opportunities to re-assess goals throughout the development process.

3.3.4 Comparative evaluation

Comparative evaluation requires standardized measures that can be compared across systems. Communities can identify and validate measures. If such measures are implemented in a consistent manner, they enable comparisons between systems. Test beds are another way to compare measures and to compare performance of different functions and algorithms.

3.3.5 Analytical Evaluation

Analytical evaluation makes the analyst think deeply about the design and about users, which can yield insights and long-term learning that inform future design decisions.

3.3.6 Cognitive Walk-through

Cognitive walk-through is a review technique, in which evaluators play a role of the user and "walk through" the interface in an attempt to complete certain information seeking tasks. Evaluators attempt to simulate the cognitive activities of the user and predict how he or she will react to different interfaces. The cognitive walk-through method proved to be very valuable for identifying ways to reduce clutter, reduce the number of links and make links more visible, and reduce the amount of text on the web site. It also identified problems with terminology and questioned whether certain color schemes might present difficulties for color-blind people.

3.3.7 Heuristic Evaluation

Heuristic evaluation is another usability inspection method that evaluates the design of a user interface based on established usability principles. HE is a checklist-based approach to assessing the usability of an interactive system. In the original version of this technique, the analyst (or team of analysts) works through every page or screen of a system, asking team questions about that system. Digital library research, planning or deployment of digital libraries all can benefit from evaluation whether formative, summative, iterative or comparative. There are couple of studies which have been carried out so far in Heuristic

evaluation area where user interface of digital libraries is studied by couple of users and have arised to different conclusions such as how the user interface should be of a particular digital library.

3.4 Evaluation of Digital Libraries (DL)

Digital libraries "give us opportunities we never had with traditional libraries or even with the web"[9]. With the growing importance of the provision of online services and resources, there is a need to establish methodologies by which it is possible to evaluate and measure the performance of digital libraries, the information they contain, and the services they deliver against set of standards.

Evaluation in context to digital libraries means finding how the system performs, how the system functions, how the user interface of the digital library is, how are the search features provided by the digital libraries. Evaluation can be done on different levels, in relation to different choices of objectives, using a variety of methods. Thus evaluation of a digital library may serve many purposes ranging from understanding basic phenomena (e.g., human information-seeking behavior) to assessing the effectiveness of a specific design. Digital libraries are complex systems, hence while evaluating digital libraries several questions needs to be answered such as

- What should we evaluate?,
- For what purpose do we evaluate,
- Who should evaluate?,
- At what level one should evaluate,
- Upon what criteria do we evaluate?

Considering the above aspects, there are very few studies which are carried out so far in the evaluation of digital libraries as well as evaluation of the open source software available for creating digital libraries. The literature that reports evaluation, is not large, because there are not many evaluation efforts around the world. The proceedings of the two major

conferences on digital libraries, the European Conference on Digital Libraries (ECDL) and the joint ACM/ IEEE Conference on Digital Libraries (JCDL) contains not more than 5% of the papers or posters that pertain to evaluation of whatever aspects of digital libraries[10].

Digital libraries have brought out several issues such as technical complexities, variety of contents and varied users using digital libraries. Due to these complexities there is a need to carry out evaluation of digital libraries. Due to lack of tools and methodologies, DL evaluation efforts have stopped behind. Still the comprehensive state-of-the-art of digital library evaluation is not sufficiently developed[11].

Since 1950's evaluation of IR systems used 'relevance' as the basic criteria for evaluation[12]. Libraries used variety of other standardized criteria for evaluation of different components of library, such as a collection, services etc. On the other hand for digital library evaluation methods have not yet developed and there are yet no standardized criteria available for digital library evaluation.

Although published research on digital libraries has increased, it mostly focuses on technical issues and digital library use patterns. In order to improve the design of existing and future digital libraries, there is a need to identify what criteria need to be applied in the evaluation process. So far little has been done on the identification of evaluation criteria. Moreover, these evaluation criteria are identified by researchers, not by users. One more reason for less evaluation studies is the lack of expertise, the lack of readily available metrics and test beds, and the lack of comparative data on uses, users, and usability. Most evaluation studies carried out today address only usability studies.

The most notable evaluation was done by the Alexandria Digital Library Project (ADL) at the University of California, Santa Barbara. The study used several methods to evaluate user views including: online surveys, ethnographic studies, focus groups, and user comments. The approach involved a series of user studies, involving different user communities and concentrating on different design features as related to their usability and functionality. User logs were also studied as part of the evaluation. The evaluation concentrated on users and their interactions through the interface, with usability and functionality as the main criteria[13].

3.5 Evaluation of Open Source Digital Library Software (OSS-DL)

The evaluation of digital libraries is yet in the primitive stage and the evaluation of Open Source Digital Library Software is a very new concept. There are very few studies carried out so far on evaluation of open source digital library software due to the fact that these software are very complex and their functionalities are different. Each software will have different backend database, different web server configuration, each software would have been written in different programming language etc. Hence evaluation study will be complex as well as difficult task to carry out[14].

Open source software are making their place in libraries and libraries all over the world are trying to find which software satisfies which functions, features, how to install the software, how is the performance of the software etc. which software is useful for a type of collection or the organization needs. Librarians from developing countries especially will be benefitted if more number of evaluation studies are available and if are carried out by the Librarians. The librarians will have detailed knowledge of the evaluation studies if the studies are conducted systematically covering all aspects of digital libraries.

In the present study, the evaluation criteria for evaluating digital libraries presented here are defined on the basis of an extensive literature search carried out on Internet and on online full text databases for searching relevant articles on the evaluation part and after going through the literature and available sources an extensive list of evaluation criteria which can be used to evaluate the available Open Source Digital Library Software was prepared. The criteria defined here form the basis of earlier studies.

The majority of research on digital library evaluation focuses on how users use a digital library, essentially usability studies, to either recommend design principles or improve the existing user interface and search features but very few studies have been carried out on evaluating Open Source Digital Library Software covering identifying their functionalities, performance etc. The details of earlier studies on OSS-DL is already mentioned in this thesis in Chapter 1.

One of the objectives of the study is to identify software available under open source license for creating digital libraries then to define an extensive list of evaluation criteria while

evaluating digital libraries and evaluate all these software against the criteria identified as there are no standard evaluation criteria which are yet defined for evaluating OSS-DL.

The defined evaluation criteria have been used as a checklist while evaluating each software. Checklist is defined as "a specialized set of questions designed to help checkers find more defects"; whereas another largely accepted definition is that a checklist is a "list of questions that clarify and standardize the inspection process and guide inspectors to pay attention to the right things"[15]

The evaluation criteria defined here are grouped under broad functions of the digital library such as content acquisition, content management, metadata submission, information search & retrieval, authentication and authorization, Ease of Deployment, Usability, Copyright Issues, Digital Preservation etc.

These broad criteria defined here are defined on the basis of different functions carried out in digital libraries as mentioned in the following figure (Fig.3.1). There are several activities which are carried out in digital libraries. These activities are more or less same as in traditional library activities. In traditional library, we select material, acquire it, organize it, catalogue it and then make available to users through OPAC along with preservation aspect. Similarly in digital environment, digital objects are selected first then they are accessioned, catalogued and made available to the users by different value added web services. On the basis of this workflow, evaluation criteria are grouped under broad and narrow subject categories.

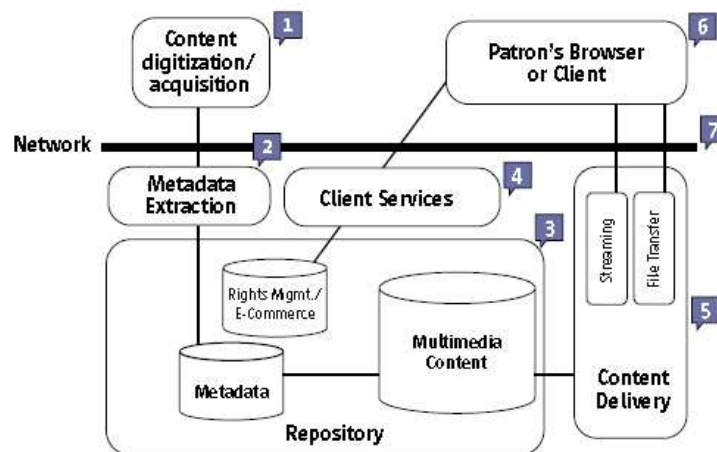


Figure 3.1: Digital Library Components

3.6 Evaluation Criteria

The following sections define different evaluation criteria for evaluating OSS-DL.

3.7 Content Acquisition

Digital libraries offer new opportunities to provide access to diverse resources beyond those which are held in traditional library environment. Content acquisition is an important part of any digital library. Like traditional libraries, while building digital libraries, only those contents are acquired in any digital library which are pertinent to the individual libraries information needs. Not all digital documents are added in any digital collection.

Content is thus most important aspect of any digital library and quality of content is a primary factor that sets any digital library apart from the majority of the material available freely on the web.

Digital libraries tend to serve a particular community/ communities. Each digital library has its own collection development policy and procedures. Digital libraries carry out all tasks that are similar to traditional libraries and they go beyond traditional library activities as they can provide more value added services than traditional libraries.

Every digital library have selection policies. Selection process is an intellectual judgment which includes what to include and save in any digital library and what to exclude. In general, selection criteria include an approval of the content -its subject and discipline – in relation to the collection goals of the digital archives.

Avoiding redundancy of information is important for effective selection. Selectors need to know before accepting any digital object in the digital library or repository, whether a copy of the digital object is already available in the digital library or not? It is also necessary to have rich understanding of the software and hardware dependencies of digital information objects before adding any document in the digital repository.

Once an information object is identified for inclusion in digital archives, it needs to be accessioned, catalogued and preserved in the digital archives. The accession process involves

both describing and cataloging selected objects, including their provenance and context information, and securing them for storage and access for future.

Content acquisition is an important aspect in any digital repository/ digital library/ institutional repository. Hence while evaluating digital library software, it is necessary to know whether the software supports the following content acquisition features:

- 3.7.1 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.?)
- 3.7.2 Does the software have capability to define the domain/ scope of Digital Library for whom it is intended for?
- 3.7.3 If yes, what level of the users it is intended for?
- 3.7.4 Which digital document file formats does the system support? (proprietary as well as open source) for e.g.:-
 - (a) Text Formats (ASCII, UNICODE, RTF)
 - (b) Image Formats (TIFF, GIF, JPEG, PNG)
 - (c) Presentation Formats (Adobe Post Script, Adobe PDF)
 - (d) Structured Formats (HTML, SGML, XML)
 - (e) Audio and video formats (Wave, Real, MP3, AVI, MPEG)
- 3.7.5 Does the system carry out duplicate checking while uploading the data? What happens if the same digital document is submitted multiple times?
- 3.7.6 How does the system allow to upload the digital objects, whether normal digital object or compressed digital objects?
- 3.7.7 Does the software support uploading digital objects from existing URL?
- 3.7.8 Does the system support distributed/ community based acquisition?
- 3.7.9 Does the software support metadata/ content import? In which file format data is required for importing?
- 3.7.10 Does system support importing of bulk metadata/ content of digital objects?

- 3.7.11 Does the software support import of automatic metadata for the digital objects that are added into the repository?
- 3.7.12 Does the software support metadata/ content export? In which file format data is exported?
- 3.7.13 Does the system support export of bulk metadata/ content of digital objects?
- 3.7.14 Does the system allow adding past versions of the digital document files?
- 3.7.15 Does the software support its own accession number for each document that gets added into the repository?
- 3.7.16 Timeliness is an important variable in many information seeking activities. Maintaining DL is more difficult than just building collections hence site must be "weeded" out regularly as material added in the digital libraries will lose value over time and documents must be needed to be withdrawn or discarded regularly. DL software tools must support as much of the time-stamping as possible and alert staff and users when information in the digital library should be updated or removed. Hence it is necessary to know whether software support weedout policy?
- 3.7.17 Does the system support to delete items from the collection/ repository, move digital objects from one collection to other collection, cross-list digital objects across different collections?
- 3.7.18 Persistent Identifiers in Digital Libraries: The digital object should have unique identification number over web irrespective of their location. There are different persistent identifier schemes which assign unique numbers to digital documents that get added into the digital library. The following are some of the schemes[16]:
- (a) Uniform Resource Name (URN)
 - (b) Internationalized Resource Identifier (IRI)
 - (c) Persistent Uniform Resource Locator (PURL)
 - (d) Digital Object Identifier (DOI)
 - (e) Archival Resource Key (ARK)
 - (f) CNRI Handle System <http://hdl.handle.net>

(g) OAI Identifier

In digital libraries evaluation, it is necessary to know which of the above persistent identifier scheme does the system support?

- 3.7.19 Does the system's persistent URL have capability to handle an object's change in location and state in future with same identifier number?

3.8 Content Management

This requirement is related to the ease with which content is created, submitted, organised and reviewed within the digital library system by different users. The digital library software should be capable enough to handle variety of users performing different tasks. The software should support to create proper workflow for adding any digital document into the repository. Within this criterion it is necessary to know how submission process is handled in each DL software:

- 3.8.1 Does the system have proper workflow in place which handles different submission processes?

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

3.8.2.1 System allows to create overall system administrator

3.8.2.2 System allows to submit documents by users via web

3.8.2.3 System allows to set up submission rules

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

3.8.2.5 Allows multiple collections within same installation of the system

3.8.2.6 Home page for each collection/ community

3.8.2.7 Allows to set different policies for different collections

3.8.2.8 Allows to set different authorization policies for different collection/ items added into the repository

- 3.8.2.9 Allow to edit metadata submitted by users
 - 3.8.2.10 Sends email notification to users/ submitters/ authors
 - 3.8.2.11 Sends email notification to metadata reviewers
 - 3.8.2.12 Sends email notification to reviewer
 - 3.8.2.13 Sends email notification to administrators
 - 3.8.2.14 Allow users to review completed content
 - 3.8.2.15 Allow users to review uncompleted content
 - 3.8.2.16 Allow content administrator to review submissions
 - 3.8.2.17 View pending content administration tasks
- 3.8.3 Does the system support to upload documents into the repository through Librarian's interface as well?
- 3.8.4 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?
- 3.8.5 Are the supported file formats well documented in the system? Can new file formats be added or removed by the end user ?
- 3.8.6 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?
- 3.8.7 Does the repository software maintain audit logs that identify by whom and when documents are submitted into the repository?
- 3.8.8 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?
- 3.8.9 Does the system support digital objects to be grouped into organized structure (i.e. linear/ hierarchical/ graphical structure)?
- 3.8.10 Does the system support to generate authority files?
- 3.8.11 Does the system allow digital object to be a member of multiple collections i. e. content mapping?

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

3.8.13 Does the system support to show strength of each collection?

3.9 Metadata Submission and Support

The key purpose of metadata is to facilitate and improve the retrieval of information. [17]. Metadata plays an important role in any digital library software. In digital libraries, to support navigation and management of different types of digital objects additional categories of metadata have emerged. Different subject domains will have different metadata schemas. Dublin Core is becoming a defacto standard for metadata entry in any digital library software. The digital library software should have capabilities to support different metadata schemas for adding variety of digital documents. While evaluating digital library software it is necessary to know which metadata schema does the software support along with other features:

3.9.1 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?

3.9.2 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, how the digital document is created, what is file size, in what file format digital document is in, what is needed to view the digital document) for each digital object that gets added into the repository?

3.9.3 Does the software support to add structural metadata (such as description, owner, data type, date deposited, version number, date 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?

3.9.4 Does the software support adding/ editing/ deleting metadata fields?

3.9.5 Are help messages given in each metadata field while entering data?

- 3.9.6 Does the metadata entry has one field mandatory? What are the mandatory fields in each software?
- 3.9.7 Does the software support Unicode character set for metadata entry?
- 3.9.8 Does the software allow to create and manage a variety of different metadata schemas to meet different needs across the subject disciplines ?
- 3.9.9 Can the metadata fields be customized?
- 3.9.10 What type of Metadata Schemas are supported by the software by default?
- (a) DC (Dublin Core)
 - (b) METS (Metadata Encoding and Transmission Standard)
 - (c) MODS (Metadata Object Description Schema)
 - (d) EAD (Encoded Archival Description)
 - (e) TEI (Text Encoding Initiative)
 - (f) Learning Object Metadata (LOM)
 - (g) CDWA (Description of Works of Arts)
 - (h) MARC 21
 - (i) VRA Core Categories
 - (j) Multimedia Metadata Initiative
 - MPEG-21
 - MPEG-7
 - (k) Metadata for datasets: FGDC (Federal Geographic Data Committee)
 - (l) Geospatial Metadata: (Content Standard for Digital Geospatial Metadata)
 - (m) Preservation Metadata Initiative (PREMIS)
 - (n) SCORM
- 3.9.11 Does the software have default metadata entry templates?
- 3.9.12 Are the metadata formats documented?

- 3.9.13 Does the software support different interface for metadata entry and whether customized data gets added into the system?
- 3.9.14 How does the software verify that submitted objects and metadata are correct?
- 3.9.15 Are there automated checks of the metadata, such as to verify that a date entered into a field really is a date string?
- 3.9.16 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?
- 3.9.17 Does the software support real time updating and indexing of accepted contents?
- 3.9.18 Does the software allow for metadata extensibility and complexity?
- 3.9.19 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?
- 3.9.20 Does the repository software support provenance metadata? What is included in the provenance?
- 3.9.21 Does the software support metadata crosswalk?
- 3.9.22 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?
- 3.9.23 Does software have ontology support?
- 3.9.24 How does the repository verify file types, for example, if JPEG images are submitted to the repository, does the software verify that it really is a valid JPEG, or does it just blindly take the submitters word?
- 3.9.25 While exporting metadata whether software supports checksums?
- 3.9.26 RAP Support: When digital object is added into the repository the Interactions such as depositing digital objects or accessing digital objects in repositories is accomplished using a repository access protocol (RAP), which all repositories generally

support. Kahn and Wilensky (1995) specified a framework for naming digital objects and accessing them through a machine interface. This Repository Access Protocol (RAP) provides an abstract model for the services needed in order to add, modify or delete records stored in a digital library[18]. A digital object stored in a repository, and whose handle has been registered with the handle server system, is called a registered digital object. A simple repository access protocol (RAP) is supported by each repository. RAP allows for accessing a stored digital object or its metadata by specifying its handle, a service request type and additional parameters. Each repository must support a simple protocol to allow to deposit and access digital objects or information about digital objects from that repository. This is called Repository Access Protocol. It is necessary to know whether the software support RAP protocol?

3.10 Classification

Classification can be defined as the act of creating a logical organization of terms or classes. Classification is a powerful intermediation service useful for discriminating information resources at both broad and specific levels. Classification and indexing schemes are used to collect related content into groups that are intuitive to a user.

Digital libraries require both broad categorization to identify relevant and specific categorization to select specific resources within libraries. Digital libraries should support many levels of representation to support cross-collection and within-collection information seeking. In addition, digital storage enables the development of machine-based classifications increasing the efficiency of classification activities.

- 3.10.1 Does the system allow digital objects to be grouped into classification/ subjects?
- 3.10.2 Which classification system does the software support?
- 3.10.3 Does it allow to group documents as per the class number?
- 3.10.4 It is necessary to know whether system supports hierarchical browsing of collections through class number.

3.11 Information Search & Retrieval

Information search and retrieval is an important part in any digital library software. Search enables quick retrieval of information. Search services help users to select relevant information from digital library. Digital library's search service provides fast access to exact information which user is looking for. The success of a search service in any digital library relies on the implementation of a powerful retrieval engine and a flexible user interface as well as a good metadata support. The search interface allow users to do "across database" searching without having to modify a query[19].

Search service also covers searching beyond text to multiple media formats, including images, sound and video. The retrieval formats should be flexible and should provide users to manipulate the search process and results by reviewing search history, adjusting search strategies, editing and sorting search results and choosing preferable delivery formats. Users should also be able get statistical analysis of the searches they have carried out (e.g. Tracking most frequent queries or terms in queries used).

Many digital libraries provide different search options and users can conduct searches on various fields as well as provide facilities for federated search or search across a number of digital libraries. Most digital libraries offer search by Boolean operator, keyword, phrase and field searches.

In the case of information retrieval, evaluation is often focused on the effectiveness of a result set in a specific search.

Browsing and searching are two major paradigms for exploring digital libraries. Boolean, proximity and truncation search facilities are commonly available in digital libraries. They are often provided by DLs as separate services. Searching is popular because of its ability to identify information quickly. On the other hand, browsing is useful when appropriate search keywords are unavailable to users (e.g. a user may not be certain of what he/ she is looking for until available options are presented during browsing).

While evaluating the search and retrieval facility of the selected software, for the study the following criteria should be considered:

3.11.1 Browsing

The browse feature is important for retrieval of information that is not already known in part. Browsing enables people to look through a digital library and discover things that they had no previous knowledge of.

Users should be able to browse digital objects by

3.11.1.1 Author/ Creator/ Contributor

3.11.1.2 Title of the document/ Title of the book/ Title of the article etc.

3.11.1.3 Issue Date/ Date of Publication

3.11.1.4 Collection

3.11.1.5 Communities

3.11.1.6 Subject browsing

3.11.1.7 Publisher wise browsing

3.11.1.8 Table of Contents browsing

3.11.1.9 Multi-Dimensional Browsing: Multi-dimensional browsing allow users to move along any of the navigational dimensions, or a combination thereof. By navigational dimension means a hierarchical structure used to browse digital objects. Typical DLs provide a directory-style browsing interface (as in Yahoo! or Open Directory) with levels in the hierarchy displayed as click able category names and DL items in that category. In ETANA-DL, a user can browse through three dimensions such as space, object and time[20]. While evaluating the existing DL software it will be seen whether multi-dimensional browsing is supported or not.

3.11.2 Searching

Searching in any digital library is one of the important aspect. Hence it is necessary to know what type of search features are supported by each digital library software. While evaluating the software it is necessary to do functional testing of the software i.e. determining the

extent to which a digital library, in whole or in part, is able to perform desired operations (e.g., basic search, search via multiple languages etc.). The digital library software should support extensive search and retrieval. Boolean operators - AND, OR and NOT - are used to combine words or phrases in a search expression. The following are some of the criteria which are identified while evaluating searching feature of each software:

- 3.11.2.1 Searching content via different metadata fields such as author, title, subject, publisher etc.
- 3.11.2.2 Full text searching
- 3.11.2.3 Boolean (AND, OR, NOT) searching
- 3.11.2.4 Basic search
- 3.11.2.5 Advanced search
- 3.11.2.6 Does the system allow to search the database putting boolean operators such as !, |, &, ~, +, -, etc. as well as "must contain", "should contain", "may contain", "not contain", as operators in place of AND, OR, and NOT Boolean operators respectively.
- 3.11.2.7 Truncation/ Wild card searching: Truncation searches allow users to search for different word variants with a single search expression where the truncation symbol stands for one or more characters in the search term. There are three types of truncation: left truncation, right truncation and middle truncation[21]. Right truncation matches any number of characters at the end of the word, while left truncation starts with any number of characters followed by the search word. Middle truncation matches words starting and ending characters with any intervening characters Middle truncation is also referred as 'embedded truncation'. Various operators, such as '*', '#', '?', are used for truncation.
- 3.11.2.8 Exact words/ phrases searching: A query may be entered in quotes to search for an exact match of the phrase.
- 3.11.2.9 Proximity searching: A proximity operator searches both words in a field or text with a fixed number of intervening word(s) between them. The proximity search operators used are 'ADJ', 'NEAR' and 'WITH'.

- 3.11.2.10 Stemming: Stemming searches look for other grammatical forms of the search terms. For example, a stemming search on the 'fly' would also find 'flies'.
- 3.11.2.11 Fuzzy searching: Fuzzy search expands the search by generating similarly spelled words to the specified word or phrase.
- 3.11.2.12 Phonetic searching: Phonetic search looks for a word that sounds like the word one is searching for and begins with the same letter.
- 3.11.2.13 Case sensitive or case insensitive while searching: Whether software supports case sensitive . Searches especially useful searching for acronyms.
- 3.11.2.14 System should support term weighting: In a search expression, users can specify that some terms should count more than others. For example, if a user is looking for documents about both 'Apple' and 'Pear', he or she might want to give preference to the word 'Apple' over the word 'Pear'. Term weighting allow the retrieval of documents with higher weightage.
- 3.11.2.15 System should support search history option , reuse of query, query save option.
- 3.11.2.16 Boosting the term
- 3.11.2.17 Range searching
- 3.11.2.18 Use of thesaurus or permuted index for searching
- 3.11.2.19 Expand search
- 3.11.2.20 Lateral searching
- 3.11.2.21 Search support for special collections
- 3.11.2.22 Searching via table of content and classification numbers
- 3.11.2.23 Ability to browse subject/ author authority files
- 3.11.2.24 PIC variant searching (plurals, international spelling variants, and compound words)
- 3.11.2.25 Support multilingual search and retrieval
- 3.11.2.26 Allow refinement of searches with qualifiers

- 3.11.2.27 Ability to group: prior to search, user can delimit pending results' listings, and group results by item characteristics (type, format)
- 3.11.2.28 Spell checker support
- 3.11.2.29 Refine searches: Once results are retrieved, users have the ability to focus and reduce number of items retrieved

3.11.3 Sorting

The system should support sorting results by

- 3.11.3.1 Author
 - 3.11.3.2 Title
 - 3.11.3.3 Issue Date
 - 3.11.3.4 Relevance
-
- 3.11.4 Which search engine toolkit is used for searching the database?
 - 3.11.5 How is the recall and precision ratio while searching the documents?
 - 3.11.6 How is the speed of the retrieval of the documents?
 - 3.11.7 Does the system provide help for the users while searching digital documents?
 - 3.11.8 How are the results displayed to the users? Whether users can download contents after searching? How are results displayed such as metadata, filetype, file size?
 - 3.11.9 While viewing the results does the system provide access control?
 - 3.11.10 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?
 - 3.11.11 Does the system support advanced features such as my binders, my searchers, my basket, etc.

- 3.11.12 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)?
- 3.11.13 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)?

3.12 Access Control, Privacy and Management

In any digital library software it is necessary to know how users are administered in the software. The software should have capability to store all details about users their current state and their roles to play in the digital library. A good digital library software should have all features dealing with administration of users. While evaluating access control it is necessary to see whether software support following features:-

- 3.12.1 How users are managed in a digital library software?
- 3.12.2 Does the system support to add/ edit/ delete user profiles?
- 3.12.3 Does the system support to define different roles for different users?
- 3.12.4 Does the system keep detailed information of each user that registers in the repository such as name, email, phone number, address etc.?
- 3.12.5 How the permissions are given to users who have registered into the repository ?
- 3.12.6 Does the system have provision to identify active and inactive members and take necessary action against inactive members?
- 3.12.7 User Roles
- 3.12.7.1 Does system allow to create different groups for different users
 - 3.12.7.2 Does system allow to create different groups of users

- 3.12.7.3 Does system allow definition of different user groups
- 3.12.7.4 Does system limit access by role
- 3.12.7.5 Does system allow collection to be customized for each role/ group
- 3.12.8 Does the system allow access limits at different levels?
 - (a) File/ object level
 - (b) Collection level
 - (c) User/ Group level
- 3.12.9 How passwords are administered in the system?
- 3.12.10 Does the system assign passwords for each user?
- 3.12.11 Does the system allow users to select passwords?
- 3.12.12 Does the system have mechanism to retrieve forgotten password?
- 3.12.13 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.
- 3.12.14 Does the system have any mechanism to provide access to the digital library via
 - (a) IP source address filtering
 - (b) Proxy filtering: Proxy server is a server that acts as an intermediary between a workstation user and the Internet so that the organization ensure security, administrative control and caching service.
 - (c) Credential-based access (i.e. Only certified users are allowed)
- 3.12.15 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?
- 3.12.16 Does the system support access to the collection for a specific period?

3.13 Authentication and Authorization

Authentication and authorization is an important aspect in any digital library software. The system should be able to protect information against malicious use and corruption and must ensure privacy of its digital contents[22].

The system should also help to provide open access to information so that vendors and information producers can add/ update information and services any time. Since DL is a new and emerging area of research, there has been very little prior work that addresses security for DL.

Establishing access controls involves setting terms and conditions for authorized use by authorized users or classes of users.

Authentication, which may make use of cryptography, provides verification that a digital object is what it purports to be and contains the contents that the author/ creator or publisher originally intended.

- 3.13.1 Does the system support to provide setting different authorization policies?
- 3.13.2 System should support restricting access to contents through password authentication, IP filtering and proxy filtering. Personal authentication technology distinguishes individuals and confirms that a given individual is the person who has registered in the system earlier.
- 3.13.3 System should be able to support access based on browsing/ viewing metadata, certain records, full text item or collection through different authorizations with different privileges for different communities or individuals.
- 3.13.4 System should also be able to allow to set different roles for different members for different items/ collections.
- 3.13.5 Does the system support a super user/ admin user who is overall manager/ administrator of the software?
- 3.13.6 Does the system use authorization to support different roles within the system (for eg submit user, metadata editor, reviewer, approver, overall administrator etc.)

- 3.13.7 Does the system use an external authentication mechanism such as LDAP (Lightweight Directory Access Protocol). Through, LDAP network users on the existing system can directly be made DL members.
- 3.13.8 Does the system display only those pages that are pertinent to each user that he/ she is able to see?
- 3.13.9 Does the system support the feature that groups are assigned to one or more roles?
- 3.13.10 Does the system support the feature that users are assigned to one or more groups or collections?
- 3.13.11 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
 - (b) Administrator Role: must be able to add/ delete/ create users and groups
 - (c) Contributor Role: must be able to submit metadata and content, in addition to general users' role
 - (d) General User
 - (e) Submitter
 - (f) Metadata Reviewer
 - (g) Max Role: must be able to do all activities within the system
- 3.13.12 Does the system use authorization to verify users with campus security systems (NetID)?

3.14 Interoperability

The digital library needs to be interoperable with other systems to which it is connected. This allows each system to evolve independently without sacrificing their ability to communicate with each other. The repository software should support two basic interoperability protocols, namely, Z39.50 and OAI-PMH (Open Archives Initiative Protocol for Metadata Harvesting).

- 3.14.1.1 It is necessary to know whether the software supports Z39.50 protocol and OAI-PMH protocol?
- 3.14.1.2 Which harvesting software is used by the Digital Library Software for metadata harvesting?
- 3.14.1.3 SRU/ SRW: The SRW/ SRU services (Search/ Retrieve Web and Search/ Retrieve URL, respectively) define a method for interacting with and retrieving information from remote databases.

SRU/ SRW are standard search protocols for Internet search queries, utilizing CQL (Common Query Language), a standard query syntax for representing queries. SRU/ SRW are modern version of Z39.50. SRU/ SRW, are standards for information retrieval. SRU/ SRW display results in XML format. SRU stands for Search/Retrieval via URL and SRW stands for Search/Retrieval via Web Service. Z39.50 allow to search by opening a telnet connection. In SRU/ SRW, instead of opening a telnet connection to remote computers, SRU/ SRW is usually communicated via the web (HTTP). Instead of a dialog of commands sent back and forth between two computers resulting in a stream of relatively unstructured data, SRW/ U sends a single URL to a remote computer and the remote computer sends back an XML stream. Because the XML is usually returned to a Web browser, and since Web browsers are able to "transform" XML into HTML with a built-in technology called XSLT, the search results are displayed to the user. SRU/ W is much simpler than Z39.50[23]. SRW/ U is used for searching indexes (, and to some degree databases as well). It is possible to use it to search just about any list of stuff such as Books, Journals, Articles, Microforms, Reference sources, Names, Addresses, Telephone numbers, Internet resources, etc.

- 3.14.1.4 OpenURL: The OpenURL resolver helps end users to directly open full text of the documents through citation links. If a user is reading an article and if he/ she refers to the citations of that article and directly want to open the article from citations software should have capability to open that article. This is called as OpenURL. While evaluating digital library software it is necessary to know whether software support OpenURL?

3.15 Ease of Deployment of Each Software

Though open source software are becoming popular, the installation of many open source software are still not very easy. Each installation demands installing allied software, configuring web servers, configuring backend databases, setting environment variables etc. and many a times installation is very difficult. Hence it was necessary to know what is installation procedure of each selected software? While evaluating the installation aspect it is necessary to know other aspects as well.

The following points will be considered while evaluating installation of each software:

3.15.1 Software Installation

3.15.1.1 Packaging and installation steps

3.15.1.2 Automatic installation script

3.15.1.3 Disk space used

3.15.1.4 Time required to install each software

3.15.1.5 Documentation help available and useful for installation

3.15.1.6 Environments needed to set for installation of each software

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

3.15.1.8 Programming languages used

3.15.1.9 Database supported at back end

3.15.1.10 Web Server used

3.15.1.11 Java servlet engine needed

3.15.2 System Support/ Maintenance

- 3.15.2.1 Documentation/ manuals
- 3.15.2.2 Mailing lists/ discussion forums
- 3.15.2.3 Wiki pages
- 3.15.2.4 Help desk support
- 3.15.2.5 Ease of system administration (ability to configure for different users)

3.15.3 Hardware/ Software Requirements

- 3.15.3.1 What are the minimum hardware requirements for the repository software?
- 3.15.3.2 How actively is the software maintained by its owners or contributors?
- 3.15.3.3 Does the software have procedures in place to monitor and receive notifications when hardware technology changes are needed?
- 3.15.3.4 Does the software creators have a process to stay current with the latest operating system security fixes?

3.15.4 Security

- 3.15.4.1 Does the system have security methods such as
 - Data encryption (supports encryption of data while transmitting the content) Encryption is the process of transforming information (referred to as plain text) to make it unreadable to anyone except those possessing special knowledge, usually referred to as a key[24]
 - Digital signatures
- 3.15.4.2 Server security (Does software have secure database connection?)
- 3.15.4.3 Ability to restrict access repository at item level (For eg to view metadata but not content)

3.15.5 System Level Security

- 3.15.5.1 Fixity - such as checksums or digests
- 3.15.5.2 Management of users accounts and rights to specified locations within the DL
- 3.15.5.3 Logging and auditing events
- 3.15.5.4 Does the software support any security if OS does not have any firewall implemented?

3.15.6 General features related to technical aspect

- 3.15.6.1 Does the software have any back end maintenance?
- 3.15.6.2 Does the software have RSS support?
- 3.15.6.3 Does the software have easy mechanisms to upgrade the software from old versions to new versions?
- 3.15.6.4 Does the software support migration of data from one repository software to another repository software?
- 3.15.6.5 Does the system support any ranking algorithms for the documents added in the repository?
- 3.15.6.6 Are there any known issues/ bugs in the system?
- 3.15.6.7 Whether a novice user can easily publish content?
- 3.15.6.8 How often user needs to use database schema for making any change?
- 3.15.6.9 Whether user can reorganize database?
- 3.15.6.10 Does system support automated content acquisition, harvesting and automatic meta-data generation, automatic subject indexing/ classification?
- 3.15.6.11 Does the system support to store metadata records separately from the actual contents?
- 3.15.6.12 Does the software support storing varied file formats?

- 3.15.6.13 Does the system keep the original file's name, size and created date?
- 3.15.6.14 While using the software if there are any errors occur, does the system allow to submit the error report?
- 3.15.6.15 How many developers are working for the software?
- 3.15.6.16 How many users are using the software/ strength of community?
- 3.15.6.17 Does the software provide its history on the web page for evidence of sustainability and vitality?
- 3.15.6.18 Total number of downloads of the software?
- 3.15.6.19 Is the software really under Open Source License terms and conditions?
- 3.15.6.20 Are there any costs involved for using OSS-DL?
- 3.15.6.21 Does the repository software have a clear policy regarding software upgrades?
- 3.15.6.22 Whether the software has a long-term leadership and consulting services?
- 3.15.6.23 After releasing the software whether it was tested and whether it has integrated new features or not?
- 3.15.6.24 How is the mailing list used by the users?
- 3.15.6.25 Whether software has conducted any workshops, training sessions on local, regional, national and international levels?
- 3.15.6.26 Whether software is built according to open standards?

3.15.7 Architecture of the Software

The Architecture concept refers to the Digital Library System entity and represents a mapping of the functionality and content offered by a Digital Library onto hardware and software components. There are two primary reasons for having Architecture as a core concept: (i) Digital Libraries are often assumed to be among the most complex and advanced forms of information systems; and (ii) interoperability across Digital Libraries is recognized as a substantial research challenge.

- 3.15.7.1 How is the architecture and design of the software whether it is simple or complex?
- 3.15.7.2 Does the architecture support 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).
- 3.15.7.3 Does system support grid architecture? How grids can help to disaster recovery of data?

3.15.8 Scalability

Many digital libraries will have collections in terms of gigabytes, terabytes, petabytes etc. Millions of items will be added in different repositories. Considering the future needs, the software should be scalable in terms of storage and upgrade. In future the digital libraries will be with multimedia intensive collections. Hence large number of data, audio/ video files will be needed to be transferred to users depending on the network bandwidth available. Each individual digital item added in the repository will be very large size. Hence it is necessary to know whether software is scalable in terms of storage and retrieval?

3.15.9 Extensibility

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

3.15.10 Storage

Does the system support data compression for e. g. 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 compressed file formats?

3.15.11 Backup/Restore facility

- 3.15.11.1 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?
- 3.15.11.2 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?
- 3.15.11.3 In the event of a disaster, what recovery plans are supported by the software?

3.16 User Friendly Interface

The user interface requirement covers the flexibility in customizing the interface to suit the needs of different digital library implementations as well as the support for multilingual access. With multilingual access, the user is able to specify the language for the DL's user interface as well as the cataloging information stored within it[25]. Generally, the user interface acts as a bridge between the user and the system-environment so that the user can interact with the DL software.

For evaluating user interface design the following tasks can be evaluated:

- 3.16.1 Whether end user can easily customize the user interface "look and feel"? Whether end user can change the header, theme, footer, overall web interface of the system?
- 3.16.2 Can user interfaces be localized in their language?
- 3.16.3 How is user interface for the experienced or non experienced users?
- 3.16.4 How is the user interface designed for the use of Librarian or for the use of users?
- 3.16.5 Whether system provide web based interface for all its functionalities?
- 3.16.6 Does the system provide online help?
- 3.16.7 Does the system provide multilingual access support?
- 3.16.8 Does the system allow users to submit feedback on system problems?

3.16.9 Does the system notify user's current state of the task?

3.17 Usability

According to Dumas and Redish[26], usability means people who use the product can do so quickly and easily to accomplish their own tasks.

International Standards Organization ISO[27] defines usability "as the extent to which a product can be used by specific users to achieve specific goals with effectiveness, efficiency and satisfaction in a specified context of use".

Usability measures the quality of a user's experience when interacting with a product or system -whether a web site, a software application, mobile technology, or any user-operated device[28]. In general usability refers to how well users can learn and use a product to achieve their goals and how satisfied they are with that process. Usability testing measures the quality of experience a user has when interacting with a software's web interface, including factors of ease, efficiency, memorability and satisfaction[29]. Usability testing uses metrics such as time taken by the system to completing a particular task, accuracy, satisfaction, and errors.

Blandford (2004)[30] grouped usability into two kinds: empirical and analytical. Empirical technique involve testing systems with users, whereas analytical techniques involve usability personnel assessing systems using established theories and methods.

There are two common approaches to usability i. e. heuristic evaluation and cognitive walkthrough. As conceived by Jacob Nielsen (1993)[31] the heuristic evaluation method employs a set of principles. The checking/ testing of following principles ensures that the system helps the users to conduct their tasks in an efficient, effective and satisfactory way[32].

Usability evaluation in digital libraries can also be carried out by using Nilesen's 10 principles.

In digital libraries user needs to be able to navigate through a digital library with a certain sense of ease, or else they may become frustrated and decide to go elsewhere. In digital libraries usability studies are now becoming more popular and implemental ways[33].

Though usability has been widely used in digital library evaluation, there is yet no uniform definition of what does it cover in digital library evaluation context.

While evaluating digital library software it is necessary to evaluate the following tasks:

- 3.17.1 What user skill levels are expected to use the web interface?
- 3.17.2 Is the user able to navigate and find his/ her ways to information quickly and easily?
- 3.17.3 How are the help features provided with the software? Does help feature allow user to feel comfortable within a digital library? Does the help feature provide general as well as technical answers to the problems?
- 3.17.4 Does the user needs to know which player or viewer is necessary to view multimedia or other digital objects from the system?
- 3.17.5 Does the user interface provide facility to give user opinion on the web as well as does the software support users to fill up online user feedback form?
- 3.17.6 Does the system support to show usage statistics such as usage patterns and which material is used more by which group, how often, when, for what reasons/ decisions etc:
 - (a) Usage patterns
 - (b) Use of materials
 - (c) Usage statistics
 - (d) Who uses what, when
 - (e) For what reasons/ decisions

3.17.7 Transaction Log Analysis

Transaction log analysis is a way to track how users are using a digital library. Analysis of transaction logs is one evaluation method that can be helpful to DL managers for making managerial decisions and establishing priorities, as well as indicating need for system enhancements[34]. Transaction logs provide a useful resource for the remote evaluation of web-based information systems due to their ability to record every action that occurs during the user's interaction with a digital resource. Log files

track and document information seeking behavior of users; network traffic statistics that offer data on where users originate, browses which sites etc. Logs can also allow meaningful interpretation and comparison of DL transactions.

When using transaction logs for evaluation, the main participants under survey are the user and the system, as well as the content that is being searched, read, manipulated, or created. The interaction between the system and the user can be examined and captured at various levels of abstraction. The transaction log analysis can provide an effective mechanism to detect data corruption or loss.

3.17.7.1 Does the system keep error log files?

3.17.7.2 Does the system keep user logs ? A user log is a technology for recording a user's interaction with the system. Typically this means installing a special ("instrumented") software application on the user's machine that "watches" everything the user does and records each word the user types and each command the user chooses[35].

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

3.18 Copyright/ Policy Issues

Copyright is an important area in digital libraries. No digital document can be added in the digital repository unless it is free out of copy right. Each document that gets added into the repository should clearly state the copyright policies for that document. The digital library software also should support mentioning details about copyright policies. SPARC[36] the Scholarly Publishing and Academic Resources Coalition, is an international alliance of academic and research libraries working to correct imbalances in the scholarly publishing system. Developed by the Association of Research Libraries, during 1998, SPARC has become a catalyst for change. (<http://www.arl.org/sparc/about/index.html>). Its focus is to create a new parallel publishing model along with commercial publishers. Leading academic organizations have endorsed SPARC. SPARC helps libraries to identify which publishers allow to keep their material on open access repositories. Hence while evaluating the selected software it is necessary to evaluate:

- 3.18.1 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?
- 3.18.2 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?
- 3.18.3 If repository ingests digital content with unclear ownership/ rights, does it have policies addressing liability and challenges to those rights.

3.19 Advanced Features

- 3.19.1 Whether software has long term leadership and consulting services?
- 3.19.2 Does the system provide citation data?
- 3.19.3 Does the software have defined road map for the future?
- 3.19.4 Does software support indexing of all documents that are added in the repository in a distributed way, without central control?
- 3.19.5 Virtual Collection Support: Does software supports to generate virtual collections / special collections across several content providers? The special/ virtual collections in digital libraries have more importance. Traditional libraries often contain, in addition to their main holdings special collections. A special collection is generally defined as a group of related materials that is given some form of special treatment. The special treatment might be due to the rare or delicate nature of the material (such as rare books, antique maps etc)[37]. In digital environment special collections are distributed across many servers, can be owned by different organizations and can be displayed in many different orderings and arrangements. Although it is common for traditional libraries to create and maintain special collections, many digital library do not attempt to provide a similar service. A new user who approaches a digital library can be better introduced to the digital library collection through the more easily exploreable partitioned set or resources in a virtual collection. For e.g. Open Video Project provides "Special Collection Spotlight" option where all special video collection is made available. <http://www.open-video.org/collections.php>

- 3.19.6 Load Balancing: Does software have any mention of load balancing? Load balancing refers to distributed processing and communications activity evenly across a computer network so that no single device is overwhelmed. Load balancing is important for networks where it is difficult to predict the number of requests that will be issued to a server. Busy web sites employ two or more web servers in a load balancing scheme. If one server starts to get swamped, requests are forwarded to another server with more capacity. Load balancing can also refer to the communications channels themselves. It is necessary to know how the software functions when number of records are more or more number of people try to access the software?
- 3.19.7 Visualization: Does the software support visualization feature? Visual interface to Digital libraries enable powerful data analysis and information visualization techniques. Visualization in digital libraries is implemented for information retrieval. Visualization tools offer the user more control over the information retrieval process by allowing the user to open icons to view documents, select subsets of documents, and to view document set interrelationships. Each document gets represented by an icon and the display is a graphical one. Each document also offers a multidimensional view of the data set to yield information for user decision-making in regard to item selection or query reformulation. The entire document set is presented to allow for browsing, and organized in ways to accommodate the user's information needs[38].
- 3.19.8 Personalization: Does the system support Personalization feature? Some digital libraries allow users to develop personal profiles and allow to save relevant or favorite items to these profiles for easy access. The digital library could use characteristics from personal profiles to recommend potential items to like-minded users. The personal digital libraries have been found as an important feature because of the unique way different users select, collect, and organize their resources.
- 3.19.9 Translation Service: Does the system provide translation service? Translation services add value to information transfer by transforming a document into another language, or another format, thus making it understandable to the user or beneficiary. Searching in the DL can be executed across continents, networks and systems. Query statements and results should be translated back and forth between different

search systems. A number of translation services offer the translation of documents into SGML and HTML format.

- 3.19.10 Community Services: Does the software support to provide community services such as allowing members of the digital library community to exchange ideas, make announcements, write reviews etc.
- 3.19.11 Page Ranking: Does the software support page ranking such as supported by commercial digital libraries .

3.20 Digital Preservation

The goal of digital preservation is to maintain the ability to display, retrieve, and use digital collections over the years in the rapidly changing technological and organizational infrastructures and elements. One of the central challenges to long term preservation in a digital repository is the ability to guarantee the interpretability of digital objects for users across time. This includes a guarantee of integrity, authenticity, confidentiality and accessibility to the digital data.

Print materials can survive for centuries and even millennia without direct intervention. In contrast, digital materials may need active management and preservation in order to survive even a decade. The preservation requirements consider storage, backup and long-term preservation issues of content and metadata.

When the digital documents are added in the digital repository it is necessary to know that the existing software supports the preservation of the digital documents over long term. There are several factors which are necessary to be considered in digital preservation part.

Digital preservation is one of the challenging areas and many digital library software have started looking into this aspect very recently. One of the hypothesis of this work is the existing digital library software do not support digital preservation aspect. Hence considering this hypothesis it was necessary to define evaluation criteria based on "digital preservation" and evaluate each software against "digital preservation" point of view.

The following are some of the criteria which are defined from digital preservation point of view:

- 3.20.1 Does the software support any digital preservation strategy and if yes, does it explicitly support any particular preservation strategy, such as described by PREMIS, bit-level preservation, format migration, format normalization, emulation, or restrictions on submission formats?
- 3.20.2 Whether system preserve file's original identities such as its name, size and created date?
- 3.20.3 Does the system have any data integrity check for digital documents that are added into the repository?
- 3.20.4 Does the software have quality control measures to ensure integrity, and persistent documentation identification for migration purposes?
- 3.20.5 Does the repository software preserve pre-existing persistent identifiers for submitted packages or objects?
- 3.20.6 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.
 - authenticity - validating that the digital object is in fact what it should be, and has not been altered.
 - preservation activity - documenting the actions taken to preserve the digital object.
 - technical environment - describing the technical requirements, such as hardware and software, needed to render and use the digital objects.
 - rights management - recording any binding intellectual property rights that may limit the repository's ability to preserve and disseminate the digital object over time.
- 3.20.7 How does the software manage compound objects (where multiple file formats of same object are linked together)?

- 3.20.8 Does it have mechanism to keep licensing conditions for individual images/ objects in the repository?
- 3.20.9 Does software support to track and manage copyrights and restrictions on use as required by contract or license?
- 3.20.10 Where does the repository software store the actual digital files and the metadata?
- 3.20.11 Can the repository software automatically validate checksums on a periodic basis?
- 3.20.12 Are the checksums cryptographically signed to prevent tampering?
- 3.20.13 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?
- 3.20.14 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?
- 3.20.15 Does the repository software provide audit logs of all events that have occurred in the life cycle of a package? What events are logged?
- 3.20.16 If repository ingests digital content with unclear ownership/ rights, does software has policies addressing liability and challenges to those rights?
- 3.20.17 Does the software have mechanisms to change its preservation plans as a result of its monitoring activities? In what format are the logs? Are they easily processed by human or machine? Are the logs tamper resistant?
- 3.20.18 Does the system have any mechanism in determining when objects in digital archives should migrate to new hardware and software?
- 3.20.19 Can the software support scheduled events such that a human can be notified on a preset schedule to manually check for format obsolescence?
- 3.20.20 Can new file formats be added or removed?
- 3.20.21 Does the software have ability to handle variety of file formats and does it also supports file format versioning?

- 3.20.22 Does the software has mechanisms in place for monitoring and notification when format (or other representation information)obsolescence is near/ or are no longer viable?
- 3.20.23 Can the software monitor any standard format registries in order to ascertain format obsolescence?
- 3.20.24 What standards does the repository software use to describe file formats and does the software record representation information? Does it use Internet MIME Types?
- 3.20.25 Does it use format registries? If yes, which format registry is used such as PRONOM or GDFR or DCC to represent format information?
- 3.20.26 Does the system support automatic format registration. For unknown formats does the system send any message to the submitter requesting for additional information?

3.21 Conclusion

With the present state-of-the-knowledge, no evaluation study can cover all aspects involved in digital library software. Thus, there is no complete "evaluation of open source digital library software". There would be only an evaluation of some of the elements in their construct. An attempt has been made to prepare an exhaustive list of evaluation criteria while evaluating open source digital library software, though it cannot be claimed that this list is a complete list and without any flaws and covering all aspects of digital libraries from all angles.

All the selected candidate software have been evaluated on the basis of the above defined evaluation criteria. The next chapters deal in detail with introduction of each selected digital library software/ Institutional Repository Software/ Open Repository Software that has been considered for the present study and observations/ results of evaluating each software against the defined set of evaluation criteria.

Bibliography

- [1] Lakos, A. (1999). The missing ingredient: culture of assessment in libraries. *Performance Measurement and Metrics*, 1(1), p.3-7, <http://www.aslib.com/pmm/1999/aug/opinion.pdf>. (Browsed on 1st March 2007).
- [2] Xie, H. (2006). Evaluation of digital libraries: criteria and problems from users' perspectives. *Library and Information Science Research*, 28(3), p.433-452.
- [3] Witten, Ian H. & Bainbridge, D. (2007). A retrospective look at Greenstone: lessons from the first decade. *JCDL'07*, June 18-23, p.147-156.
- [4] Fuhr, N. [et al.] (2007). Evaluation of digital libraries. *International Journal of digital libraries*, Springer, <http://www.springerlink.com/content/w2w0j4h272k26812/>. (Browsed on 1st June 2007).
- [5] Fuhr, N., Hansen, P., Mabe, M., Micsik, A. and Solvberg, I. (2001). Digital libraries: A generic classification and evaluation scheme. *Springer Lecture Notes in Computer Science*, 2163, p. 187-199. <http://www.springerlink.com/content/w22qmkk09ncy59av/>. (Browsed on 15th June 2007).
- [6] Fourth DELOS Workshop (2003). Evaluation of Digital Libraries: Test beds, Measurements, and Metrics. Final report to National Science Foundation, Digital libraries Program. *NSF report on DELOS workshop on DL, evaluation*, 21 p. http://www.dli2.nsf.gov/internationalprojects/working_group_reports/evaluation.pdf. (Browsed on 2nd June 2007).
- [7] Saracevic, T. & Covi, L. (2000). Challenges for digital library evaluation. *Annual Meeting 2000 of the American Society for Information Sci-*

- ence http://www.is.informatik.uni-duisburg.de/courses/dl_ss04/folien/saracevic00.pdf. (Browsed on 20th May 2007).
- [8] Fourth DELOS Workshop, *ibid.*, p. 5.
- [9] King, N., Ma, T.H.-Y., Zaphris, P., Petrie, H. and Hamilton, F. (2004) : An incremental usability and accessibility evaluation framework for digital libraries in Brophy, P., Fisher, S. and Craven, J. (Eds), *Libraries without walls 5: The Distributed Delivery of Library and Information Services*, Facet, London, pp. 123-31.
- [10] Saracevic, T. (2004). Evaluation of digital libraries: an overview. http://www.scils.rutgers.edu/tefko/DL_evaluation_Delos.pdf. (Browsed on 30th June 2007).
- [11] Saracevic, T. (2000): Digital library evaluation: toward an evolution of concepts, *Library Trends*, 49(3), Fall 2000, p. 350-369.
- [12] Saracevic, T. (2004). *ibid.*, p. 5.
- [13] Saracevic, T., Covi, L. (2000). Challenges for digital library evaluation. *Annual Meeting 2000 of the American Society for Information Science*.
- [14] Saracevic, T. (2005). How were digital libraries evaluated? Presentation at the course and conference Libraries in the Digital Age (LIDA 2005), 30 May - 3 June, Dubrovnik, Croatia. http://www.scils.rutgers.edu/tefko/DL_evaluation_LIDA.pdf. (Browsed on 20th September 2007).
- [15] Tervonen, I., Kerola, P. (1998). Towards deeper co-understanding of software quality. *Information and Software Technology*, 39(14/15), p.995-1003.
- [16] Prasad, ARD & Guha, Nabonita (2005). Persistent identifiers for digital resources. *DRTC-HP International Workshop on Building Digital Libraries using Dspace, Paper J*, DRTC, Bangalore, 11th March, Available at <https://drtc.isibang.ac.in/handle/1849/246>. (Browsed on 20th November 2007).
- [17] Cathro, W. (1997). Metadata: an overview, paper submitted at Standards Australia Seminar.

- textitMatching discovery and recovery, August, <http://www.nla.gov.au/nla/staffpaper/cathro3.html>. (Browsed on 20th March 2007).
- [18] Madalli, Devika P. (2003). Digital libraries and digital library initiatives. *DRTC Workshop on Digital Libraries: Theory and Practice, Paper A*, DRTC, Bangalore, March, 7 p. Available at <https://drtc.isibang.ac.in/handle/1849/10>. (Browsed on 10th September 2007).
- [19] Brewer, A. (1996). The role of intermediary services in emerging digital libraries. *DL 96*, Bethesda MD USA, p.29-35
- [20] Shen, R., Vemuri, N. S., Fan, W. [et al.] (2006). Exploring digital libraries: integrating browsing, searching and visualization. *JCDL-06*, p. 1-10.
- [21] Meyyappan, N., Chowdhury, G. G., Foo, S. (2000). A review of the status of 20 digital libraries. *Journal of Information Science*, 26(5), p. 337-355.
- [22] Ferrari, E., Adam, N. R., Atluri, V., Bertino, E., Capuozzo, U. (2002). An authorization system for digital libraries. *The VLDB Journal*, 11, p.58-67.
- [23] <http://techessence.info/node/48>. (Browsed on 15th September 2007).
- [24] <http://en.wikipedia.org/wiki/Encryption>. (Browsed on 1st October 2007).
- [25] Witten, I. H. & Bainbridge, D. (2002). *How to Build a Digital Library*, Morgan Kaufmann, San Francisco, CA.
- [26] Dumas, J. S. and Redish, J. C. (1999). *A practical guide to usability testing*, revised edition, Exeter, England, Portland, Or, Intellect Books, p.4.
- [27] International Standards Organization. (1994). *Ergonomic requirements for office work with visual display terminals. Part 11: Guidance on usability (ISO DIS 9241-11)*. London: International Standards Organization.
- [28] <http://www.usability.gov/basics/whatusa.html>. (Browsed on 20th October 2007).
- [29] Jeng, J. (2005). Usability assessment of academic digital libraries: effectiveness, efficiency, satisfaction and learnability. *Libri*, 55, p.96-121.

- [30] Blandford, Ann, Keith, S., Connell, I., Edwards, H. (2004). Analytical usability evaluation for digital libraries: A case study. *JCDL04*, June 7-11, p.27-36.
- [31] Nielsen, J. (1993). *Usability engineering*, San Francisco: Morgan Kaufmann, 362 p.
- [32] Fuhr, N. (2007). Evaluation of digital libraries. *International Journal on Digital libraries*, Springer, 18 p.
- [33] Battenfield, B. (1999). Usability evaluation of digital libraries. *Digital libraries: philosophies, technical design considerations, and example scenarios*, D. Stern (ed.), Binghamton, NY: Haworth Press, p.39-60.
- [34] Klas, C., Albrechtsen, H., Fuhr, N. [et al.] (2006). A logging scheme for comparative digital library evaluation. *ECDL-2006*, LNCS 4172, edited by Gonzalo, J., p.267-278.
- [35] <http://faculty.washington.edu/farkas/TC407/Evaluation>. (Browsed on 10th October 2007).
- [36] <http://www.arl.org/sparc/>. (Browsed on 14th September 2007).
- [37] Geisler, G., Giersch, S., McArthur, D. [et al.] (2002). Creating virtual collections in digital libraries: benefits and implementation issues. *JCDL 2002*, July 13-17th, p.210-218.
- [38] Koshman, S. (2006). Visualization-based information retrieval on the web. *Library and Information Science Research*, 28, p.192-207.