J. Basic. Appl. Sci. Res., 2(11)10843-10852, 2012

ISSN 2090-4304
Journal of Basic and Applied Scientific Research
www.textroad.com

Evaluation of Three Open Source Software in Terms of Managing Repositories of Electronic Theses and Dissertations: A Comparison Study

Mohamad Noorman Masrek[1], Hesamedin Hakimjavadi[2]

[1] Accounting Research Institute and Faculty of Information Management, Universiti Teknologi MARA (UiTM), Shah Alam, Malaysia
[2]Faculty of Information Management, Universiti Teknologi MARA (UiTM), Shah Alam, Malaysia

ABSTRACT

Electronic Theses and Dissertations (ETDs), as a new generation of scholarship resources, are gradually increasing in number and quality at higher academic institutions. Meanwhile, by introducing various types of software solutions for managing Institutional Repositories (IRs), selection of appropriate solutions has become a time-consuming process for institutions. The goal of this paper was to appraise 59 features of three widely utilized open source IR solutions (DSpace, EPrints, Fedora) from the perspective of managing ETDs, via an in-depth evaluation of their important functionalities in this regard. For this purpose, all applications were installed and the features were tested in a test-bed environment (a benchmark machine) with a predefined set of ETD collections and registered users. Findings related to assessment of each feature were presented in the tabular format. Our comparison indicated that, although all three solutions are capable of managing ETD systems, in most of the comparative areas that are vital for an ETD repository DSpace was ahead of EPrints and Fedora.

KEY WORDS: Institutional Repository, DSpace, EPrints, Fedora, Open-source software assessment.

INTRODUCTION

Nowadays, there is a broad consensus on the vital importance of openness and dissemination of scientific information resources over the Web. Institutional Repositories (IRs) are one of the most reliable types of these sources. Furthermore, among all types of IRs, the motion of generating Electronic Theses and Dissertations (ETDs), as a new genre of scholar documents, has achieved significant progresses during the last decade, and universities are providing free access to a huge number of ETD collections through their portals.

At present, there are a number of software solutions promoted to manage repositories systems, and most of them can also handle ETDs[1]. Due to the different nature and complexities of these packages, there are limited guidelines, which can be used by academic institutions to choose the most appropriate application for their ETD management system implementation [2-4].

Mining the literature unveiled that there are several evaluations conducted to compare features and functionalities of free and commercial software for establishing IRs[5-10]. However, it seems that none of them has focused on ETD and its unique characteristics[11]. In addition, most of these studies were meant to provide general overview of various software products. Hence, an in-depth study evaluating a wide range of characteristics and functionalities is still very scarce. This evaluation could also be considered as a new update of the previous comparison studies in this context.

The goal of this paper is to appraise important functionalities of three open-source software (OSS), from the perspective of managing ETDs, via an in-depth evaluation of their features. At the theoretical level, this research attempts to propose some revisions to previous IR software evaluation methods in order to adapt them to the current requirements of the ETD systems, with an emphasis on the resource integration functions.

The three evaluated OSS were chosen because they were found the most popular software for the implementation of IRs [12]. Additionally, according to[13], 67% institutions that have implemented ETD have chosen these applications. These percentages were also confirmed through another global survey on adoption of OSS in this field[14].

The process of software selection for sustainable systems is multi-dimensional. Selection of an ETD software solution for an IR that maintains a diverse range of scholar documents (reports, articles, journals, etc.), is not just based on its ability to meet the ETD-specific requirement, but the other characteristics related to the interoperability of systems. Then, although the focus of this work was on the features related to the ETDpublishing, in order to provide a comprehensive and useful comparison, in addition to those features, a number of interoperability-related key features of each softwarewere evaluated as well.

The conduct of this study involves the use of literature review and experimental comparison. Working with "Features and functionalities of Institutional Repositories"[15] as a guideline, 16 major criteria were selected for evaluation of these applications. In addition to these criteria, some unique characteristics that affect the process of

***Corresponding Author:** Hesamedin Hakimjavadi, Faculty of Information Management, Universiti Teknologi MARA (UiTM), Shah Alam, Malaysia, Tel: (+1)806-448-4826, email: hesamhakim@gmail.com

ETD integration [11]have also been considered for the evaluation. Every major criterion was broken into a number of sub-criteria for in-depth evaluations. All applications were installed and tested in a test-bed environment (a benchmark machine) with a set of pre-defined ETD collections, and a number of registered users. The results obtained from the test have been compared with the previous studies [e.g. 16, 17-19]

Technical building locks, Characteristics, and Challenges

Each ETD repository solution regardless of the software that handles ETDs, consists of at least five essential component, including:

i. interface for adding document and metadata
ii. interface for searching and browsing through the content
iii. database for storing and managing the content
iv. administrative interface
v. additional features such as integration with other repository systems [20]

In addition to the inherited general characteristics of peer reviewed scholarly document[21], ETDs have a number of unique characteristics, including:

- availability of contents in several languages
- large book-size documents in addition to related multimedia files
- availability of full content in various formats (XML, PDF, etc.)
- a multiplicity of bibliographic references
- rich sets of metadata elements that are different in quality
- variety in range and scope of user interests [22]

The flexibility of workflow and the ability to adapt to the different implementation models are the key factors in ETD repositories. Describing three of the main implementation models of ETD repositories (distributed, semi distributed, and semi centralized models), Tenant (2002) states that "there are nearly as many models as there are institutional repositories" [23].Furthermore, Gibbons suggested two other featuresto be evaluated specifically in ETD management systems, including multi-file deposit structure and complexity in access control [15].

In terms of resource integration and information retrieval methods, one of the challenges of ETD repositories is enhancing metadata schemain order to improving subject access [24]. Additionally, in relation with subject classification schemas, ETDs' topics are highly specialized and frequently interdisciplinary, and consequently are not often classified deep enough[25].

It has been indicated that using web2.0 techniques for integration and retrieval of ETD records is not efficient. Because ETDs are usually not read and annotated by many people and then the social tagging approach is not a useful method for increasing the visibility of ETDs [24].

Related Comparison Studies

In 2004, University of Queensland provided a short report of evaluation of 14 Free OSS for managing IRs, that theircommon denominator was their compatibility with OAI-PMH (Open Access Initiatives Protocol for protocol[26].

In the same year, The Open Society Institute released the third edition of “A Guide to Institutional Repository Software” [16]. This comprehensive guide coveredOAI compliant repository packages, including Archimede, ARNO, CDSware, DSpace, EPrints, Fedors, i-Tor, MyCoRe, and OPUS.

In an effort to assist regional universities in adopting best practice for running their IRs, RUBRIK Toolkit was launched by Regional Universities Research Infrastructure Collaboratively in regards to the process of selecting repository software.[27]. Another study in this context is the “Recommendations on National Library of Medicine Digital Repository Software” conducted by NLM Digital Repository Evaluation and Selection working Group [18]. In 2007, Networking of Swiss Academia (SWITCH) evaluated Fedora and DSpace for establishing an object repository project for e-learning contents of the Swiss universities [9].Recent study conducted by *Repositories Support Project* (RSP) has provided a short repository software survey to compare main features of eleven free and commercial packages [19].

Regular IRs software packages

Open Society Institute (OSI) has introduced ten OAI compliant OSS that could be used to handle IRs [16]. Additionally, there are also a number of commercial packages as well as hosting services available to support IRs implementation.

In 2007, Primary Research Group conducted an international survey of 253 institutional repositories [13]. This survey demonstrated the usage distribution of different IR software, based on which, DSpace was the most common software (37%), followed by Fedora (17%) and EPrints (13%). To this effect, these three web applications will be evaluated for the purpose of this study.

DSpace was developed jointly by HP and MIT as an OSS. The aim of the project was to create a package that could provide an institutional repository solution, which addressed the problem of electronic publishing and digital preservation as a central theme [28].

GNU EPrints is general digital repository management software that was developed by the University of Southampton. Its developer's goal was creating a flexible platform for building high quality and scalable repositories [29].

Fedora (Flexible Extensible Digital Object Repository) is designed to be a foundation for interoperable web based electronic libraries [20]. In the context of IRs, Fedora can be used as a basic environment which full-featured IR system can be built upon it. Fedora was developed by University of Virginia and Cornell University. Since Fedora provides a web service, rather than, web-based UI front-end, in this evaluation Fedora is tested with Fez as its web interface (Fez is one of the most common web interface provided for Fedora).

METHOD

Methods of Evaluation of IR Software Packages

Varioussets of criteria have been proposed for assessing software to be considered when selecting solution for managing IRs. Goh & Chua (2006) defined 12 categories of items, each with varying degrees of importance, including: preservation, interoperability, metadata, search, access control and security, report and inquiry, user interface, content acquisition, standards compliance, automatic tools and support, and content management.Then, they assigned aset of criteria to each category to assess it. They also assign weightings to each category and criterion, based on the importance of each [5].

Kaczmarek & Hswe (2006) proposed a framework for evaluating repository software applications based on Using the Audit Checklist for the Certification of a Trusted Digital Repository[7].At the same year,Wyles applied another set of criteria for evaluating six open source application, including: DSpace, Fedora, EPrints, ARNO, CDSware, and i-TOR [17]. The selected criteria were scalability, ease of working on code-base, extensibility, security, interoperability, ease of deployment, ease of system administration, internationalization, open source, quality, and configurability of workflow tools, strength of community.

Criteria for Evaluating Solutions

The conduct of this study involves the use of literature review and experimental comparison. Working with "Features and functionalities of Institutional Repositories"[15] paper as a guideline, 16 major criteria were selected for evaluation of these applications. In order to establish sustainable and interoperable IRs, Gibbons (2004) suggests features and functionalities that should be appraised, including metadata and protocols, format types, environment and infrastructure compatibility, deposit structure, batch importing or exporting, persistent linking, search engine, type of license (open source or commercial), versioning, access control, user interface, authentication or off-site access, editors and administrative levels, clients, usage statics, and extensibility[15].

We adopted criteria set suggested by Gibbons for this study because among above mention methods proposed for evaluating IR software packages, this criteria set considers the ETD-specific aspects of software solutions. Moreover, for features that were not fully covered in this set, guidelines and findings from previous studies were employed[16-19].

Following the method proposed in *An Evaluation of Open Source Software: for Building Digital Libraries*[30], the16 suggested major features were broken into 59 sub-features for in-depth evaluations (See

Table 1 to Table 12).

In order to test these features, the last stable version of the software packages at the time of evaluation (i.e. DSpace 1.6.1, EPrints 3.2.4, and Fedora/Fez2.1.RC3) were installed in a test-bed environment with 108submitted ETDs and 21 registered users (Including super admin, admins, editors, end users) in each system. ETD records was extracted from five repositories including: OhioLINK - Electronic Theses and Dissertations Center(34 records), Electronic Theses and Dissertations at Virginia, Tech (26 records), CaltechTHESIS (21 records), Durham e-Theses (21 records), and White Rose E-theses Online (6 records).The study performed on a HP personal computer with Intel Processor - dual core 2.4 GHz, 4GB RAM, 160 GB SATA Hard Disk, and a 10-100 network adaptor. OpenSUSE 11.1 (Linux kernel 2.6.27) used as operating system. The results obtained from the test were checked and compared with the previous studies.

Features and Functionalities of ETD Repository Management Systems

In this part, the set of 16 features and functionalities[15] as well as 59 derivative sub-features assessed in these software packages are briefly reviewed.

Meta data

Providing the metadata for ETD records in various schemas improves the relevancy of results retrieved by the semantic search. Indeed, in order to improve the visibility and accessibility of the TDs, providing multiple classification criteria is crucial. Today the metadata enhancement is one of the big challenges in the context of ETDs[31]. The metadata enhancement can take place at three levels of data structure, categorical data and factual data [24]. Each digital library of ETDs follows its standards, policy and data model for assigning metadata to

ETDs. The flexibility and extensibility of software to handle metadata schemas is a key factor that must be considered in the selection process.

In order to evaluate the metadata-related functionalities of software, we compare seven features in these three packages, including: Metadata schema support, Metadata checks, Extendable metadata schema, Metadata Export/Import, Controlled vocabulary, Social tagging, and Name authority control. The results of assessment of these features are represented in Table.

Machine to Machine Interoperability

Interoperability among different levels of information systems is one of the most basic factors considered by decision makers when selecting platform, software, standards, and metadata schema for information systems. It becomes more important when dealing with open digital repositories, in which integration and interoperability are considered as an advantage over their traditional corresponding.

Table 2 represents the results of compared features, including: Compatibility with OAI-PMH, Data harvesting Features (OAI-PMH service provider), SRU/SRW (version 2.0), Support Z39.50 and Z39.78, Unicode, PREMIS, Support of OAI-ORE and SWORD.

Subject Classification

Due to the highly specialized and frequently interdisciplinary nature of TDs' topics, librarians, even in traditional OPAC systems should follow a narrow and multilayer yet widely accepted subject classification system for categorizing topics of TDs. It is suggested to provide a narrower subject classification (comparing to commonsubject classificationamong the other fields of scholarly documents), in order to improve the access to TDs[25]. On the other hand, by using different subject classification systems the interoperability between systems will be decreased and retrieving records from multiple repositories based on their subjects would be more complex. Table 3 illustrates the results of comparing the three repository software solutions in terms of supporting widely accepted subject classifications of LoCClassification, LoCSubject Headings, DDC (Dewey Decimal Classification), UDC (Universal Decimal Classification).

Format Type

Diversity of format types that can be stored and managed in a repository is one of the interoperability-related aspects, especially in ETD repositories, because ETD records are instances of resources that consist of several files (e.g. full-text documents, images, multimedia, etc.)[22]. Additionally, the full contents of ETDs are available in various formats (e.g. XML, PDF, HTML, etc.). The evaluated features in terms of document format compatibility are: Accepted document types (Text, Image, Video, Audio), Special Document types, Extensibility (See Table 4)

Administrators and Editors Levels

Distribution of administrators' and editors' tasks in some customizable levels improves the flexibility especially in implementation of workflows for an ETD management system. For instance, in a regular ETD repository, once submitted, a dissertation is immediately placed in a virtual holding area and is not accessible by end users before the committee approve the document as an electronic dissertation. Before this step, based on the institution policies, someone may check the document to ensure that preferred format is utilized, and another person verifies plagiarism. Usually, after passing the approval step, another person assigns appropriate metadata to the document and makes it ready for publishing via ETD system.

Evaluated features in this regard were: monitoring system functionality, managing user permissions, administrating multiple collections and submission processes, workflow administration, Data recovery, and file migration (Batch processing). The results obtained from the evaluation are available in Table 5.

Monitoring system functionality allows a system administrator to monitor the functionality of the entire system and to generate customized report.

Administrating multiple collections and submission processes: Multiple collections refer to the ability of a repository management system to handle several collections with separate submission workflows in a single repository. This function is useful in ETD management systems, because each degree, and in some cases, each faculty has its specific workflow (submission and approval stages) for theses and dissertations.

Import/Export

The Import/Export function usually refers to interoperability within the same type systems (e.g. DSpace to DSpace). In fact, the bulk import/export between different systems often is conducted through protocols such as OAI-ORE. The evaluated features in this regard were: Batch import or export, Upload/Download compressed file, Upload to/Download from URL, Bulk import/export for document, Bulk import/export for metadata. The results of evaluation are available in Table 6.

User Validation

Utilizing user validation services in repositories are an example of integration with external service providers. Shibboleth[32] and Lightweight directory access protocol (LDAP) [33]are instances of such systems that digital libraries are configured to work with them. In fact, by integrating the local user validation system with centralized validation mechanisms, once a user is validated by the system, he or she can access to all integrated repositories. In this way centralized (e.g. Athens) as well as distributed (LDAP) validation mechanisms improve the integration of repositories in terms of user validation. The evaluated features in this relation were: Username and password authentication, Centralized System (Athens), and Distributed Systems (Shibboleth, LDAP). The results are available in

Table 7.

Access control

Access control simply could be defined as the ability of a system to limit users or groups to access to some files or function of repositories, based on the access policy. Gibbons (2004) propounds some possible scenario for different types of access control policies in ETD collections to demonstrate the complexity of this function in ETD repositories[15].

The evaluated features in the context of access control involve: Control access based on the user type, Control access based on object (document) specifications (See Table 8).

Control access based on the user type: This feature allows the administrator to limit access to the certain content based on the user's level of authorization [16]. This function could be used, for example, to restrict access to ETDs that are part of a patent or are restricted by research funding organization.

Control access based on the object (document) specifications: Due to their multi-file nature, ETDs are extreme examples of complex level of access management. For instance, a thesis might contain a copyrighted video, which should not be accessible by public users.

Language

One of the characteristics of ETDs is their availability in more than one language [11]. In many cases, especially in European and other non-English countries, titles and abstracts of ETDs, in addition to their local language, are also available in English. Evaluated features in connection with language are represented in Table 9.

Versioning

The versioning feature (See Table 10) refers to the ability of the system to store, track and retrieve multiple versions of documents.

Environment and infrastructure compatibility

In order to secure the stability of an open-source-based repository system, prior to the software selection, a manager needs to be ensured about the sufficiency of experience of the related staffs in terms of programming and database management skills for the selected platform. Furthermore, to achieve the best performance, the selected software should be compatible with institution's infrastructure and computer environment (such as operating system platform, server types, security policies, etc.). In order to select the most compatible software, each solution should be checked from the platform-related aspects such as: Compatible operating systems, Compatible databases, development languages and hardware. In the Table 13 only software-related aspects have checked (examination of each solution from hardware aspects is out of scope of this study)

RESULTS

In this section, the evaluated features and functionalities of the selected software and the results of evaluation are presented in a tabular format. These features were described briefly in the previous section, and their related tables were mentioned (See *Features and Functionalities of ETD Repository management systems*)

Table 1: Comparison of metadata features

Feature	DSpace	EPrints	Fedora
Metadata schema support	qualified Dublin Core (Default), Dublin Core, METS (mapping between various metadata schemas and QCD is available)	DC, METS, MARC21,	Dublin Core ,METS, MARC21, MRCXML, ONIX, MODS, EAD, TEI
Metadata checks*	manual metadata check (batch processing is not supported)	several levels of metadata check as a parts of submission workflow	Add-ins
Extendable metadata schema	Extensible with some changes in code-base	Fully extensible to support all types of metadata schema	Fully extensible to support all types of metadata schema
Export to/Import from other metadata formats	YES (through fully customizable XML)	YES (through fully customizable XML)	YES
Controlled vocabulary	YES	N/A	YES
Social tagging	N/A	N/A	N/A
Name authority control	YES (OCLC Research has provided name look-up service as a metadata creation support to be integrated with templates in DSpace [24]	N/A	N/A

***Metadata check**: This feature refers to the quality control system for incoming metadata sets in orderto check whether the associated document meets minimum metadata requirements or not. If metadata set was incomplete, it would be rejected to submitter, or sent to a metadata provider to be enhanced. This quality control system is important for ETD collections that are OAI compliance.

Table 2: Machine to machine interoperability 10847

Feature	DSpace	EPrints	Fedora
Compatibility with OAI-PMH	Version 2.0	Version 2.0	Version 2.0 (Add-ins: OAI-provider tools)
Data harvesting Features (OAI-PMH service provider)	YES	N/A	YES
SRU/SRW (version 2.0)	Version 1.2	N/A	Version 1.2, Add-in (provided by VLTS)
Support Z39.50 and Z39.78	N/A (Planed for future versions)	N/A	N/A
Z39.78	N/A	N/A	Optional
Unicode	Supported (by configuration of database management system)	YES	Supported as content characters but not as file name
PREMIS	N/A	N/A	Supported (but FEZ does not render this schema)
Support of OAI-ORE	YES	YES	Optional
Support of SWORD	YES	YES	YES

Table 3: Subject classification systems

Feature	DSpace	EPrints	Fedora
LoCClassification	YES	YES	N/A
LoCSubject Headings	N/A	Add-ins	N/A
DDC	N/A	Add-ins	N/A
UDC	N/A	Add-ins	N/A
Extensibility in terms of subject classification systems	Fully extensible for providing any user-defined classification systems	Availability of several plug-ins for adding different subject classifications	Fully extensible for providing any user-defined classification systems

Unlike common digital library solutions, none of the evaluated software widely supports these subject classifications by default. However, their customizability allows third party developers to add these features to repositories.

Table 4: Format Types

Feature		DSpace	EPrints	Fedora
Accepted document types	Text	PDF, Postscript, DOC, RTF, HTML, TXT, XML	PDF, Postscript, DOC, RTF, HTML, TXT, XML	PDF, Postscript, DOC, RTF, HTML, TXT, XML
	Image	JPEG, GIF, PNG, TIFF	JPEG, GIF, PNG, TIFF	JPEG, GIF, PNG, TIFF
	Video	AVI, MPEG	Add-in	AVI, MPEG
	Audio	AVI, MPEG	Add-in	WAV, MP3
Special Document types		websites, CAD, Drawings, 3D objects		learning objects (SCORM)
Extensibility		Extensible for all types of documents	Extensible for all types of documents	allows user and administrator to store and manage all types of content

Table 5: Administrators and editors levels

Feature	DSpace	EPrints	Fedora
Monitoring system functionality	YES	N/A	N/A
Managing user permissions*	YES (complex and flexible)	YES (Inflexible)	YES (XACML based)
Administrating multiple collections and submission processes	YES (Different collections with different workflow and different homepages)	YES (Different collections with different homepages)	YES (Different collections with different workflow and different homepages)
Workflow administration	Optional, fully customizable	Built it feature	Optional (FEZ), outside of repository
Data recovery**	Export/Import (metadata, content, user information)	Export/Import (metadata, content, user information)	Regenerate, Export/Import in three levels: Archive, Migrate, Public Access
File migration (Batch processing)	Add-in	Add-in	Built-in

***Managing user permissions*:** The complex structure of user administration In DSpace, allows an administrator to define privileges for all users based on their types. In the case of Fedora, the major feature of security architecture is built upon "eXtensible Access Control Markup Language" (XACML) and a XACML-based policy enforcement module. All types of access controls in Fedora repositories are conducted through XACML. The flexibility of XACML policies allows administrators to edit security options on the item, group, and collection levels. Although EPrints support access control function on all levels, it seems that this feature is not as flexible as it is in other listed software in this paper.

****Data recovery*:** There are three main different ways in Fedora to export backup files and recover data. The first approach is *Archive*, which exports xml file containing encoded data and metadata. The second method is *Migrate*, which is designed for transforming objects or collection to another Fedora repository web server. In this method exported XML file just includes metadata and links to documents. And the third approach is *PublicAccess* that its function is similar to *Migrate*, but it is designed to transfer objects between Fedora and other types of repository web servers. In addition to these methods, Fedora developers have introduced FOXML that can regenerate the entire system.

Table 6: Import from/ Export to features

Feature	DSpace	EPrints	Fedora
Batch importing or exporting	YES	YES	YES
Upload/Download compressed file	YES	YES	YES
Upload to URL/Download from URL	YES	YES	YES
Bulk import/export for document	YES	YES (with some shortcomings in practice)	YES
Bulk import/export for metadata	YES	YES	YES

Table 7: User validation

Feature	DSpace	EPrints	Fedora
Username/Password	System assigned password, Forgotten password, Edit profile	Returning forgotten password, Edit profile	Returning forgotten password
Centralized Systems: Athens	N/A	N/A	Add-in
Distributed Systems: Shibboleth	Built-in	Add-in	Add-in
LDAP	Built-in	Built-in	Built-in

Table 8: Access Control

Feature	DSpace	EPrints	Fedora
Control access based on user type	YES (Fully configurable)	YES	YES (Trough XACML)
Control access based on object (document) specifications	YES (Fully configurable)	YES	YES (Trough XACML)

Table 9: Language

Feature	DSpace	EPrints	Fedora
Multi language support in interface	Built-in	Built-in	Optional
Multi language support in submitted documents	Built-in	Built-in	Built-in
Support Unicode	Utf-8 encoding (through database functions)	Utf-8 encoding	Utf-8 encoding
Other language related files(help pages, input forms, email templates)	Available	Available	Developing

Table 10: Versioning

Feature	DSpace	EPrints	Fedora
Versioning	N/A (planed for version 2.0)	Supported through Version Identification Framework (VIF)	Fully Supported (In addition Fez provides *version viewing* and *version of content* components to improve the capability of versioning

t Linking, Search and retrieval.

Table 11 to Table *15* represent the comparison of six general features of evaluated OSS that are not specific to ETD repository functions, but are important specifications that should be considered in the process of decision making for selecting software. These features are: Interface, Open source license, Environment and infrastructure compatibility, Persistent Linking, Search and retrieval.

Table 11: Interface

Feature	DSpace	EPrints	Fedora
Apparent structure of front-end page	Fixed, Configurable through code base	Fixed, Configurable through code base	Fully customizable through different web-based UI front end applications
Theme of interface	Customizable	Customizable	Customizable
User view (ability of customizing user's view of the specific objects (page, document, collection)	Add-in	N/A	YES
Customizability of header and footer of static and dynamic pages	Customizable header for each collections, Fixed footer	Customizable header and footer for each collections/pages	Customizable header and footer for each collections/pages

Table 12: Open source license

Feature	DSpace	EPrints	Fedora
Distribution	Free, Open Source	Free, Open Source	Free, Open Source
License	BSD license	GNU Public License	Apache License V2.0

Table 13: Environment and infrastructure compatibility

Feature	DSpace	EPrints	Fedora
Minimum hardware requirements	Not specified	Not specified	Not specified
SAN architecture support	YES	YES	YES
Operating System	Linux, Sun Solaris, IBM AIX, BSD, HP/UX MS Windows Mac OS	Linux (Debian/Ubuntu and Redhat/Fedora) MS Windows(XP, Vista) Mac OS	UNIX-like platforms (Linux, Sun Solaris, IBM AIX, etc) Mac OS MS Windows (not stable)
Web Server	Jakarta Tomcat 5.0 or latest 4.1.x release, every servlet container or http server that can do SSL (such as Jetty and Caucho Resin)	Apache web server	Jakarta Tomcat 5.0 or latest 4.1.x release
Default Database	PostgreSQL	MySQL	MySQL
Compatible Databases	Oracle, DB2, MySQL	Oracle	Oracle, PostgreSQL, McKoi, MS SQL and DB2
Programming Language	JAVA	Perl, (JavaScript and AJAX as scripting language)	JAVA ((JavaScript and AJAX as scripting language)
Extensibility	Extensible for all types of documents	Extensible for all types of documents	allows user and administrator to store and manage all types of content

Table 14: Persistent Linking

Feature	DSpace	EPrints	Fedora
Persistent Linking	YES	YES	YES (Based on progressive method of URN linking)

Table 15: Search and retrieval

Feature	DSpace	EPrints	Fedora
Full text search	YES	YES (With some changes in code base)	N/A (Generic Search Service, which is a part of the Fedora Service Framework supports full-text search
Standard Boolean search	YES	N/A	YES
Federated search server	N/A	N/A	N/A
Search configuration by user	N/A	N/A	N/A
External search engine*	YES	YES	YES
Brows by	Author, Title, Collection, Year (Extensible)	Author, Title, Collection, Year, Subject, Academic Unit (fixed)	Author, Title, Collection, Year, Subject, Academic Unit (extensible)

* It is a time-consuming process to test which repository could be better indexed through Googlebot, and needs a long-time observation of several repositories. However, based on short-term observations, it seems that, Google can crawl the DSpace archives easily, and indexes more number of DSpace-based records than others (probably due to the straight-forward mechanism of assigning URLs to records and collections). It also seems that EPrints and Fedora, in this regard are in the next levels respectively.

Advantages and disadvantages of evaluated OSS

This section involves the general results of the evaluation of softwares' functionalities achieved through the process of implementation and usage of DSpace, EPrints, and Fedora.

Strengths of DSpace

- Manageability and customizability of workflows are more flexible and more developed than other packages
- More structured data model
- Ability to address the long-term preservation of digital objects
- Security related functions in DSpace is more progressive than other solutions
- More capable of integration with other information systems
- Could be considered as an easy-to-use and low-cost solution for a wide range of institutions
- Installation and maintenance steps (backup, recovery, import, update, etc.) are explicit and can be conducted easily
- An active community of experienced users and developers supports the system development
- Expectancy of future expansion
- Good choice for quick deployments
- Installation, administration, and maintenance processes are well-documented.

Weaknesses of DSpace

- Code base of DSpace is complex to be modified for low level customizing.
- The scalability challenge is a major risk of selecting DSpace for sustainable ETD repositories. In fact DSpace consists of many tools and applications and its structure and code base is complex for being specialized for new or special needs of an institution.
- The developer team of DSpace has decided to re-architect this software for next year and it may affect variety add-ones that are based on the current version.

Strengths of EPrints

- Several separated installations could be merged to establish an integrated ETD repository.
- Robust and stable system that needs minimum maintenance
- Less complex compared to others
- Complete solution for handling wide range of pre-print and post-print research documents
- Many plug-ins are available to improve different functionalities of EPrints.
- A uniform and well-documented code base makes it easier to work on for low level customization.

Weaknesses of EPrints

- Indexing process is slow compared with other packages.
- The number of file formats that are supported by default is limited.

Strengths of Fedora

- More flexible and extensible solution compared to other systems and a good choice for managing complex objects and relationships
- The infrastructure consists of two main components: the storage system that is placed in the back-end and the interface for users and administrators in front-end. This system architecture makes Fedora a flexible and extensible solution.
- Availability of progressive Application Programming Interface (API) makes it possible for various types of client applications to establish a conversation to Fedora and request services.

Weaknesses of Fedora

- Fedora is a web service in its nature and needs a number of additional web applications and tools for being utilized as a complete ETD management system (e.g. FEZ)
- In terms of implementation, Fedora is the most complex system among compared solutions because of its complicated steps of installation, which requires an expert administrator for implementation.
- In comparison with other systems, Fedora is not well documented.

Conclusion

The evaluation of the three open source IR solutions depicted that, in the most of interoperability and integration-related functions, DSpace was more comprehensive and ready-to-be-used solution for ETD repositories. It was also observed that, in the most of the comparative areas that are critical for ETD management system such as metadata enhancement features, compatibility with interoperability protocols, complex submission

workflow, and multi-file deposit structure, DSpace is ahead of EPrints and Fedora. DSpace also has the largest number of installations and takes advantage of the most open development and support community.

The evaluation also demonstrated that in a number of cases, there was not any great distinction among packages (e.g. in terms of OAI-PMH compliancy). Additionally, in few features, EPrints and Fedora were assessed as reasonable choices for implementing ETD repositories. For instance, in terms of the important function of document versioning, EPrints was more complete than the others and was far ahead of DSpace. However, these few features are not those that directly affect the essential ETD-specific functions of repositories.

Finally, it is worth noting that all of packages are suitable for implementing sustainable ETD repository systems and in a number of cases, Fedora and EPrints are utilized for handling a huge number of ETDs in addition to the other types of research documents.

Further Works

Our selected criteria for comparing IR software solutions were mostly based on unique characteristics of ETDs. However, in most cases, institutions utilize specific IR software for managing different types of scholarly documents (including ETDs). Providing a comprehensive and extensible framework for assessing wider range of IR software to be utilized in different fields (not just ETDs), could help universities and institutions to select the most appropriate and best adapted solution for their special needs and policies. Furthermore, although the three evaluated software are broadly utilized to manage not just ETDs, but also various type of IRs, forobtaining a reliable results, the other available solutions including hosted service and commercial packages should be involved in a more comprehensive comparison study.

REFERENCES

1. Lynch, C.A. and J.K. Lippincott, *Institutional Repository Deployment in the United States as of Early 2005*. D-Lib Magazine, 2005. **11**(9).
2. Chawner, B., *F/OSS in the library world: an exploration*, in *Proceedings of the fifth workshop on Open source software engineering*2005, ACM: St. Louis, Missouri. p. 1-4.
3. Mckay, D. and S. Burriss, *Improving the Usability of Novel Web Software: An Industrial Case Study of an Institutional Repository*, in *Proceedings of the 2008 international workshops on Web Information Systems Engineering*2008, Springer-Verlag: Auckland, New Zealand. p. 102-111.
4. Westrienen, G.v. and C.A. Lynch, *Academic Institutional Repositories*. D-Lib Magazine, 2005. **11**(9).
5. Goh, D.H.L., et al., *A checklist for evaluating open source digital library software*. Online Information Review, 2006. **30**(4): p. 360-377.
6. Jose, S. *Adoption of open source digital library software packages: A survey*. in *International Convention on Automation of Libraries in Education and Research Institutions (CALIBER)*. 2007.
7. Kaczmarek, J., et al., *Using the Audit Checklist for the Certification of a Trusted Digital Repository as a Framework for Evaluating Repository Software Applications*. D-Lib Magazine, 2006. **12**(12).
8. Powell, A., *Notes about possible technical criteria for evaluating institutional repository (IR) software*. 2005.
9. SWITCH, *Evaluation of an Open-Source Repository System*, 2008.
10. RSP, *Repository Software Survey November 2010*, 2010, Repositories Support Project
11. Goncalves, M.A., et al. *MARIAN: Searching and querying across heterogeneous federated digital libraries*. in *First and DELOS Network of Excellence Workshop on Information Seeking, Searching and Querying in Digital Libraries*. 2000. Zurich, Switzerland.
12. Mittal, R. and G. Mahesh, *Digital libraries and repositories in India: an evaluative study*. Program: electronic library and information systems, 2008. **42**(3): p. 286 - 302.
13. Primary Research Group, *The International Survey of Institutional Digital Repositories*, 2007, Primary Research Group.
14. Jose, S. *Adoption of Open Source Digital Library Software Packages: a Survey*. in *Proceedings CALIBER 2007: 5th International Convention on Automation of Libraries in Education and Research Institutions*. 2007. Chandigarh, India.
15. Gibbons, S., *Featurs and functionalities*. Library Technology Reports, 2004.
16. Crow, R., *A Guide to Institutional Repository Software*, 2004, Open Society Institute.

17. Wyles, R., *Technical Evaluation of Research Repositories*, in *Research Repositories in New Zealand*2006.

18. NLM, *Recommendations on NLM Digital Repository Software*, 2008, National Library of Medicine.

19. RSP, *Repository Software Survey March 2009*, 2009, Repositories Support Project

20. Roberts, C. *Choosing an IR Platform*. in *INSTITUTIONAL REPOSITORIES (IR) : a workshop on creating an information infrastructure for the scholarly community*. 2007. Johannesburg.

21. Blustein, J., *Automatically generated hypertext versions of scholarly articles and their evaluation*, in *Proceedings of the eleventh ACM on Hypertext and hypermedia*2000, ACM: San Antonio, Texas, United States. p. 201-210.

22. Goncalves, M.A., et al. *MARIAN: Searching and querying across heterogeneous federated digital libraries*. in *Workshop on Information Seeking, Searching and Querying in Digital Libraries*. 2000. Zurich, Switzerland.

23. Tennant, R., *Institutionalrepositories.* LibraryJournal, 2002. **127**(15).

24. Koch, T. *Electronic thesis and dissertation services: Semantic interoperability, subject access, multilinguality*. in *E-Thesis Workshop*. 2006. Amsterdam

25. Hoover, L. and R.E. Wolverton, *Cataloging and Treatment of Theses, Dissertations, and ETDs.* Technical Services Quarterly, 2003. **20**(4): p. 3-57.

26. Taylor, C., *Criteria for choosing repository software* 2004, University of Queensland.

27. RUBRIK, *Repository Software*, 2006, Regional Universities Research Infrastructure Collaboratively: Sidney, Australia.

28. Jones, R., *The Tapir: Adding E-Theses Functionality to DSpace*. ARIADNE, 2004. **41**.

29. EPrints. *Open Access and Institutional Repositories with EPrints*. 2009 [cited 2009; Available from: http://www.eprints.org.

30. Barve, S. and A. Prasad, *An Evaluation of Open Source Software: for Building Digital Libraries*2010, Berlin: VDM Verlag.

31. McCutcheon, S., et al., *Morphing metadata: maximizing access to electronic theses and dissertations.* Library Hi Tech, 2008. **26**(1): p. 41 - 57.

32. Needleman, M., *The Shibboleth Authentication/Authorization System.* Serials Review, 2004. **30**(3): p. 252-253.

33. Howes, T.A., M.C. Smith, and G.S. Good, *Understanding and Deploying LDAP Directory Services*2003, New York City: Addison-Wesley Professional.