Jimma University
College of Natural Science
Department of Information Science
Postgraduate Program
Electronic and Digital Resource Management

Research Project

ENHANCING DIGITAL INFORMATION INTEROPERABILITY AND USER INTERACTION AMONG ETHIOPIAN HIGHER EDUCATION DIGITAL LIBRARIES (EHEDLS)

By:

Ziena Ghidey

Jimma, Ethiopia

November, 2014

Jimma University
College of Natural Science
Department of Information Science
Postgraduate Program
Electronic and Digital Resource Management

Research Project

ENHANCING DIGITAL INFORMATION INTEROPERABILITY AND USER INTERACTION AMONG ETHIOPIAN HIGHER EDUCATION DIGITAL LIBRARIES (EHEDLS)

By:

Ziena Ghidey

Principal Advisor: Michael B. Spring (PhD,)

(University of Pittsburgh, USA)

Co-Advisor: Kedir Mohammed (Msc)

(Jimma University, Ethiopia)

DECLARATION

This Independent research project en-titled *"Enhancing digital information interoperability and user interaction among Ethiopian higher education digital libraries (EHEDLs)"* is my original work and all the sources of materials used for the thesis have been duly acknowledged.

____

Zena Ghidey Alemayehu

November, 2014

This thesis has been read and approved as the requirements of the Department of Information Science in partial fulfillment for the award of the Master of Science in Electronic and Digital Resource Management (EDRM), Jimma University, Jimma, Ethiopia.

Principal Advisor:

Michael B. Spring ______________________

Co-advisor:

Kedir Mohammed ______________________

Examiner:

____

ACKNOWLEDGMENTS

I would like to thank many people for their kindness and assistance during this study. I sincerely thank my advisors Michael B Spring and Kedir Mohammed for their guidance and encouragement. Special thanks go to **Spring** whose brilliance inspires me in all that I do. I owe a lot for his patience, friendly approach, invaluable comments and heartwarming personality. He spent his precious time in commenting my work and showing me the right directions that I found very important for the accomplishment of this thesis. I truly appreciate the support of AAU, ASTU, JU, MU and BDU digital librarians in this study one way or another.

Above all, I thank God who surrounded me with so many wonderful people.

Contents

LIST OF ABBREVIATIONS

AAU – Addis Abeba University

ASTU – Adama Science and Technology University

CGI – Common Gateway Interface

CORBA – Common Object Request Broker Architecture

DCMI – Dublin Core Metadata Initiatives

DL – Digital Library

DTD – Document Type Definition

EHEDLs – Ethiopian Higher Education Digital Libraries

EHIDL-AP – Ethiopian Higher Institutions Digital Library – Application Profile

EHIDLs – Ethiopian Higher Institutions Digital Libraries

ETGLS – Ethiopian Government Locator Service

GNU – Gnu's Not Unix!

GSDL – Greenstone Digital Library

HERQA -- Higher Education Relevance and Quality Agency

HTTP – Hyper Text Transport Protocol

JU – Jimma University

METS – Metadata Encoding and Transmission Standard

MODS – Metadata Object Description Schema

MOE – Ministry of Education of Ethiopia

MU – Mekele University

NISO – National Information Standards Organization

OAI – Open Archive Initiatives

OAI-ORE – Open Archive Initiative- Object Reuse and Exchange

OAI-PMH – Open Archive Initiative – Protocol for Metadata Harvesting

PKP OHS – Public Knowledge Project Open Harvest System

RDF – Resource Description Framework

SDLIP – Simple Digital Library Interoperability Protocol

SGML – Standard Generalized Mark-up Language

SOAP -- Simple Object Access Protocol

SRW/U -- Search/Retrieve Webservice/ by URL

SWORD – Simple Web-service Offering Repository Deposit

URI – Uniform Resource Identifier

WWW – World Wide Web

XML – eXtensible Markup Language

XSLT – eXtensible Stylesheet Language Transform

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

When currently some Ethiopian Higher Educations are implementing digital libraries for their actual clients and interoperability is one of the advancement features of digital libraries, extending this service is this study purpose. Therefore this study identifies Ethiopian Higher Education Digital Libraries tools they used; service they provide as well as identifies stakeholders, developers and end users so that discussing three interoperability approaches used by digital libraries. Because of simplicity and suited for current EHIDLs over other protocols OAI-PMH metadata harvesting approach is deployed to EHIDL. Thus first we proposed EHIDL-AP metadata schema; then for DSpace and Greenstone digital library tools demonstration about how to use EHIDL-AP schema and implement data provider also presented; those metadata created by those institutions service provider using PKP OHS tools will harvest, normalize and make available for EHIDL users. Service provider (federated system) of EHIDL architecture, installation overview, adding EHIDL-AP schema, adding data providers, harvesting metadata, searching and browsing functions also discussed. Finally the study conclusions presented with a long side of recommendations for future works to be done to maximize the semantic level of interoperability among EHIDLs.

CHAPTER ONE

INTRODUCTION

1.1 Background of the study.

Universities help to shape national capability and interaction in an increasingly competitive global world. Institutions of higher education, especially those in Africa, are core institutions that link nations to the emerging global forces through the knowledge domain. As digital resources become more available, scholars will demand more and more digitally enabled resources and services. Most African universities do not yet have sufficient infrastructure to utilize digital resources (Gelaw, 2006).

The university environment in Ethiopia is changing. There is renewed recognition of the role that universities play as drivers of national development. Their transformation has included investment in electronic infrastructure and connectivity as well as attention to e-learning and related approaches as key tools to enhance the quality of higher education and make it more accessible.

The government currently pursues an economic growth strategy based on agriculture-led development. This thrust is complemented by efforts to enhance overall labour productivity through better education and health services, to foster an emergent private business sector, and to reform aspects of the civil service. For this strategy to be successful, the country's higher education system will have to produce graduates with the technical knowledge and research skills to support economic diversification. This is one of the drivers of the present reform (World Bank, 2003).

Ethiopian higher education proclamation 650/2009 states "it has a curriculum that match the national standards set by the Ministry, the necessary academic staff, institutional governing structures as provided for by this Proclamation, teaching materials, classrooms, libraries, laboratories, and other appropriate discipline-related facilities"(part two section 11 sub section d). Ethiopian Libraries must demonstrate their value and document their contributions to overall institutional effectiveness and be prepared to address changes in higher education. The government oversees the importance of modern libraries in the provision of user needed services.

Interoperability is becoming a paramount issue as the internet unites digital library systems of differing types, run by separate organizations which are geographically distributed all over the world. Federated digital library systems, in the form of cooperating autonomous systems are emerging in a bid to make distributed collections of heterogeneous resources appear to be a single, virtually integrated collection.

The benefits to users include query processing over larger, more comprehensive sets of resources as well as the promise of easier to use interfaces that hide systems, syntax and structural differences in the underlying systems (DELOS, 2005).

Even though some universities of Ethiopia implement library automation project to support the overall managerial function of the library and digital library project to manage electronic and digital resources they have using open source softwares and tools. There are no federated based initiatives done to maximize interoperability with them. Therefore this study tried to assess their digital library management to design and implement federated system to Ethiopian Higher Institutions.

1.2 Statement of the Problem

Ethiopia has embarked on a higher education expansion and reform programme of impressive dimensions. Expansion will create new universities, establish three system support agencies, mount new courses, and triple enrollments. Reforms introduce increased institutional autonomy and curriculum revisions (Saint, 2004). Libraries are expected to provide information resources and services with appropriate technologies to meet information needs.

While digital repositories and libraries have been around for some time, the practice of building digital repositories/libraries in Ethiopia is a recent phenomenon. This has been due to lack of relevant IT skills, the cost of commercial technologies, etc. Recent interest from many professionals in the area and the proliferation of free open source technologies to build and manage local resources, has led to a number of initiatives by higher education and research institutions. (ADLSN n.d.)

Interoperable systems broaden choice and open up new perspectives for researchers, governments and citizens across a spectrum of disciplines and domains. Interoperability is key to improve digital libraries, enabling wider collaborations and ensuring that a broader spectrum of resources are available to a wider range of people whether for simple consumption or to enhance research activities. The importance of interoperability is well known and many attempts have been made in the past to provide solutions that enhance interoperability (DL.org n.d.).

Interoperability is a broad problem domain. It is typically investigated within a specific scope, such as within a particular community (e.g., libraries, commercial entities, scientific communities), within a particular classification of information (e.g., electronic records, technical reports, software), or within a particular information technology area (e.g., relational databases, digital imaging, data visualization).

Current research on interoperability in digital library architecture addresses the challenges of creating a general framework for information access and integration across many of the above domains. A common goal of these efforts is to enable different communities, with different types of information and technologies, to achieve a general level of information sharing and, through the process of aggregation and computation, to create new and more powerful types of information (Payette, 1999).

At national level, the national information policy draft paper stated a goal of transforming traditional library systems into a network of electronic libraries. An effort to expand the reach of the currently limited reference materials available to Ethiopian university students and academics, the Ministry of Education is finalizing preparations to launch a nationwide electronic library system (university world news, 2009). However, the project is done by independent company there is no work has done to integrate universities libraries resources, instead of focusing on infrastructure installation and electronic book enriching on the central server.

As Ethiopian higher education institutions initiate projects on digital libraries and repositories, there is little understanding of how to make systems interoperable to allow for sharing resources and services at a national level. The focus of this research was three folds:

- To identify available digital repository standards available and in use in Ethiopian institutions so as to identify the best array of interoperable protocols that will allow optimal acquisition and sharing of digital resources to reduce costs and increase the number of needed resources critical to the realization of national goals for knowledge and information sharing.
- To identify stakeholders, developers, and end users of a coordinated digital repository.
- Based on the above findings design and implement a coordinated system with the expectations of meeting the needs of users and stakeholders missions of EHEDLs. Therefore maximizing resources and service sharing among respected libraries.

For effective collaboration of Ethiopian digital libraries a critical evaluation on digital library standards employed, services provided, intended clients, the technologies they used, and their flexibility to integrate with other systems was done. Choosing among different interoperability protocols requires an analysis of ease of use for the developers as well as the best fitted to help the country higher institution quality assurance of the education.

1.3 Objectives of the study

1.3.1 General Objective

The main objective of this study was to identify the best set of interoperability protocols based on current Ethiopian higher institution digital repositories standard used; and also design and implement a coordinated system for utilization of digital resources considering the stakeholders and end user behavior.

1.3.2 Specific Objectives

- To identify digital information resources and services provided to university community as well as the nation.
- To identify the best array of protocols that will allow optimal acquisition and sharing of digital resources to reduce costs and increase the number of available resources.
- To identify stakeholders, developers, and end users of a coordinated digital repository.
- To design a coordinated system that will maximize utilization of the system.
- To implement a coordinated system to EHEDLs for interoperability of their resources.
- To provide information to librarians, researchers and university community that will improve the use of the digital repositories.

1.4 Significance of the Study

Data from many countries show a positive correlation between increasing higher education access and economic growth as expressed by increasing per capita income and/or human development index (UNESCO/OECD, 2003). Implementation of development strategies and policies will succeed if higher education institutions through their functions of teaching, training, research and services play their essential roles as a factor in sustainable development. As an information service provider, the library has an important role in assuring quality education in the national level will take place.

Digital information resources are increasingly the backbone of an academic library and so integration, collaboration with other related system and organization is crucial. This study addresses the needs of individual group of community (both undergraduate and postgraduate students, instructors, researchers, as a whole university community) and the nation's citizens more broadly.

No single university library collection can fulfill all the needs of its users. As Ethiopian universities are expanding, and new universities are established it is increasingly important to share information resources the ***dominant universities*** already have.

As more universities engage in research, there is a growing need for transparency and access to the results. Today, their findings have limited disseminations to the relevant community (Yizengaw, 2004). Accessing up-to-date information resources across to the whole institution's members is critical. So that this study plays an important role in increasing the availability of this information.

The cooperative efforts of libraries will create cost savings and efficiencies of scale, bring wider recognition for libraries, and provide cooperative intelligence for better decision-making, and provide the platform on which libraries can innovate.

1.5 Scope of the study

This study was done only for Ethiopian governmental higher education institutions which are under one government supervision. This is because of public universities are under in one division (Ministry of Education), digital resources legal rights and other issues could be handled with the supreme of the MoE. While the effort will be limited to these institutions, we made note of what Universities in other parts of the world have done to increase interoperability and we also share the results of this study widely both within Ethiopia and with other African national groups.

This study was not create or dictate new digital libraries for universities. Instead it helps universities who have already built or are planning their digital libraries to identify the best protocol to enhance interoperability. The study also identifies stakeholders, developers, and end users of coordinated digital repositories with an eye to designing and implementing a coordinated system based on the suited identified protocols.

CHAPTER TWO

LITERATURE REVIEW

This chapter presents literature review related on the topic of Ethiopian higher institution digital libraries (EHIDLs), digital library interoperability, and digital information user interaction regarding current issues and trends aspects.

2.1 Ethiopian Higher Institutions Digital Libraries

Studies by the Organization for Economic Co-operation and Development (OECD, n.d.) confirm that an even greater positive correlation exists between the quality of schooling provided and increased levels of economic growth. Therefore, it is clear that education, especially high quality education, will be a major factor in Ethiopia's goal of achieving real and long lasting economic growth. In Ethiopia, quality of education is the single biggest challenge and priority of the MOE, as well as all donors and other stakeholders. The concern over low quality education exists from the pre-primary levels up to tertiary education. United States alone has invested $200 million in the education sector in Ethiopia over the past fourteen years (USAID, 2012).

Getachew & Gojeh, (2007) and Tadesse & Gojeh, (2007) studies on libraries or information centers in Ethiopia; conducted in the public sector and Universities, revealed that there is a dire need for qualified and experienced professional library and information center staff at higher levels to head and provide effective and efficient library and information services in the public sector and University libraries in Ethiopia respectively. However in how way the librarians are implementing this issue never described.

A global rise in the price of periodicals coupled with devaluation of local currency resulted in the need for more funds to subscribe even to the same periodicals libraries have. Budget reductions make it difficult even to sustain subscription to the journals they currently receive. Gelaw & Kelly (2006) pointed out that advancement of digital technologies is shaping creation, access, use and preservation of information resources in ways that are so profound that traditional methods and concepts of access and organization are no longer effective. Web applications, together with the growing numbers of digital library initiatives, are making resource integration much easier and are providing scholars with access to more diverse information sources and services. Digital libraries and supporting technologies have now matured to the point where their contents are incorporating complex and dynamic resources and services. In reviewing the literature of the past few years, there is no shortage of views on the role of digital libraries.

By knowing the significance of digital libraries, some universities of Ethiopia are under the process of building digital library and integrating their library automation using open source software. Addis Ababa University already launched in webspace for electronic dissertations done by its students and faculty members, others launch their digital library to be accessed locally for their university community only.

2.2 Digital Library Interoperability

NISO, (2004) define interoperability, as the ability of multiple systems, using different hardware and software platforms, data structures, and interfaces, to exchange and share data. Interoperability is also defined as the compatibility of two or more systems such that they can exchange information and data and can use the exchanged information and data without any special manipulation (Taylor 2004).

Interoperability is one of the most heavily discussed issues in digital library research (Shiri, 2003). It is becoming a paramount issue as the internet unites digital library systems of differing types, run by separate organizations which are geographically distributed all over the world. Federated digital library systems, in the form of cooperating autonomous systems are emerging in a bid to make distributed collections of heterogeneous resources appear to be a single, virtually integrated collection. The benefits to users include query processing over larger, more comprehensive sets of resources as well as the promise of easier to use interfaces that hide systems, syntax and structural differences in the underlying systems (DELOS, 2005). Varatharajan & Chandrasekhara, (2006) also discuss about the importance of interoperability of digital libraries.

2.2.1 Types of Interoperability

Interoperability occurs at different levels: from the bit stream layer up to semantic interoperability. In these subsection types of interoperability are described from the basic level to the highest abstraction level.

a) Technical Interoperability

Technical interoperability may be ensured if two systems follow the same technical specifications for processing an identifier string, where the scope of the likely identifiers to be encountered is reasonably predictable. In certain cases, rules may exist for directly incorporating an identifier from one scheme in the syntax of another scheme.

b) Semantic Interoperability

Semantic interoperability refers to the meaning of information to its human users, as opposed to the simple physical transfer of data. Interoperability at this level can fail if different users, or groups of users, use different terms for the similar concepts, or use similar terms to mean different things.

c) Political/Human Interoperability

While crossing organization to organization the interoperability has to be discussed and agreement signed properly. The participating organizations have to establish a discussion forum to facilitate sharing of information in interoperability and related issues.

d) Inter-community Interoperability

Many factors are contributing to the blurring of boundaries between communities. Digital libraries in different communities have to be specified clearly while crossing the community's access and limitations. As a result it is increasingly important that information systems be designed to interoperate across these boundaries. In the area of resource discovery, one of the main mechanisms for facilitating this interoperability is metadata standard and harvesting system which provides for consolidated resource discovery across all inter-communities.

e) Legal Interoperability

While the Internet makes it easy to physically publish and access information, there are many important legal aspects which constrain and influence how information can and should be made available and used. These include laws related to copyright, content regulation, privacy, freedom of information, telecommunications regulation, e-commerce and trade practices. Activities which may be legal in one context or jurisdiction may not be permitted in another.

f) International Interoperability

Each of the key issues identified, above, is magnified when considered on an international scale, where differences in technical approach, working practice, and organization have evolved over many years. Online technologies facilitate access to resources from anywhere in the world, and make resources available to an international audience. However, this brings with it a need to ensure that interoperability issues are addressed in an international as well as particular country level.

The Digital Library Technological and Methodological Cookbook (2011) presents a rich array of best practices and pattern solutions to common interoperability issues faced when building interoperable Digital Libraries of Europeana. There are so many interoperability protocols and projects are doing on very different abstraction levels. Discussing some of the major initiatives and projects which are focusing on digital information or libraries is the major task of this research even though the scope is limited for EHIDL's.

Z39.50

The Z39.50 protocol is an application layer protocol that supports distributed search and retrieval between structured network services. This protocol stipulates data structures and interchange rules that allow a client machine to search and retrieve records from databases on a server machine, across different platforms. It is widely used by librarians and very often integrated into library systems and personal bibliographic reference.

Open Archives Initiatives (OAI)

In October of 1999 the Open Archives Initiative (OAI) was launched in an attempt to address interoperability issues among the many existing and independent DLs. The focus was on high-level communication among systems and simplicity of protocols. The OAI has since received much media attention in the DL community and, primarily because of the simplicity of its standards, has attracted many early adopters. (Gowda & Bhandi, 2006)

Open Archives Initiative - Protocol for Metadata Harvesting (OAI-PMH)

The OAI-PMH is first of all a protocol which underlies an architecture based on metadata harvesting. And It was designed to be easy to implement (based on widely accepted standards such as HTTP, XML and Dublin Core) and highly efficient.

The OAI-PMH defines various roles in an architecture built on metadata. **A data provider** makes its metadata available for use in one or more description formats. **A service provider** launches a programme called harvester to visit a data provider and collect metadata in the format it requires, if this is available, at least in unqualified Dublin Core. **An aggregator** gathers metadata from various data providers and makes them available in an OAI repository.

OAI-PMH only manages data transfer; it is not a cross-searching protocol since it does not support querying functionalities. However, querying can be processed when making the OAI repository or within the OAI service by rebuilding a finding aid on the server of the service provider. Cross-searching functionalities may be added and used as complementary functions to an OAI architecture.

Open Archives Initiative – Object Reuse and Exchange (OAI-ORE)

Open Archives Initiative Object Reuse and Exchange (OAI-ORE) defines standards for the description and exchange of aggregations of Web resources to expose the rich content in these aggregations to applications that support authoring, deposit, exchange, visualization, reuse, and preservation. OAI-ORE is not the replacement of OAI-PMH. OAI-PMH will continue to exist as one approach to interoperability. OAI-ORE will complement with richer functionality, when this is desirable.

The OAI-ORE data model is built on the foundation of the WWW architecture, RDF, Cool URIs and Linked Data. This abstract data model includes Aggregation, Aggregated Resource, Resource Map (ReM) and a Proxy.

ResourceSync

The Web is highly dynamic, with resources continuously being created, updated, and deleted. As a result, using resources from a remote server involves the challenge of remaining in step with its changing content. In many cases, there is no need to reflect a server's evolving content perfectly, and therefore well established resource discovery techniques, such as crawling, suffice as an updating mechanism. However, there are significant use cases that require low latency and high accuracy in reflecting a remote server's changing content. These requirements have typically been addressed by ad-hoc technical approaches implemented within a small group of collaborating systems. There have been no widely adopted, web-based approaches.

NISO and OAI have recently launched the ResourceSync project that aims at designing an approach for resource synchronization that is aligned with the Web Architecture and that is targeted at the needs of different communities.

Simple Web-service Offering Repository Deposit (SWORD)

SWORD was funded to take the deposit API activity into a more formally funded project, to ensure that the ideas and enthusiasm already captured could be used to produce concrete outputs. Led by UKOLN, the project was a partnership between CASIS at Aberystwyth University, the University of Southampton and Intrallect. The project aims were simple – to agree on a protocol or specification for deposit, to implement a deposit interface into DSpace, Fedora, EPrints and IntraLibrary and to produce a prototype, "smart" deposit client for testing the Implementations. SWORD is a lightweight protocol using profile of the Atom Publishing Protocol.

Search/Retrieve Webservice/ by URL (SRW/U)

SRW, the Search/Retrieve Webservice, is an XML oriented protocol designed to be a low barrier to entry solution to performing searches and other information retrieval operations across the internet. It uses existing, well tested and easily available technologies such as SOAP and XPath in order to perform what has been done in the past using proprietary solutions. The design has been informed by 20 years of experience with the Z39.50 information retrieval protocol.

The protocol has two ways that it can be carried, either via SOAP or as parameters in a URL. This second form is called SRU -- SearchRetrive by URL. Other transports would also be possible, for example simple XML over HTTP, but these are not defined by the current standard.

The Dienst Protocol

Dienst is an architecture and protocol for digital libraries across multiple servers. Initially called the Computer Science Technical Report Project (ARPA) funded project originated from the need to create a digital library of Computer Science technical reports.

The protocol supports an info service, an index service, a repository service, a query mediator service, collection service, a registry service and additionally there is a user interface service for interaction between the above mentioned services and their protocols.

The services are defined individually, and when combined they create a distributed digital library that provides functionality for deposit and storage of digital resources, as well as access to those resources by discovery and browsing.

Simple Digital Library Interoperability Protocol (SDLIP)

SDLIP is a protocol that defines simple interfaces for interoperability between data Providers. It was part of a Stanford University project called Stanford Digital Libraries Technologies. SDLIP's main goals were the simplification of both client and server side implementations; server support for stateful and stateless operations; dynamic load balancing for server; support for thin clients; and implementations via both distributed object technology (CORBA and HTTP/CGI).

SDLIP operations are divided across three interfaces, namely: the *Search interface* that allows the submission of search queries, the *Result Access* interface that allows the client applications to access results from a search request and the *Source Metadata interface* that allows clients to question a library service proxy about its capabilities.

2.3 Digital information user interaction

The effect of information sources structure have been acknowledged in the information behavior literature as one of the many intervening variables in the information seeking process, that can be ‘‘supportive of information use, as well as preventive’’ (Wilson, 1999).

Sastry & Reddy (2009) pointed out digital libraries are inherently interactive systems with a constant growth in the number of end users. They must not only rely on effective and sophisticated retrieval mechanisms but also provide efficient interaction with the end users.

Computer and information scientists should be among the first to experiment with digital libraries. The Association for Computing Machinery (ACM), as well as other associations and publishers, are becoming involved in Project Envision, a research effort supported by the National Science Foundation to build “a user-centered database from the computer science literature” (Brueni et al., 1993). The user interaction with content and system has been used in the area of information architecture for the purposes of effective DL design. Toms (2002) has used this tripolar structure as a platform, where users’ information seeking behavior takes place, and upon this platform proposes ways of extracting meaningful design suggestions. Several empirical studies have demonstrated that usability problems, i.e. design inefficiencies, can disrupt user’s information searching activity and affect information task accomplishment. Xie (2006) observed users like to apply the least effort principle to finding useful information to solve their problems.

In order to develop usable digital libraries and to improve system design, researchers have addressed user behavior and user requirements in different contexts including academic environments, schools, government departments and business.

Content in the 'hidden Web' needs a specific set of user interactions in order to access it and such access is difficult to automate. Some headway has been made with this problem by attempting to replicate these human actions with software agents that can detect HTML forms and learn how to fill them in, using what are known as hidden Web agents (Masanès,2006). One alternative requires direct collaboration with a site’s owner, who agrees to expose the full list of files to an archive process through a protocol such as OAI-PMH. Another alternative, which saves the site’s owner from setting up a protocol and which is useful for websites that offer a database gateway which holds metadata about a document collection, is to extract (deep mine) the metadata directly from the database and archive it, together with the documents, in an open format. In effect, the database has been replaced, at the archive, by an XML file.

Rao, (1995) stated that effective information access involves rich interactions between users and information residing in diverse locations. They have developed a number of techniques that support various aspects of the process of user/information interaction.

Regardless of users' information technology (IT) backgrounds, their expectations of digital libraries' functionality are the same. However, based on users' previous experiences with digital libraries, their requirements with respect to specific features may change. So involving users in digital library design should be an integral step in the process of building a digital library in addition to the classic roles of evaluation and testing (Kani-Zabihi et al. 2006). A key part of good design is tailoring the evaluation to fit the particular circumstances and to fit information needs of the primary audiences for the study and address a real, known need (Feinstein, 2002; Mathison, 2001; Patton, 1997).

In the DL field, interoperability as mentioned by Suleman in his theses refers to the ability of a DL to work cooperatively with other DLs in an attempt to provide higher quality services to users (Suleman, 2002). The goal of interoperability factually is to build coherent services for users, from components that are technically different and managed by different organizations. So, as Paepcke et al. (1998) expressed in their article, interoperability is a central concern in building DLs (Paepcke et al., 1998). Accordingly, interoperability is one of the most important issues in DLs and should be of concern at the time of building or improving DLs.

Federal programmatic support for digital library research was formulated in a series of community-based planning workshops sponsored by the National Science Foundation (NSF) in 1993-1994. The first significant federal investment in digital library research came in 1994 with the funding of six projects under the auspices of the Digital Libraries Initiative (now called DLI-1) program. These DLI-1 research and development projects were jointly funded by a federation comprised of the National Science Foundation (NSF), the National Aeronautics and Space Administration (NASA), and the Defense Advanced Research Projects Agency (DARPA). (Griffin, 2000).

In 1998, at the cessation of the DLI-1 program, federal funding for the DLI-2 program was instituted with support from NSF, NASA, DARPA, the National Library of Medicine (NLM), the Library of Congress (LC), the Federal Bureau of Investigation (FBI), and the National Endowment for the Humanities (NEH). Also, in 1998, the Corporation for National Research Initiatives (CNRI), under DARPA support, funded the three-year D-Lib Test Suite program which provided continuing funding for several of the digital library Testbeds created under DLI-1. In aggregate, between 1994 and 1999, a total of $68 million in federal research grants were awarded under DLI-1 and DLI-2 (Fox, 1999).

The OAI protocol (OAI, 2002) is the most widely discussed and investigated standard for cross repository interoperability Gowda & Bhandi (2006) shows a debate why not using Z39.50 protocol which achieves federated searching and they come up with many reasons why another protocol is needed.

Gradmann on interoperability challenges on the digital libraries, (2009) defined the characteristics of Interoperability aspects of Selected Frameworks for DL modeling of DELOS, DRIVER, OAI-ORE, DCMI abstract, JISC Information Environment, JCR and iRODS. Paepcke et al (1998) discuss the interoperability space and its solutions.

Suleman, (2002) design an Open Digital Library (ODL) based on OAI-PMH. Paihama et al (2012) also evaluate user understanding of Xsearch vs. SRU and Xharvester vs. OAI-PMH by 27 computer students at university of Cape Town. They hypothesized that simpler interoperability protocols and standards will lead to an increased level of adoption by making it is easier for programmers to understand and implement them and therefore leading to more interoperable systems.

The point is for Ethiopian higher institutions to be part of this advantageous technological enhancement carefully selection of interoperable protocol is needed to meet the national mission.

CHAPTER THREE

RESEARCH DESIGN AND METHODOLOGY

Research in the information systems field examines the technological system and the social system; in addition, it investigates phenomenon that emerge when the two interact (Lee, 2001). For the proper works to be done this research used design science research. Hevner & Chatterjee (2012) defined design science research as a research paradigm in which a designer answers questions relevant to human problems via the creation of innovative artifacts.

This study focuses on identifying the best array of interoperable protocol from those used in the current Ethiopian Higher Institutions Digital Libraries (EHIDLs).; It also examines services they provide with accordance of suited to stakeholders, developers and end users of the anticipated coordinated system.

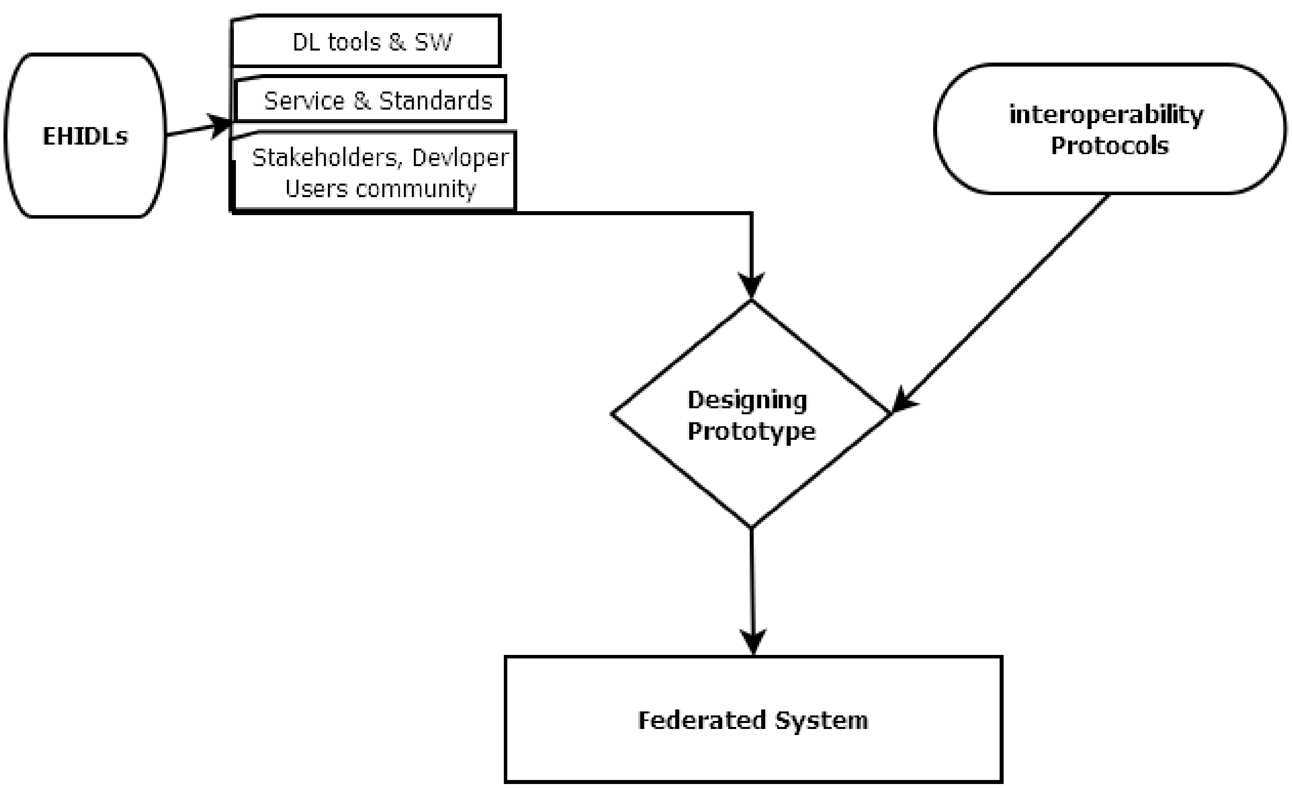


Figure 1 Diagrammatic Overview of the Overall Research Design & Methodology.

The diagram shows that first of all Ethiopian Higher Institutions Digital Libraries evaluations that *include DL tools or software they used; identify their DL standard they used; types of service they provide; and identify their DL stakeholders, developers and end users were done.*

Above all describing the three interoperability approaches with their demerits also done. And based on discussions as well as the identified DL standards and services; and the identified stakeholders, developers and end users consideration, thus selecting OAI-PMH harvesting approach interoperability protocol was done. Then designing a prototype that describes how components of the interoperability architectures should be looks like with demonstrations is presented. Finally federated (coordinated) system for EHIDLs users is demonstrated.

3.1 Methods of EHIDLs Evaluation

There are 31 public universities in the country; some with well-established resources, including delivering a postgraduate program. Many are new and under development. Their digital library projects status are different. Using MoE higher education web portal universities library pages were assessed whether they have or under process digital library or repositories tasks were done in their university, in addition to that their willingness to cooperate to this study.

Only for those who have DLs and agree to cooperate (Jimma University, Addis Abeba University, Adama Science & Technology University, Mekele University and Bahir Dar University), to digital librarian or the individual responsible for designing the system a questionnaire that deals with digital library standards they used and types of service they provide sent. The finding also presented in chapter four section one current state of EHIDLs with detail descriptions of the two digital library tools (Greenstone and DSpace) which are used by institutions of Ethiopia; how they works as well as how institutions also used those tools.

Identifying the pioneer stakeholders, developers and end users of the anticipated federated system was the basic one to choose the interoperability protocols for implementing federated system. And it was done by using secondary data collections evaluations with the subordinate interview of selected university librarians (Judgmental sampling technique is also used) about who are the stakeholders of the system, who are their developers (i.e level of education, expertise status, technical and professional skills). Who are their end users (teachers, students, researchers etc)? Their end users difference on information or/and computer literacy, cognitive styles, cultural (background), skill levels and expectations on digital library aspect etc are a valuable metrics for both interoperability protocol selection for coordinated system.

Another important issue in digital library is metadata standards, therefore the overall summary of metadata standards available for the digital library designed by wide initiatives as specific Ethiopian Government Locator Service (ETGLS) proposes also presented. However we propose Ethiopian Higher Institutions Digital Library – Application Profile to handle the metadata concern for the federated system.

3.2 Methods of Interoperability Protocols Designing and Implementing

Existing interoperability protocols are, arguably, overly complex as a result of each protocol being designed by a different group, providing a single service, and having its own syntax and vocabulary (Paihama et al, 2012).

Different interoperability protocols are designed for different purpose. This study not covers all of them instead it focuses on interoperability protocols which are designed for digital libraries or repositories. Still this study goal was to identify the best suite of interoperability protocols for EHIDLs so that to design and implement it, choosing and selecting of interoperability protocols was mainly depends on the below criteria:-

- The protocol which is compatible with each EHIDLs standards they used, services they provide, digital library software they build;
- The protocol that fulfill the identified stakeholders goals and missions;
- Considering the developers level of technical and professional knowledge;
- Easy to use by end users of EHIDLs.

Consider all that currently there are three interoperability approaches deployed, all of them have their own advantages and demerits. But for this study Open Archive Initatives for Metdata Harvesting Protocol (OAI-PMH) metadata harvesting approach deployed with reasons.

In chapter five the proposed system design presented. The reason why the new EHIDL-AP is needed with elements name; controlled vocabulary used; and other guidelines also stated. Because of OAI-PMH was used for federated system data providers and Service providers architectures also described. How the proposed system implements also described in chapter six starting from EHIDL-AP metadata encoding for both Greenstone and DSpace tool users; how both DL tools users implement Data providers (metadata repository) and implement federated system using PKP OHS tool.

CHAPTER FOUR

STATE OF THE ART

For effective interoperability to be accomplished there are so many aspects needed to be considered and evaluate and beside that, there are various initiatives or tasks forces are done so many jobs to enhance interoperability on digital library resources and services provisions. Complex of digital library interoperability leads them to work in different level of interoperability architecture of they want to accomplish the tasks they want to address too. Here in this chapter presents starting from Ethiopian Higher Institutions Digital Libraries current trends and practices in handling resources and service provisions, standards they used; and also in global current motivates of creating digital library standards in metadata aspects and in general any activities to support interoperability activities in specific digital libraries are discussed.

4.1 Current State of EHIDLs

Below table 1 shows universities of Ethiopia who implement digital library the tool or software including the version as well as in which operating system is installed described. And as table shows most of them use Greenstone digital library software and some universities also implement DSpace institutions repository tool to manage the research/project done by their undergraduate and postgraduate students. They all installed on Linux kernel operating system distributions.

Universities	Tool or Software	Operating System
Adama Science & Technology	GSDL v2.85	Opensuse (Linux)
Addis Abeba	DSpace v4.2 GSDL v2.86	Debian (Linux)
Bahir Dar	GSDL v2.85	Ububtu (Linux)
Jimma	GSDL v2.86 DSpace v4.1	Fedora (Linux)
Mekele	GSDL v2.86	Debian (Linux)

Table 1 Universities of Ethiopia Digital library tools they used

So that EHIDL uses DSpace and Greenstone digital library software or tools, how those tools are handle collection building of resources; metadata standards they are using; interoperability protocols/projects proposed or accomplished to enhance the system describe in detail in section 4.1.1. After all implementing DL using those tools almost all institutions only provide A search of a library's collection services apart ASTU is working on to add instructional services in the near future. Table 4.2 elaborate types of services and how many universities are providing the service or not.

Services	Yes	No
A search of a library's collection	5	0
Reference and Question-answering Services	0	5
Filtering and Selective Dissemination of Information	0	5
Instructional Services	1*	4

Table 2 services provided by institutions digital library.

Other issues related digital library standards they used and finally describing the activities or jobs done which are tried to support interoperability for resources sharing among themselves also described in section 4.2. This observation is very central to this study by giving a track how to design the federated system that will enhance the current system they are using.

4.1.1 Digital Libraries softwares and tools

The open source and free software philosophy helps to improve education in developing countries. Among others quality open source digital library softwares availability support academic institutions of Ethiopia's to manage their electronic resources. Based on this study observation from which universities use digital libraries Greenstone and DSpace are the two digital library softwares they used. For that matter how those digital libraries are working more specifically their general ***system architecture***; how they handle ***collection or resources building process,*** methods they used for ***end-user interaction, interoperability protocol*** they support briefly described with how those selected universities used those tools in aspect of metadata schema and other customization they work to enhance the interoperability issues also addressed.

Greenstone Digital library

Greenstone is a suite of software for building and distributing digital library collections produced by the New Zealand Digital Library Project at the University of Waikato, and developed and distributed in cooperation with UNESCO and the Human Info NGO. It is open-source, multilingual software, issued under the terms of the GNU General Public License. Greenstone runs on all versions of Windows, and Unix/Linux, and Mac OS-X. (Greenstone digital library software, 2014)

System Architecture.

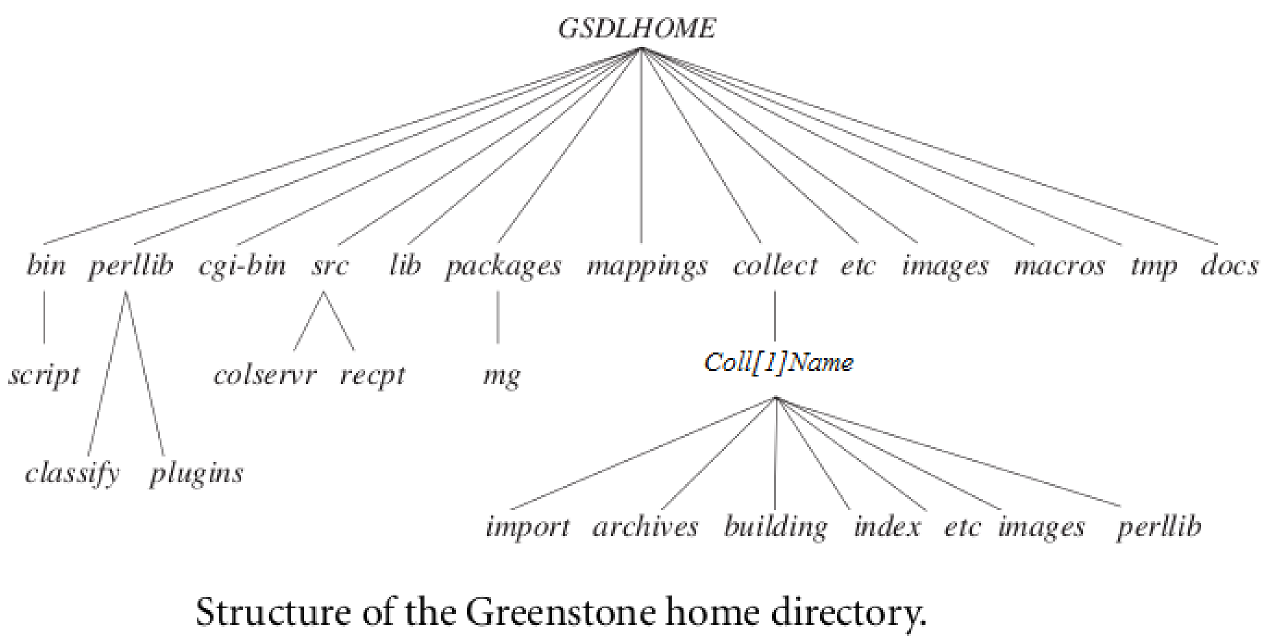


Figure 2 Greenstone digital library software architecture.

The overall structure of the greenstone DL as presented in the above ***GSDLHOME*** is the root of gsdl whole system. Under the root a ***Bin*** folder contains the programs that are used in the building process, and a *script* subdirectory that holds the Perl programs used for creating and building collections.

The ***perllib*** directory contains Perl modules that are used when building. The ***cgi-bin*** directory contains the software that implements the Greenstone runtime system. ***Src*** contains the source code. Common software that is used by both components is placed in ***lib***.

Packages hold the source code for various external software packages. The ***mappings*** directory holds Unicode translation tables. The main Greenstone ***etc*** directory holds configuration files for the entire system. It also includes initialization and error logs and the user authorization database. The main ***images*** directory stores images used in the user interface. The user interface is driven by small code fragments called ***macros***. ***Tmp*** is used for storing temporary files. The ***collect*** directory contains the digital library collections in this Greenstone repository. Each collection has the same structure such as an *import* directory where the original source material is placed and an *archives* directory where the import process's result goes. The *building* directory is used temporarily during building, whereupon its contents are moved manually into *index*. It is *index* that contains the bulk of the information that is served to users. Finally *etc*, *images* and *perllib* sub-directory could be used for collection based configurations. (Witten, Bainbridge & Nicolas, 2009)

Collection and Resources building Process

There are two main parts to the collection-building process in gsdl namely importing and building. The import process brings documents (resources) and metadata into the system in a standardized XML form that is used internally, the Greenstone Archive Format. Afterward the original material can safely be deleted because the collection can be rebuilt from the archive files. The original material is placed in the collection's *import* directory, and the import process transforms it to files in the *archives* directory. Adding new material to the collection could be done by putting it into *import* and re-execute the import process. Then the new material will find its way into *archives,* along with any files that are already there. The build process creates the indexes and data structures needed to make the collection operational. The building process does not work incrementally (although the import process does): indexes for the whole collection are built all at once. (Witten & Boddie, 2004)

For every document a system gives an associated object identifier or OID that is used to identify it within the system. This identifier is *persistent:* that is, it is intended as a permanent name for the document. Assigning object identifiers to documents is one of the import process's major functions. The OID is assigned and stored as an attribute in the document's archive file. If the import process is reexecuted, documents receive the same OID. To ensure this, OIDs are obtained by calculating a random number based on the content of the document called *hashing*. If the content changes, the OID changes, if it does not, the OID remains the same. Identical copies of a document will be assigned the same OID and will hereafter be treated by the system as one. The same document can appear in two different collections: if it does, searching both collections will return just one copy of the document. OIDs are character strings starting with the letters *HASH* (Bainbridge et al, 2004).

Another major task performed during importing process is source documents are brought into the Greenstone system by converting them to a format known as the Greenstone Archive Format. This divides documents into sections and stores metadata at the document or section level.

In GSDL the overall importing and building the collections function can be done either by using command line, greenstone librarian interfaces or remote web interfaces (collector). Choosing from the above methods are depends on the function the collections builder want to accomplish. It's easy to build simple collections by using the Collector leads you step by step through the necessary operations. No programming is required and it is specifically designed to conceal details of what happens behind the scenes. Whereas building using command line is very fast if the librarians are dealing with a collections have large size, however GLI (Greenstone Librarian Interface) is a graphical tool for building new collections, altering or deleting existing collections, and exporting existing collections. It allows importing or assigning metadata also (Witten, Bainbridge & Nicolas, 2009).

Greenstone and Metadata

Digital library resources are enriched by metadata, how metadata schema or standards are using including encoding syntax, controlled vocabulary and others are very important for metadata quality. Metadata schema in gsdl (metadata set) is mds format created by using metadataSet.dtd schema. The default metadata sets for new collections are Dublin Core (dc), the Greenstone Metadata set (gs), and the *Extracted Greenstone Metadata Set* (ex). The Extracted set is unique because it contains metadata automatically generated during the collection building process (and as such cannot be edited) and its metadata fields can be referred to without a namespace (Greenstone Wiki, 2014).

GEMS (Greenstone Editor for Metadata Sets) can be used to modify existing metadata sets or create new ones. Metadata sets in Greenstone are stored in XML files. Any metadata has a namespace, elements, sub elements, elements attribute and might have also element value. And in version 2.86 transforming xsd format of schema into mds gsdl metadata schema format using XSLT is included.

Assigning metadata value for documents in greenstone is achieved by automatic using document based plugins during building process and manually assigning metadata value can be done by using predefined metadata sets.

However in gsdl cause of some metadata schema has an attribute for elements, the value of metadata based on element attribute couldn't be accomplished even though they implement crosswalk dc elements to MODS. Metadata quality of element values (for manual assigning) using controlled vocabulary and encoding syntax are not even giving any direction how could be handled in greenstone. Moreover Metadata multi lingual aspects of language translations only provide metadata element label (name) translation not include element content value rules to be resolved.

How Greenstones Work

The ***receptionist*** accepts user input, typically in the form of keyboard entry and mouse clicks communicated via a Web browser, analyzes it, and dispatches a request to the appropriate collection server (or servers). *Collection servers* interact with the data structures that have been produced by the building process. They locate the requested piece of information and return it to the receptionist for transmission to the user's Web browser and presentation to the user. Collection servers in fact, every time any Web page is requested, the *library* program is started up, responds to the request, and then exits. This is accomplished by the CGI mechanism (Fast-CGI schema) that is widely used by Web servers to communicate with application programs. All Web pages in the user interface are created on the fly: none are stored in advance. They are generated using *macros* written in a simple language specially designed for the job (Bainbridge et al, 2004).

Interoperability Protocols

Greenstone incorporates a server that can serve any collection over the Open Archives Protocol for Metadata Harvesting (OAI-PMH), and Greenstone can harvest documents over OAI- PMH and include them in a collection. And it is also use the Z39.50 and SRU protocols. All of these facilities are provided in the Librarian interface's Download panel. Any collection can be exported to METS (in the Greenstone METS Profile), and Greenstone can ingest documents in METS form. Any collection can be exported to DSpace ready for DSpace's batch import program, and any DSpace collection can be imported into Greenstone (Bainbridge et al, ?).

DSpace Digital Library.

The DSpace is a joint project of the MIT Libraries and HP labs. DSpace is a digital asset management system. It helps create, index and retrieve various forms digital content. DSpace is adaptable to different community needs (Biswas and paul, 2010 Smith, 2002).

System Architecture

The DSpace system is organized into three layers, each of which consists of a number of components. The *storage layer* is responsible for physical storage of metadata and content. The *business logic layer* deals with managing the content of the archive, users of the archive (e-people), authorization and workflow. The *application layer* contains components that communicate with the world outside of the individual DSpace installation. Each layer only invokes the layer below it; the application layer may not use the storage layer directly, for example. Each component in the storage and business logic layers has a defined public API. The unions of the APIs of those components are referred to as the Storage API (in the case of the storage layer) and the DSpace Public API (in the case of the business logic layer) (The DSpace Developer Team, 2013).

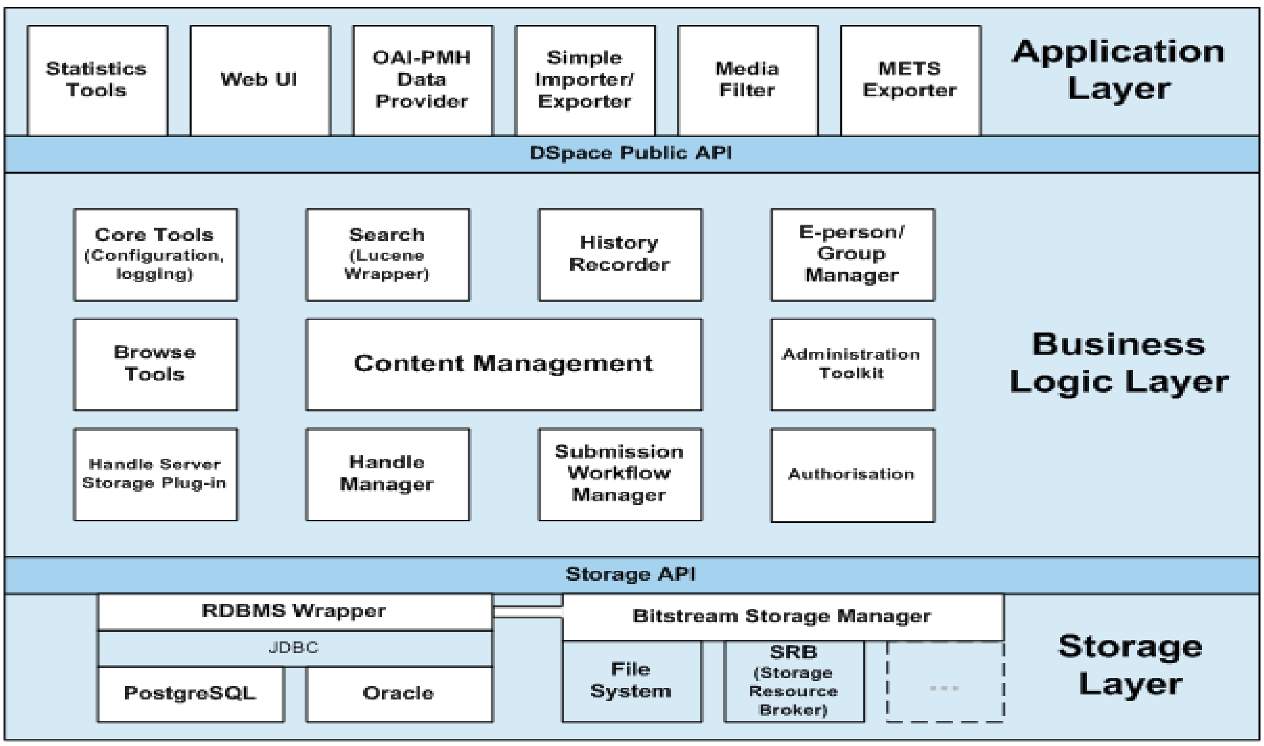


Figure 3 Dspace institutional repository tool system architecture

Collection and Resources Process

The way data is organized in DSpace is intended to reflect the structure of the organization using the DSpace system. Each DSpace site is divided into *communities*, which can be further divided into sub-communities reflecting the typical university structure of college, department, research center, or laboratory. Communities contain *collections*, which are groupings of related content. A collection may appear in more than one community. Each collection is composed of *items*, which are the basic archival elements of the archive. Each item is owned by one collection. Additionally, an item may appear in additional collections; however every item has one and only one owning collection. Items are further subdivided into named bundles of *bitstreams*. Bitstreams are, as the name suggests, streams of bits, usually ordinary computer files. Bitstreams that are somehow closely related, for example HTML files and images that compose a single HTML document, are organized into bundles. Each item has one qualified Dublin Core metadata record. Other metadata might be stored in an item as a serialized bitstream, but for every item for interoperability and ease of discovery. Dublin Core may be entered by end-users as they submit content, or it might be derived from other metadata as part of an ingest process (Tansley et al, 2003).

The basic entity in DSpace is item, which contains both metadata and digital content. Qualified Dublin Core (DC) metadata fields are stored in the item, while other metadata sets and digital content are defined as bitstreams and categorized as bundles of the item. The internal structure of an item is expressed by structural metadata, which define the relationships between the constituent parts of an item. DSpace uses globally unique identifiers for items based on CNRI Handle System. Persistent identifiers are also used for the bitstreams of every item (Smith et al, 2003).

Different DSpace Communities, representing different schools, departments, research labs and centers, have very different ideas of how material should be submitted to DSpace, by whom, and with what restrictions. Individuals from the Community are registered with DSpace, then assigned to appropriate roles

DSpace and metadata

DSpace can support multiple flat metadata schemas for describing an item. A qualified Dublin Core metadata schema loosely based on the Library Application Profile set of elements and qualifiers is provided by default. Most of this is held within DSpace's relational DBMS schema. Structural metadata in DSpace is currently fairly basic; within an item, bitstreams can be arranged into separate bundles as described above. A bundle may also optionally have a primary bitstream. This is currently used by the HTML support to indicate which bitstream in the bundle is the first HTML file to send to a browser. In addition to some basic technical metadata, a bitstream also has a 'sequence ID' that uniquely identifies it within an item (The DSpace Developer Team, 2013).

How DSpace is work

DSpace's current user interface is web-based. There are several interfaces: one for submitters and others involved in the submission process, one for end-users looking for information, and one for system administrators.

The end-user or public interface supports search and retrieval of items by browsing or searching the metadata (all fields for now, and specific fields in the near future). Once an item is located in the system, retrieval is accomplished by clicking a link that causes the archived material to be downloaded to the user's web browser. "Web-native" formats (those which will display directly in a web browser or with a plug-in) can be viewed immediately; others must be saved to the user's local computer and viewed with a separate program that can interpret the file (e.g., a Microsoft Excel spreadsheet, an SAS dataset, or a CAD/CAM file).

DSpace and Interoperability protocol

Even though both DSpace and Greenstone digital library softwares are vulnerable to interoperable with other systems and external resources, none of EHIDL's is using these features. The necessity of interoperability for EHIDLs is obvious figured out, how this will be done considering the system [DL software], the resources they have, service want to provide and other issues are the primary purposes of this project. Therefore this study considering the above criteria designs a system.

Greenstone and DSpace

DSpace is designed for the institutional setting, where members of faculty submit their documents to a common system that enforces common standards. Its model envisages "communities" (e.g., schools, departments, centers, labs, and programs) that contain one or more "collections" of digital "items" Greenstone is designed to allow non-specialist users to produce single, individualized, collections. Its model envisages a "librarian" who is creating collections from existing "resources" (comprising both "items" and metadata resources) and distributing them over the Web or on removable media, possibly in an international setting (Witten et al, 2005).

DSpace supports a hierarchical form of metadata that can be attached at the document level, whereas in Greenstone each item of metadata is flat, but metadata can be attached to individual sections within a document. These differences are reflected in the METS files the two systems generate. To support METS interchange, such differences must be reconciled. Witten et al, 2005 proposed another alternative solutions present StoneD a bridge between the production versions of Greenstone and DSpace that allows users of either system to easily migrate to the other, or continue with a combination of both. This bridge eliminates the risk of finding oneself locked in to an inappropriate choice of system.

4.1.2 Stakeholders, Developers and End Users of EHIDLs

Higher education is provided by universities, university colleges and specialized institutions. They are under the responsibility of the Ministry of Education. Junior colleges and colleges offering diploma program are also under regional governments and private providers. The first stage of university level education leads to the Bachelor's Degree after three to four years' study. University level second stage, Master's Degree; Specialization, leads to a Master's Degree after a minimum of two years' further study. University level third stage, Doctor of Philosophy, is conferred after some three years' study beyond the Master's degree (EMIS, Planning and Resource Mobilization Directorate, 2013).

	Regular	Evening	Summer	Distance	Total
Undergraduate	294,357	61,160	88,030	30,651	**474,198**
Postgraduate (Masters)					**25,103**
Postgraduate (Phd)					**3,165**
				Total	**502466**

Table 3 Enrolments in public universities of Ethiopia by Year-2005 E.C. (2012/13) from MoE Education statistical annual abstract

The above table shows that the number of students enrolled in public universities in undergraduate and postgraduate programs. Based On Ethiopian Higher institution proclamation no 650/2009 definitions "Academic staff" means members of an institution employed in the capacity of teaching and/or research, and any other professional of the institution who shall be recognized so by senate statutes;

	Dipolma	Bachelor	MD/MVD	Master	PHD	Others	Total
Ethiopian	769	6055	1,308	9,622	2,077	220	**20,051**
Expatriate	76	157	263	901	427	19	**1,848**
Total	845	6212	1571	10523	2504	239	**21899**

Table 4 Academic staff in Higher Education by level of qualification MoE Education statistical annual abstract

Sensitive to the fact that expansion of numbers alone would not satisfy the needs of the country, Higher Education Proclamation 351 (Ethiopian Federal Ministry of Education, 2003) made provision for the creation of the Higher Education Relevance and Quality Agency (HERQA) and this was established in 2003 (Higher Education Proclamation no.351/2003) with the aim of safeguarding and enhancing the quality and relevance of higher education in the country. Its mission includes: ensuring that accredited HEIs are of an appropriate standard; establishing that the programs of study offered by these HEIs are of an appropriate quality and relevance to the world of work and the development needs of the country; and supporting the country's higher education sector in enhancing the quality and relevance of its education provision.

HERQA has therefore undertaken several activities to date, including: pre-accreditation and accreditation of a number of programs in private higher education institutions; external quality audits in all the public and some private higher education institutions training of its staff, both locally and abroad, on issues of quality and relevance assurance and enhancement; convening of consultative and training workshops with stakeholders (public and private institution leaders, managers and academic staff and representatives from government organizations and professional associations); development of draft benchmarks for selected subjects; and the publication of procedures for external quality audits and accreditation processes.

And libraries of Ethiopia universities also try their best to implement the government policies and proclamations by providing library and information services to their actual patrons. Based on the interview done by university librarians of ASTU their Digital library most of collection developed by former president of the university after then some staff members donate their resources they get from a broad during their study. They also started scan recommended books to enrich to their DL hence they also considering to allocate budget for electronic resources in the near future. And most of universities use the DL to be accessed only on their IP domain only as security purposes and copy right issues. MoE subscribe INASP ejournal databases so that universities will use.

Some digital librarians of Ethiopians universities take some seminar, workshop and training funded by university budget or other donors held in Ethiopia as well as a broad about how to build digital library. However most of them are using forum as support since they are using opensource DL tools. Because of loosely connection with other universities librarians' cooperation among them are not done.

4.2 Digital Library Standards

4.2.1 Metadata Standards

Digital library technologies are by now well established and understood throughout the higher education. Making effective use of these resources is dependent on the creation of good quality metadata that will help to easily retrieve by end users and to manage these resources by librarians. NISO, 2004 define metadata is structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage an information resource. Or in other words metadata can be **descriptive** describes a resource for purposes of discovery and identification**, structural** indicates how compound objects are put together and **administrative** provides information to help manage and preserve the resources in the repository.

A single organization or corporate organizations can create a metadata standard called metadata schema. A metadata schema provides a formal structure designed to identify the knowledge structure of a given discipline and to link that structure to the information of the discipline through the creation of an information system that will assist the identification, discovery, and use of information within that discipline.

Metadata schemes generally specify names of elements and their semantics. Optionally, they may specify content rules for how content must be formulated (for example, how to identify the main title), representation rules for content (for example, capitalization rules), and allowable content values (for example, terms must be used from a specified controlled vocabulary). There may also be syntax rules for how the elements and their content should be encoded. A metadata scheme with no prescribed syntax rules is called syntax independent. Metadata can be encoded in any definable syntax. Many current metadata schemes use SGML (Standard Generalized Mark-up Language) or XML (Extensible

Mark-up Language).

Dublin Core Metadata Initiatives (DCMI) one of the most popular, developed Dublin core to be simple and concise, and to describe Web-based documents.as well as other types of materials used Qualifiers can be used to refine (narrow the scope of) an element, or to identify the encoding scheme used in representing an element value. However METS developed to fill the need for a standard data structure for describing complex digital library objects. METS provides a document format for encoding the metadata necessary for management of digital library objects within a repository and for exchange between repositories. In other hands, rich description of electronic resources is a particular focus of Metadata Object Description Schema (MODS). Scientific and technical authors, particularly those using mathematical notation, often favor a widely used generalized document-processing system called TeX (pronounced *tech*), or a customized version called LaTeX. This freely available package contains a subsystem called BibTeX that manages bibliographic data and references within documents.

Other initiatives focus metadata standards for specific format (technical types of metadata) of resources such as the Tagged Image File Format (TIFF), EXIF, XMP/IPTC, and MIX for image metadata: embedding metadata in the image file itself or separating the metadata from the image data. Although audio resources are often described using external metadata schemes, such as Dublin Core, several audio formats contain embedded metadata ID3, **AES Core Audio,** Technical Metadata for Textual Objects (textMD), MPEG -7 for multimedia.(Witten et al, 2009)

Digital information is fragile; it can be corrupted or altered, intentionally or unintentionally. It may become unusable as storage media and hardware and software technologies change. Many organizations internationally have worked on defining metadata schemes for digital preservation, including the National Library of Australia, the British Cedars Project (CURL Exemplars in Digital Archives), and a joint Working Group of OCLC and the Research Libraries Group (RLG). PREMIS (PREservation Metadata: Implementation Strategies) also sponsored by OCLC and RLG is developing a set of core elements and strategies for the encoding, storage, and management of preservation metadata within a digital preservation system. Many of these initiatives are based on or compatible with the ISO *Reference Model for an Open Archival Information System* (OAIS).

Because metadata covers too great a variety of information to specify exhaustively and categorically, the Resource Description Framework (RDF) is designed to facilitate the interoperability of metadata, particularly in the realm of the World Wide Web. RDF follows the lead of XML: rather than providing a set of possibilities, it supplies a means for describing a valid system. It is expected that communities of users will assemble to establish RDF systems suited to their collective needs. They will have to agree on a vocabulary, its meaning, and the structures that can be formed from it. This is done by specifying an RDF Schema, just as DTDs and XML Schemas are used to control XML vocabulary and structure (Tauberer, 2008).

And again, metadata is the core of any information retrieval system and so its implications for any digital library are profound: the choice of a metadata scheme underpins any such library's ability to deliver objects in a meaningful way, and greatly affects its long-term ability to maintain and preserve its digital assets. Garnter, 2008 also summarized in order to ensure digital libraries the degree of interoperability long established in traditional libraries the degree of standardization of metadata practices has become more vital.

For large digital library with diversity of resources and intended users, apart the single schema’s easier or complexness over other schemas, no single metadata schema can handle all descriptive, administrative and structural types of metadata to describe and manage digital resources. For that matter Chan & Zeng, 2006 describe various methods we could use to overcome as such problems such as a derivation is creating new metadata schema from the existed one. For instance TEI Lite is derived from the full Text Encoding Initiative (TEI), Both MODS (Metadata Object Description Schema) and MARC Lite are derived from the full MARC 21 standard; a crosswalk is mapping of metadata element, syntax and semantics of one schema to another schema; and application profile is adding, modifying and restricting element, syntax, semantics from multiple metadata schema; and other approaches can help to overcome interoperability issues in metadata level.

Almost all schemas have created crosswalks to popular schemas such as DC, MARC, LOM, or may also include crosswalks to a previous version of a schema as well as to other metadata schemas VRA Core 3.0, which lists mapped elements in target schemas VRA 2.0 (an earlier version), CDWA, and DC. A metadata framework is a reference model that provides a high-level, conceptual structure into which other metadata schemes can be placed. It also gives designers and developers a consistent, cross cutting terminology around which to discuss metadata for a particular purpose.

Each approach has its own role depending on the problem we want to address. For instance, an application profiles describes the set of metadata elements, policies, guidelines and vocabularies defined for a particular domain, implementation, or object type. Based on Heery & Patel, 2000 the basic characteristics of application are it may draw on one or more existing namespaces; introduce no new data elements; and it may specify permitted schemes and values.

MODS/METS application profiles include University of Maryland Descriptive Metadata; UVa DescMeta and Texas Digital Library profile for electronic theses and dissertations Texas Digital Library profile for electronic theses and dissertations.

Dublin core metadata standard initiatives designed DCAP to promote interoperability within the constraints of the Dublin Core model and to encourage harmonization of usage and convergence on “emerging semantics" around its edges (EUROPEAN COMMITTEE FOR STANDARDIZATION, 2003). Among others based on DC-LIB application profile they modify to their local customization such as MWDL; in fact ETD-MS also create application profile for thesis and dissertation resources by adding some element in DC-LIB elements. However MODS is intended to complement other metadata formats and to provide an alternative between a simple metadata format with a minimum of fields and little or no substructure such as Dublin Core and a very detailed format with many data elements having various structural complexities such as MARC 21.

Hodge, 2005 also stated registries are another tool for exchanging metadata. They provide information about the definition, origin, source, and location of the scheme, usage profile, element set, and/or authority files for element values. A registry maps one scheme to another so that both humans and computers can understand how they might integrate, and registries can also document rules for transforming content for an element in one system to the content required for an equivalent element in another.

Ethiopian Information Communication Technology Development Agency develop the ***Ethiopian Government Locator Service (ETGLS)***, has element set of 19 descriptors that resulted from review of International best Practices like that of New Zealand, Australian and the Dublin Core Data and Metadata Standards. ETGLS discuss each elements by element name, label, definitions, comment, obligations and if element has qualifier element the above steps will be repeated.

Based on the questionnaire distributed all of universities DL uses Dublin core (DC) metadata schema for enriching their resources. Jimma University also used METS schema in addition to DC. The reason behind they used Dublin core was simplicity. Because of Dublin core elements are optional the elements they used vary but most of them element title are filled as mandatory including creator, subject (using formal Schema such as DCC for ASTU or informal based on the university school structure) filled as recommended. But none of them use controlled vocabularies for selected elements to assure metadata content quality.

4.2.2 Interoperability Approach.

EU-NSF Digital Library Working Group described Interoperability for digital libraries are more complex than for traditional libraries for several reasons. First, there are myriad technical and engineering issues associated with connecting together networks, databases, and other computer-based systems; Second, digital libraries will provide a greater array of services than do traditional libraries; Third, the types of information available in digital libraries, and the format of this information, will be in much greater variety than is typically the case for traditional libraries. Fourth, and most importantly, digital libraries will be composed of a large number of loosely connected components.

Still the main challenge in providing interoperability among DL collections is the issue of decentralization and heterogeneity among DLs. Decentralization often means a diversity of query languages, information retrieval protocols, capabilities, attributes and organizational structures. Seamless interoperability of DLs involves reconciling heterogeneity and integrating the DLs at several levels (Adam et al. 2000).

To coup up those issues designing digital library interoperability protocols and initiatives may occur at different levels of abstraction In general there are three basic models of interoperability. (1) federated; (2) harvesting, and (3) gathering. In fact, each level has different operating procedures, standards and protocols.

4.2.2.1 Federated model

Federation refers to the case where the digital library sends search criteria to multiple remote repositories and the results are gathered, combined, and presented to user (Shen, 2006). This can be done by using client-server architecture. The server undertakes to update and respond to queries. The client undertakes to connect with end-users, receive queries from end-users, and send, receive and mix received responses from the server, finally presenting them to end-user. In fact, relations between client and server can be established by certain protocols. There is need to standardize query language and data storage following the same standard in all systems. In other situations middleware also used thus the middleware undertakes contact with servers and the user just can make access to resources in other collections along with this middleware and do not need to have any relation with server (Alipour-Hafezi et al, 2010).

Such as NCSTRL, or Networked Computer Science Technical Reference Library, is a confederation of over 100 institutions with the goal of providing a federated search service centered on computer science material. Each organization maintains its own digital library services and the interoperability is achieved by conformance to an open architecture and joint protocol, agreement on data types a metadata format. MARIAN is also a search system for digital libraries. Originally designed for library catalogs, it has been used successfully for collections of varying sizes and structures, and has been enhanced to support digital library and semantic web applications. SDLIP**,** the Simple Digital Library Interoperability Protocol, is a middleware approach to achieve interoperability developed by Stanford University. In SDLIP a wrapper or digital library proxy is defined between the search client and the ultimate information source. Between the client and the proxy SDLIP defines the transport protocol, query language, and other interface so that they can communicate. Clients use SDLIP to request searches to be performed over information sources. The transport protocol can be HTTP or CORBA based. The Alexandria Digital Earth ProtoType (ADEPT) architecture is a framework for building distributed digital libraries of georeferenced information. An ADEPT system comprises one or more autonomous libraries, each of which provides a uniform interface to one or more collections, each of which manages metadata for one or more items. The primary standard on which the architecture is based is the ADEPT bucket framework, which defines uniform client-level metadata query services that are compatible with heterogeneous underlying collections.

A federation is the conventional approach to interoperability is for a group of organizations to agree that their services will be built to certain specifications (which are often selected from formal standards). Organizations that build systems to these specifications form a federation. The problem of forming a federation is the effort required by each organization to implement and keep current with all the agreements. Since the cost of participation is high, federations have small but dedicated memberships.

4.2.2.2 Harvesting Methods

Metadata harvesting was first developed by the harvest project in the early 1990s, but the approach was not widely adopted. The concept was revived in 1998 in a prototype known as the universal preprint server (Van De Sompel et al., 2000). This prototype concluded in favor of metadata harvesting as a strategy to facilitate the creation of federated services across heterogeneous preprint systems. The OAI, which is derived from this experiment, emphasizes the core functionality that can be achieved by digital libraries sharing metadata. It minimizes the cost by using a simple protocol based on HTTP, by providing software that is easily added to web servers, and by documentation, training and support (Lagoze and Van de Sompel, 2001).

In the harvesting model, DLs – which are members of a consortium, agree to interoperate with each other. Hence, they establish a server in order to present services so each library could update their data on the server by means of a simple protocol such as HTTP. Users refer to the server to retrieve information.

Metadata harvesting, unlike meta searching, is not a search protocol; rather, it is a protocol that allows the gathering or collecting of metadata records from various repositories or databases; the harvested records are then "physically" aggregated in a single database, with links from individual records back to their home environments. The current standard protocol being used to harvest metadata is the OAI-PMH (Open Archives Initiative Protocol for Metadata Harvesting) Version 2. The challenge has been to collect these records in such a way that they make sense to users in the union environment while maintaining their integrity and their relationship to their original context, both institutional and intellectual (Woodley, 2009*)*

The most obvious service to provide would be cross-archive searching. The service provider can harvest metadata in one or more formats from multiple remote OAs and index the data according to collection, set, or specific fields within the metadata. Such an experimental search engine has already been developed at Old Dominion University (Liu, 2001) in parallel with the development of the OAI protocol. Other projects such as TORII and OLAC provide cross-archive searching as one of the services offered to their users.

Having regard to an agreed standard in storage and sharing metadata and also using open achieves for making access facility to information by server are the primary necessities of using this model. In this way, server undertakes to present services related to the DL's integrated data. Therefore, the possibility of integrated searching in many of DLs is procured for users.

4.2.2.3 Gathering Methods

It is still possible to achieve interoperability among DLs that are not prepared to cooperate in any formal manner. This can be done by gathering openly accessible information, from search interface to search results. The results gathering approach uses the distributed search approach and it does not require any prior coordination among federated digital libraries. Commercial meta web search engines like MetaCrawler are essentially using the gathering approach to provide a meta search service.

ResearchIndex (formerly known as CiteSeer) is a superb example of a digital library built automatically by gathering public accessible information" (Arms et al., 2002). In fact, this model, because of being more public, was only introduced, and except for the ResearchIndex and Exchange center software that was described in the previous section, there is no other DL project to be discussed here.

Multiple approaches have been used for enabling search across multiple repositories, such as the Z39.50 (Lynch 1997) distributed search method. Distributed searches, however, can be problematic, both for response time and uptime. Search results can only be presented to the user as quickly as the slowest search agent returns (plus some processing time if the results are to be integrated) and the composite uptime is the product of the individual uptimes.

Federation is a more expensive mode of operation in terms of network and search system constraints since each repository has to support a complex search language and fast real-time responses to queries. Harvesting requires only that individual archives be able to transfer metadata to the central DL. The frequency of queries, quantity of metadata, and availability of network resources also factor into this comparison but, in general, federation places a greater burden on the remote sites while harvesting reduces the demand on remote sites and concentrates the processing at the central DL site (Suleman and Fox, 2001).

CHAPTER FIVE

SYSTEM DESIGN

Based on the pervious chapter analysis of current state of the art of EHIDLs trends and practices as well as globe level initiatives on digital library interoperability, here present proposed system for EHIDLs to boost their service by creating a federated system.

5.1 Proposed System Architecture

The general overview of the proposed federated system for Ethiopian Higher Institutions Digital Libraries (EHIDLs) is universities who already built digital libraries will perform enerich their resources metadata accordance of EHIDL-AP metadata standards. Then they need to implement OAI-PMH data provider to send the enriched metadata (md1, md2, md3 etc) for EHIDL service provider. The service provider then gathers those metadata and those metadata will be aggregate semantically then store at EHIDL database And the service provider will implement responding for users request and responds.

5.1.1 Ethiopian Higher Institutions Digital Libraries – Application Profile (EHIDL-AP)

For semantic level of interoperability and to meet the audience of Ethiopian higher institutions digital libraries, this project believes new application profile need to be designed. For that reason this project designs Ethiopian Higher Institutions Digital Libraries Application Profile (EHIDL-AP). Because of most of EHIDL uses unqualified Dublin core elements and most of ETGLS elements also derived from Dublin core elements thus, EHIDL-AP is based on DC-LIB application profile; however there are so many reasons new application profile is needed.

1. Some elements of DC-LIB are very broad to EHIDL's they need refinement. For instance the element dc:contributor needs refinement which must specify what kind of contribution role it has. Is it the contributions is advising the researcher (in case of the creator role is research).

2. Elements of DC-LIB are very constraints to handle all EHIDLs resources types. Because of Dublin core is focusing on descriptive functions with flat format technical and Structural functions are not addressed very well. Therefore elements from other metadata standards also

included.

3. Elements contents are critical issues for high level of interoperability; even DC-LIB comments to use controlled vocabulary for selected elements so that to complete for EHIDLs projects using controlled vocabulary for selected elements must be defined.

4. Element structure, element content syntax, cardinality of the element, and the element obligation of presence are needs modifications to suit local community users and the digital resources we have.

5. After all the primary purposes of application profile is to use, add, modify elements of metadata schema and also to specify the elements content for specific project and target audience. Thus this AP helps as to perform the above tasks based on EHIDL trends and experiences.

Thus, the table shows EHIDL-AP elements with their target namespaces; element refinement; element content value might be entered either manually or automatically. However the value of the element may be selected from listed of controlled vocabulary or based on the encoding, thus for both element and element refinement controlled vocabulary is also described the rest of elements encoding syntax they will used also described. In addition to the table the whole description of EHIDL-AP elements definitions, comments, minimum and maximum availability of an element for single resources (cardinality), obligations and other clarification is described in **Appendix A**.

Element name	**Element Refinement (ER)**	**ER Controlled vocabulary**	**Element content Controlled vocabulary**	**Encoding syntax**
dc:audience			EHIDL:audienceType	
dc:contributor	mods:role	EHIDL:contributorType		xsd:string
dc:creator	mods:role	EHIDL:creatorType		xsd: string
dc:coverage	dct:spatial	LCsH:Geographical		
	dct:temporal	LCsH:temporal		

dc:date	dc:available			xsd:dateTime
	dc:modified			
	dc:created			xsd:date
	dc:submitted			
	dc:accepted			
dc:description	dc:abstract			xsd:string
	dc:tableOfContent			xsd:strngxsd:url
dc:format			dc:MIMI	
dc:language			RFC:4646	
dc:publisher	mods:place		LCsH:Geographical	
	mods:publisher			xsd:string
dc:relations	dc:isPartOf			xsd:URL
	dc:hasFormat			xsd:URL
	dc:isVersionOf			xsd:string
	dc:hasVersion			xsd:URL
	EHIDL:hasTranslations			xsd:URL
dc:right	premis:rightsBasic		EHIDL:rightType	
dc:subject	dc:LCSH		LCSH:topical	
	dc:MeSH		MeSH:SSS	
	dc:keywords			xsd:string
dc:titlte	dc:alternative			xsd:string
dc:type			dc:dcmiType	
etd:degree	etd:degree.name			xsd:string
	etd:degree.level	EHIDL:degreeType		
	etd:degree.dicipline	EHIDL:deciplineType		

	etd:degree.granter			xsd:string
	EHIDL:thesisType	EHIDL:degreeType		
dc:identifier	Doi			
	ISBN			
	ISSN			

5.1.2 OAI-PMH data provider

Most DLs are driven by databases; thus the popular search engines do not index their contents. As a result, search engines are not of much use to users who want to perform searches across multiple DLs. In order to address this need, different approaches such as Z39.50 (ANSI/NISO, 1995), Dienst protocol from Cornell University, STARTS protocol from Stanford University were taken by various communities of users (Suleman and Fox, 2001).

Because of federated search engines deficiencies; common agreement of their resource policies and more over the simplicity and advantageous of using metadata harvesting protocol is very pioneer for high quality of interoperability in digital libraries. This choice is based on the fact that the cross archive search approach does not provide the necessary scalability for large numbers of participating repositories. The lack of scalability in the cross archive search approach arises from the fact that the service performance always depend on the slowest data provider (Alipour-Hafezi et al, 2010). Now days OAI-PMH is getting popularity because of it is simplicity and both digital libraries tools (DSpace and Greenstone) already have pre-built OAI-PMH data provider, we use them for metadata harvesting in this project.

In our situations, each university who owns digital resources must implement OAI-PMH data provider. A data provider maintains a repository that allows external online access to its metadata through the OAI Protocol for Metadata Harvesting (Suleman, 2002). Accordingly universities enrich their resources metadata that described by using EHIDL-AP to be harvested by EHIDL service provider. And also repositories (data providers) that have resources for each resource (item) likewise must have unique address and for each resource metadata in dc and EHIDL-AP (OAI-PMH needs at least dc metadata and this project requires EHIDL-AP) should be assigned or mapped based on EHIDL-AP metadata reference.

The basic (minimal specification) for digital libraries to implement data provider of OAI-PMH described on repositories implementation guidelines. Moreover for this project we also mention some of the basic specification we want to consist of. Data providers' bear in mind that a resource is a representation of a single digital document which available in any format, nature of the resources is not defined in the OAI-PMH they may be digital or non-digital, However for this project nature of the resource must be only digital. An item is a constituent of a repository from which metadata about a resource can be disseminated. An item is conceptually a container that stores or dynamically generates metadata about a single resource in multiple metadata formats, each of which can be harvested as records via the OAI-PMH. Each item has an identifier that is unique within the scope of the repository of which it is a constituent. A ***record*** is metadata in a specific metadata format. A record is returned as an XML-encoded byte stream in response to a protocol request to disseminate a specific metadata format from a constituent item.

This resource must be described with the way EHIDL-AP rules, plus to that because of EHIDL-AP uses much of dc-lib crosswalk/mapping from EHIDL-AP to dc will be take place but never vice versa.

Generally OAI-PMH already create guidelines for repository implementers, we use this guidelines. However the OAI-PMH considers data providers share metadata or even the resources to any harvester who want to harvest from any data provider. Apart Addis Abeba University ETD all of digital library only accesed by their IP domain, Therefore each data provider configure to secure their digital resources to accessed only by EHIDLs service provider.

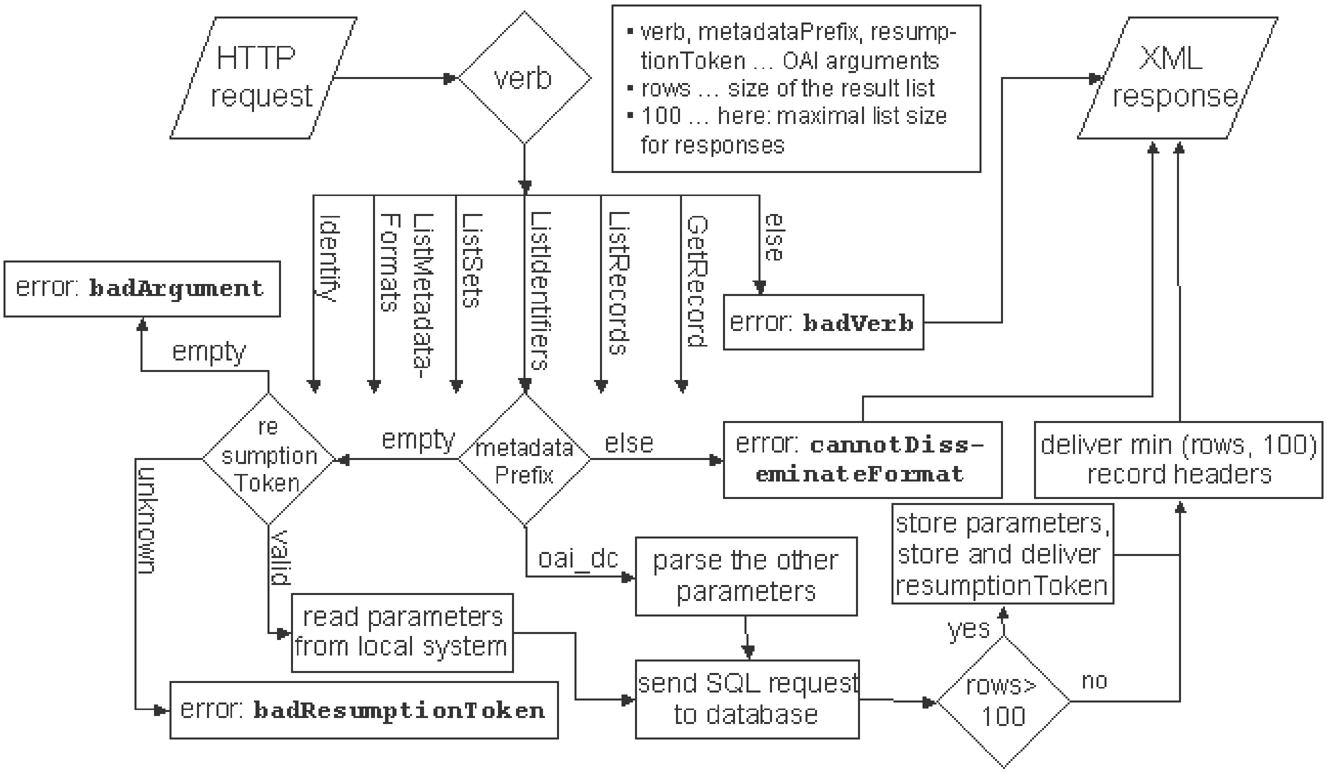


Figure 4 OAI-PMH data provider verb request type with answers captured by The Open Archives Forum

As figure shows when receiving an OAI request the Data Provider has to parse the query and firstly has to decide which of the six valid request types is issued or if the request type is illegal. The latter case (verb parameter has a nonstandard value) results in an error message to the Service Provider (**badVerb**). In case the issued request type is **ListIdentifers** the next parameter the parser has to check is **metadataPrefix** because this argument is mandatory for the request type **ListIdentifiers**. If the parameter has not been provided the only possibility for the request to be valid is to have a **resumptionToken** parameter which has to be known to the Data Provider. In this case the Data Provider reads the locally stored parameters representing the arguments of the original request and cursor information indicating how many identifiers have already been delivered to the Service Provider. If the **resumptionToken** argument is emtpy as well or has an unkown value error messages have to be generated.

The only valid value for the **metadataPrefix** parameter is **oai_dc** because the example Data Provider assumed here can only deliver metadata sets in the unqualified Dublin Core schema. If this is the case the other optional parameters have to be parsed, which in the figure has been described informally for reasons of simplicity. The possible parameters are **from**, **until** and **set**. In this process, error messages have to be issued if the parameters have illegal values or if the query contains other parameters not allowed for this request type.

Subsequently, the given parameters received by the query or - in case of a resumed resumptionToken query – read from the local system have to be assembled to an SQL query which then has to be issued to the database. If this results in more than 100 records (100 in the example is the maximum number of delivered indentifiers at once) the Data Provider has to generate a new resumptionToken and to store it locally together with the query parameters and the cursor information. The resumptionToken has to be included in the XML response to the service provider as well. Of course, the XML response also contains the identifiers returned by the database.

5.1.3 OAI-PMH Service Provider

EHIDL service provider based on PKP Open Harvesting System (OHS) is implemented to gather, harvest, normalize, store the metadata from those universities who implement data provider. Based on the retrieved metadata it will respond users query and also other service will provide. Besides the architecture of the OAI-PMH services provider is profound to describe how EHIDL OHS is work.

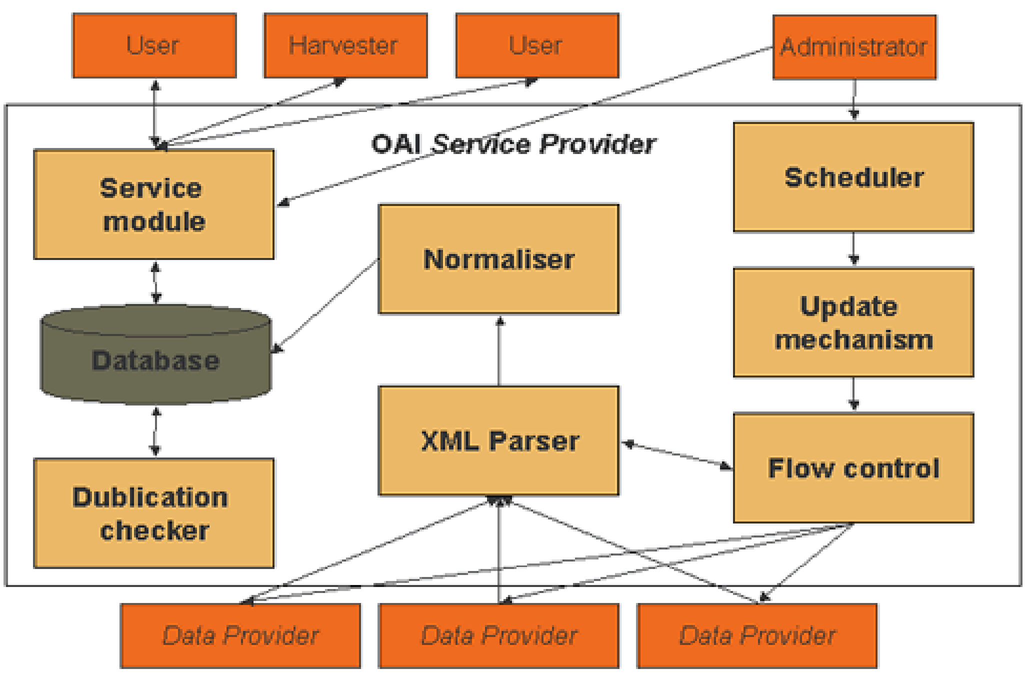


Figure 5 OAI-PMH service Provider architecture overview captured from OAI forums

Service provider first of all implements **Archive management** that involves the selection of repositories to be harvested. All institutions (Data provider) will be added, managed using it.

Request Component creates HTTP requests and sends them to OAI repositories (Data Provider). It demands metadata using the allowed verbs of the OAI-PMH. It may do selective harvesting using the **set** parameter. We used *ListIdentifer* methods with *MetdataFormat=EHIDL-AP* during data provider added for the first time.

Scheduler realises timed and regular retrieval of the associated archives. The simplest case would be manual initiation of the jobs, but this is done by automated using of a cron job in our study.

Flow Control is implemented via resumption token, partitioning of the result list into incomplete sections with a new request to retrieve more results. An HTTP error 503 (service not available) allows analysis of the response to extract a "retry-after" period. Unfortunately OHS not support resumption token thus we use command line to harvest all metadata at one time.

Update Mechanism realises the consolidation of metadata which have been harvested earlier (merge old and new data). The easiest case would be to delete all 'old' metadata from each repository before harvesting it again. A reasonable alternative is to do an incremental update (**from** parameter) – insert *new* metadata and overwrite *changed/ deleted* metadata (assignment using the unique identifiers).

XML Parser analyses the responses received from the repositories, with validation using the XML schema, and transforms the metadata encoded in XML into the internal data structure.

Normaliser transforms data in different metadata formats into a homogenous structure. It harmonises representation of, for example, date, author, language code. It may map between or translate different languages.

Database receives the output of the normaliser mapping the XML structure of the metadata into a relational database that will handle multiple values of elements. An alternative is to use an XML database.

Duplication Checker merges identical records from different data providers. One possibility for implementing this is by the unique identifier for each item (for example, by URN). However, this solution is often not easily practicable and is not risk or error free.

Service Module provides the actual service to the 'public'. The basis for a service provided is the harvested and stored records of the associated archives. That is, it uses only the local database for requests etc., and thus it does not make calls on the Data Providers during operation.

To point out those participating universities is only dealing with data providers implementations the service provider implementation is part of this study jobs

Even though different Service providers tools are available, apart PKP Open Archive Harvester and the Virginia Tech Perl Harvester. Other Harvesters, most of performed poorly with respect to installation ease, are also explored (Kellogg, 2004). Therefore this study implement service provider by using PKP Open Harvesting System by modifing and customizing to handle the above requirements. The PKP Metadata Harvester allows for the creation of a customizable, searchable, online index of metadata available from Open Archives Initiative-compliant databases and information sources. The software harvests OAI metadata in a variety of schemas, including unqualified Dublin Core, MODS, ETD-MS and MARCXML. The Harvester has a flexible search interface that allows for both simple and advanced searching using crosswalked fields from all harvested archives. It also has the ability to perform post-harvest and pre-indexing filtering/normalization on the metadata.

CHAPTER SIX

SYSTEM EXPERIMENTS AND TESTING

The proposed system that was designed in chapter four are involved different tasks to be processed to get an interoperability system. More over this chapter presents starting from creating EHIDL-AP metadata using DSpace and Greenstone institutional and digital library tools; implementing OAI-PMH data provider for both DL; Implement OAI-PMH service provider by using OHS OAI-PMH service provider to accepts user request and responds with appropriate results.

6.1 Metadata Encoding

Ethiopian Higher Institution Digital Library – Application Profile (EHIDL-AP) is created to enrich resources of Ethiopian Higher institution Digital Libraries. Even though most of Ethiopian Higher Institutions uses Dublin core metadata schema without any rules or guidelines how they are using the metadata schema. This project pointed out designing an Application profile which at least to handle descriptive type of metadata that all universities should follow that could help them to enable standard along with other universities too. Despite we present EHIDL-AP elements, contents, and others guidelines at Appendix A section; Implementing those rules with the current digital repository tools those universities used such as on DSpace and Greenstone are clearly discussed here.

6.1.1 DSpace EHIDL-AP

Because of EHIDL-AP primary purpose is to describe digital resources so that sharing metadata resources from one university to another will be done with high level of semantically. As we already described other university uses DSpace for thesis management, beside metadata DSpace uses a qualified Dublin Core metadata standard for describing items intellectually (specifically, the Libraries Working Group Application Profile). Only three fields are required by the system: title, language, and submission date, all other fields are optional (The DSpace Developer Team, 2013). There are additional fields for document abstracts, keywords, technical metadata and rights metadata, among others. However this study believe that those three dc elements are not enough to be mandatory as well as all activities of metadata creation need to consider issues of quality, data checking, error correction and the ongoing refinement of processes for error prevention so that will maximize agreement between those systems including the librarians on managing their resources across the nations (Witten, 2009). Well-structured digital library management will help the user to effectively use it the system. For that matter in this section presented creating EHIDL-AP with the rules and recommendation that proposed in chapter four and appendix B of this study for those who used DSpace institutional repository management system.

Metadata Schema Register

DSpace is installed and configured to use the Dublin Core metadata schema by default. In fact most of EHIDL-AP elements are from dc, still there are also elements that are added from various metadata schema standards with bearing in mind that even some dc elements refinements (qualifying) and using controlled vocabularies for selected elements are need to be done also. Metadata schema creation (metadata Register) in DSpace can be achieved by using DSpace web interface or by importing well-structured xml metadata schema in [DSpace]/config/register path. Creating metadata registers are recommended to use the web UI for post installation of DSpace. Thus we also use Web UI to create EHIDL-AP, since DSpace metadata registry are using xml syntax any metadata schema must have ***namespace(*** *http://moe.edu/et*) and ***Name (Prefix*** given by designer of the metadata standard schema) *EHIDL-AP* are assign to EHIDL-AP.

For each metadata element apart dc schema elements by using web UI of DSpace fields ***Element,*** injecting the metadata element name given by the source metadata schema followed by the prefix of schema and colons are done. For those Elements that have refinements using ***Qualifier*** fields they are inserting to the schema. To guide the metadata librarians general descriptions of the elements are placed on ***Scope note*** of field. Therefore the ***etd:degree*** element with *name, level, discipline, granter* and *ehidl:thesisType* qualifier are created.

For elements those that are derived from dc moving metadata elements to EHIDL-AP is applicable. But we did not use it because moving a metadata field will also remove it from its existing metadata schema, in fact any items which previously used this metadata field will be updated automatically (i.e. no existing metadata should be lost). However some elements of dc are qualified by from other metadata schema for our application profile and updating elements on dc Metadata schema registry is done. So for *dc:contributor* ***Qualifier*** *advisor, editor, illustrator;* dc*:publisher* ***Qualifier*** *place, press; and dc:relations* ***Qualifier*** *EHIDL:hasTranslations* only adding those qualifier are done.

Still producing metadata schema, adding elements or sub elements will not automatically available the elements to be used for describe (enriching resource with metadata) during submission of items to the repository, before that we need to construct controlled vocabularies that will be used for metadata elements.

Controlled Vocabularies

DSpace supports controlled vocabularies so as to confine the set of keywords that users can use while describing, searching or browsing items. Supported controlled vocabularies are expressed in a simple XML format (DSpace node schema). According to this schema all information about a term is enclosed in a *<node>* element. Only the expression of a hierarchical (narrower in meaning) relationship is allowed through the use of the *<isComposedBy>* sub-element. Furthermore, by using *<hasNote>* a simple annotation mechanism becomes possible. [sec5ds]. EHIDL-AP use controlled vocabularies derived from other standard initiatives developed. Besides using those controlled vocabularies local controlled vocabularies (created by this study) also developed according of DSpace metadata xml schema (controlledVocaublary.xsd) for validation of controlled vocabulary. And put it into *[DSpace]*/config/controlled-vocabularies/. By default sssr.xml (for subject keyword) and nsri.xml (not used) are available. Hence all controlled vocabularies audienceType, dcmitype, degreetype, deciplineType, geographical, mesh, temporal and topical that are identified in EHIDL-AP are created with the same as below described audienceType controlled vocabularies here presented. The root node label describe the controlled vocabulary terms by excluding the prefix of controlled vocabularies derived from. For now controlled vocabularies are created for metadata content value of English (default) language for other language crosswalk is already implement by service provider.

```
- <node id="audience" label="Audience Type">
  - <isComposedBy>
      <node id="eng1" label="Student"/>
      <node id="eng2" label="Teacher"/>
      <node id="eng3" label="Researcher"/>
      <node id="eng4" label="Librarians"/>
      <node id="eng5" label="Community"/>
      <node id="eng6" label="All"/>
    </isComposedBy>
 </node>
```

Figure 6 Controlled vocabulary of audienceType.xml which used for dc:audience element content

Item Submissions

Items (resources) building in DSpace can be done by bulk using the batch item importer an application that turns an external Submission Information Package (SIP) [4] (an XML metadata document with some content files) into an "in progress submission" object. And by using the Web submission UI is similarly used by an end-user to assemble an "in progress submission" object can also build the item one at time.

The *[DSpace]/config/item-submission.xml* contains the submission configurations for WU interface item submissions. The default that we used the "traditional" Item Submission Process for DSpace, which consists of the following Steps (in this order):

Select Collection (choosing the collections we want to submit items) -> *Initial Questions * (that will help to hide an used elements for descriptions)* -> *Describe (this is the metadata enrichments page that will be discuss in depth in here)* -> *Upload (browse and upload the item)* -> *Verify (check the over all job we done before publishing the items)* -> *License* -> *Complete*

Rearranging, deleting and generally modifying those steps needs to configure the item-submission.xml file with actual java class that need to be created. For this study configuring the describe page is enough.

Custom Metadata-entry Pages for Submission

Automatic metadata contents that are recommended to assign the element value automatically by the system rather than manually during the items (resource) building process will not editable like manual metadata elements content. Additionally DSpace user-interface engine skips Dublin Core fields which are not needed, according to the initial description of the item. For example, if the user indicates there are no alternate titles on the first "Initial Questions" page (the one with a few checkboxes), the input for the *title.alternative* DC element is automatically omitted, *even on custom submission pages.*

The metadata web forms are controlled by the Describe step within the Submission Process. However, they are also configurable via their own XML configuration file (input-forms.xml).

The description of a set of pages through which submitters enter their metadata is called a form. A form is identified by a unique symbolic name. The content of the form is a sequence of page elements. Each of these corresponds to a Web page of forms for entering metadata elements, presented in sequence between the initial "Describe" page and the final "Verify" page (which presents a summary of all the metadata collected)

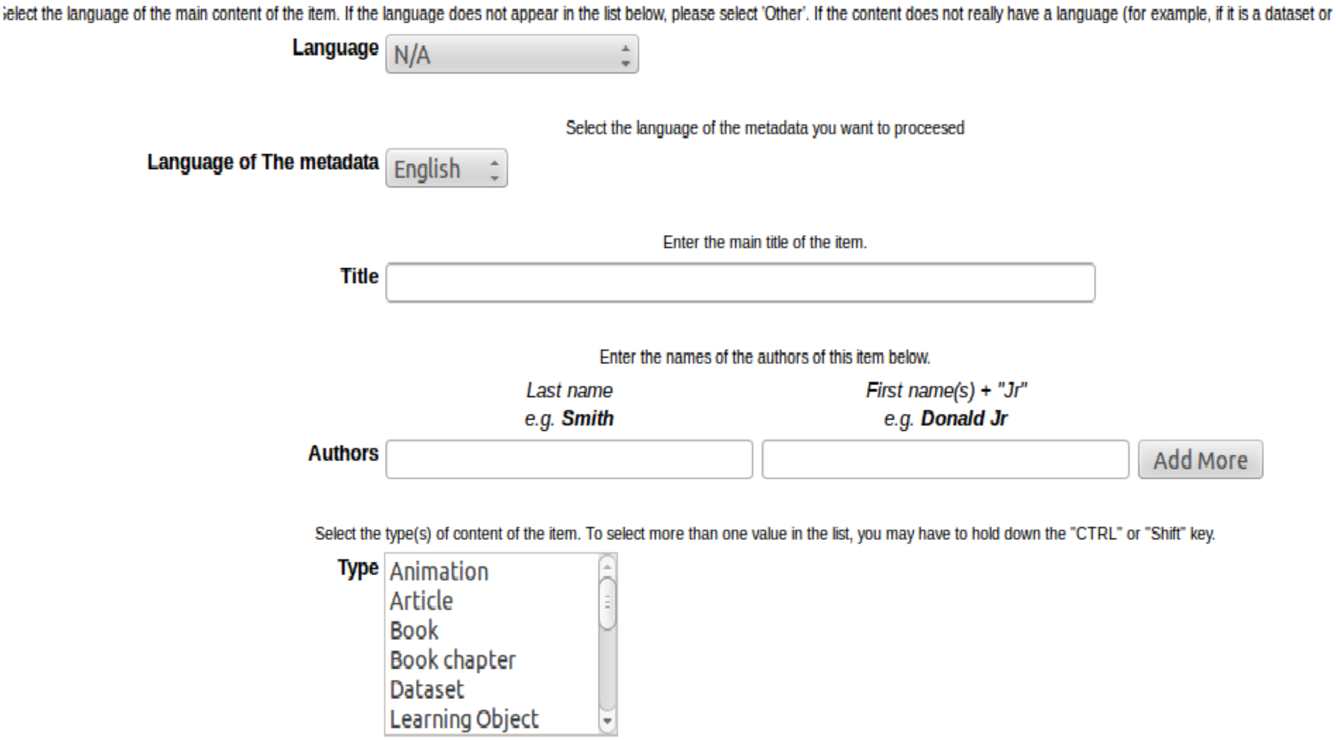


Figure 7 Dspace Describe page

Fig this is how DSpace users enrich their content in metadata starting from the content language with available listed of language and for title the user insert text, some resources might have more than two types to select and obviously single resources might be worked by more than two authors that is why add more button is used for authors element.

This study identify metadata content value enrichment structure in the way identifying element such as date and format by omitting from input.xml the value must be entered automatically. Still for the remaining element based on the necessity and some elements contents are also depending on type of resource as such as in page 1 start from language of content and metadata; creator; title; publisher; identifier; subject; type and audience are put in respectively. Then in page two description; coverage; right; source are also put however if the type is article the advisor qualifier of contributor; degrees qualifiers and dateofaccepted and date of submission will be set too.

For each elements that are identified input.xml presented below page number under field described *by* the prefix of the schema of the elements ; then *element name with qualifier if available* must be similar with metadata registry already presented. For those multiple values are acceptable *repeatable* is set true, label and hint are used for description of users. *Input-type* are configured (onebox, twobox, textarea, name, date, series, dropdown, qualdrop_value). For those who use controlled vocabularies the element *vocabulary* with the actual controlled vocabulary name (without the xml file extension) must be descibed. However we use qualdrop_value input type to for element type. If the user select *thesis* from *dc:type* elements such as *dc:contributor.Advisor; dc:date.DateofAccepted; dc:date.dateofsubmitted* and *etd:degree* with qualifiers will be available in page 2 of describe.

Generally by performing the above tasks metadata librarians easily submit the items with EHIDL-AP that helps not only by making schema to the librarians it will help them to fill the metadata value in uniform way by using controlled vocabularies to identified metadata elements.

6.1.2 Greenstone Metadata

Greenstone can ingest many other formats, such as metadata records downloaded over the Open Archives Initiative Protocol for Metadata Harvesting, MARC that used for bibliographical description of library materials, ISIS, which is widely used in developing countries; BibTeX, which pervades mathematics and computer science; and the commercial product Procite. It can also ingest metadata expressed in a spreadsheet, where the first row contains the names of the metadata fields and subsequent rows contain metadata records. (iwaten et al, 2009). Thus in those sub sections creating EHIDL-AP metadata schema (metadata set in gsdl) using greenstone GEMS tools; creating controlled vocabularies as well as mapping those controlled vocabulary to the identify metadata elements and enriching the documents with those EHIDL-AP set described.

Metadata sets (Greenstone Editor for Metadata Sets (GEMS)

All metadata fields in Greenstone belong to a *metadata set*, which is simply a pre-defined collection of metadata fields. Because sets will often have metadata fields with the same name (for instance, most sets will have a 'Title' field), *namespaces* are used to distinguish between metadata from different sets. For instance, all metadata fields in Dublin Core are preceded by dc. (dc.Title, dc.Creator, etc.).

Greenstone come up with GEMS (Greenstone Editor for Metadata Sets) for creating or modifying metadata set for new ones and for existing one respectively. GEMS can be launched from the GLI (in the Enrich panel ***Manage Metadata Sets… → Add… → New…***) or can be launched directly from ***Start → All Programs → Greenstone → GEMS.***

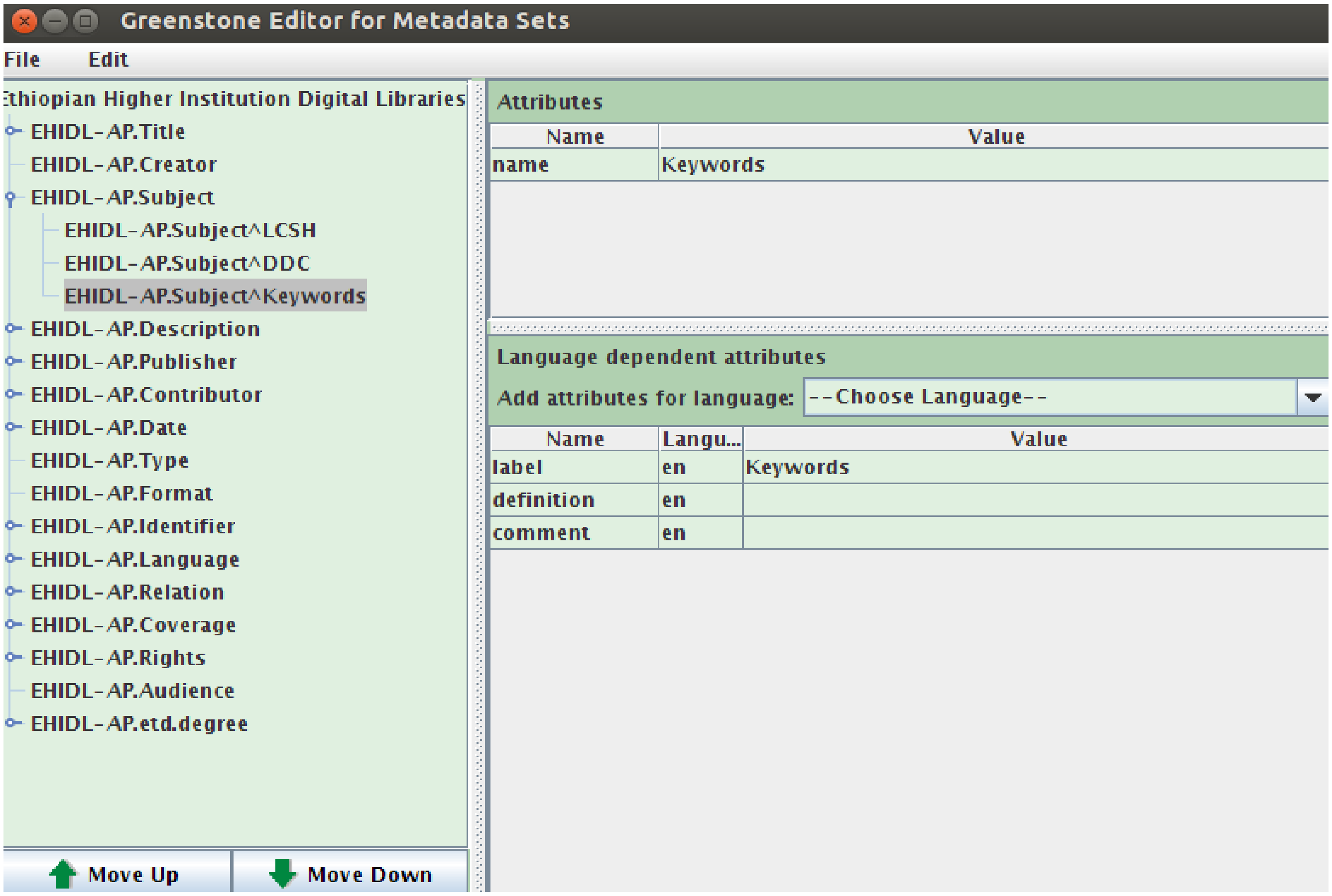


Figure 8 EHIDL-AP metadata set creation using GEMS

The above figure shows EHIDL-AP metadata set created using GEMS based on Qualifed Dublin core set (GEMS have option to create metadata set based on the created set or new framework) In fact deleting the unused dc elements and also creating sub elements such as subject^LCSH, subject^DDC are done to accommodate the rule of EHIDL-AP described in the appendix. To add a new element, right click on the name of the set and choose "*Add Element*". To add a new sub element, right click on the element and choose "*Add Subelement*". Elements and sub elements can be deleted by choosing "*Delete Subelement*" from the right click menu. Sub elements are denoted like this: *element^subelement*. The *Move Up* and *Move Down* buttons helps to re-arrange the sequence of elements display in *Enrich Panel* interface.

Another Important aspect is the created metadata set is created based on MetadataSet.dtd schema and saves as EHIDL-AP.mds. And sample mds code is shown in figure 6.8. MetadataSet attribute contact, creator, family are used only for description however the attribute of namespace is used for internal purposes. Each metadata elements are created with sub-elements of MetadataSet elements. For instance Element Name “dc:audience” uses controlled vocabularies that are specified *OptionList* sub-elements with the actual values of *Value.* Enerich Panel interfaces will shows if the Elements of dc:audience attribute *predefined* value is True and also if the users must only use those controlled vocabulary attribute of *restricted* value should be true. For Subject Elements we also create LCSH, DDC sub-elements that those also uses controlled vocabularies creations will be similar with the way the above created by considering correctly nesting the elements with xml dom syntax fragment.

```
<MetadataSet contact="zghidey@gmail.com" creator="ziena Ghidey" family="application profile" namespace="EHIDL-AP">
-<Name language="en">
   Ethiopian Higher Institutions Digital Library Application Profile
 </Name>
-<Element name="dc:audience" restricted="true" predefined="true" accumulate="unbounded">
   <Attribute name="label" language="en"> dc:audience </Attribute>
  -<Attribute name="definition" language="en">
     An entity for whom the resource is intended or usefu
   </Attribute>
  -<OptionList>
     <Value> Undergraduate </Value>
     <Value> Postgraduate </Value>
     <Value> Researchers </Value>
   </OptionList>
 </Element>
-<Element name="dc:subject" accumulate="unbounded">
   <Attribute name="label" language="en"> dc:subject </Attribute>
   <Attribute name="definition" language="en">The topic of the content of the resource</Attribute>
  -<Element name="dc:LCSH" restricted="true" predefined="true" accumulate="unbounded">
     <Attribute name="label" language="en"> dc:LCSH </Attribute>
     <Attribute name="definition" language="en">Library of Congress Subject Heading </Attribute>
    +<OptionList></OptionList>
   </Element>
 </Element>
</MetadataSet>
```

Figure 9 EHIDL-AP.mds created by GEMS

The Enrich Panel

After all we create a metadata set with the appropriate metadata elements, the next phase will be assigning metadata content for the documents. Greenstone supports command line resources building and Graphical user interface building as already described. To build collection using Graphical interfaces first we should select collection server we want to build then importing the documents (folder) we want to build. Then selecting single document or folder metadata content assigning could be done using Enrich Panel tab (Interface). By using Format and Design tab formating structure and how those collections will be displayed on the web user interface configurations are listed. Finally in the create tab there is Build and Minimal build tab are used for the new and to rebuild respectively purposes. We will Focus on Enrich Panel (because for this project the metadata aspects of Dls).

The general overview of *Enrich Panel* interfaces looks like the below figures. The number 1 with the circle background area is the imported documents or folders that will be publish after building process, below that there is *Manage Metadata Set* Buttons to add, modify or remove Metadata set that will be used. The Extracted Metadata set is the only that could not be removed, by default unqualified dc metadata set is selected; However we remove the default with EHIDL-AP by using the *Manage Metadata Set* button. Therefore each metadata elements from selected metadata set are displayed with a table first column metadata element label (name) based on the language GLI is chosen, for each metadata element a blank column is created to insert metadata value. To assign metadata value first the folder or document must be selected then click on the metadata value column and add the value, if the element have multiple values Enter after assigning the value will add that specific metadata element (circle background with number 2).

Finally there is metadata value tree panel (number 3) that shows the whole metadata value assigning to that specific element or for controlled vocabulary (OptionList) it will list the values that could be used. Click to those values will automatically assign the values to the selected elements. If the elements are restricted to controlled vocabulary assigning any other value will not be acceptable.

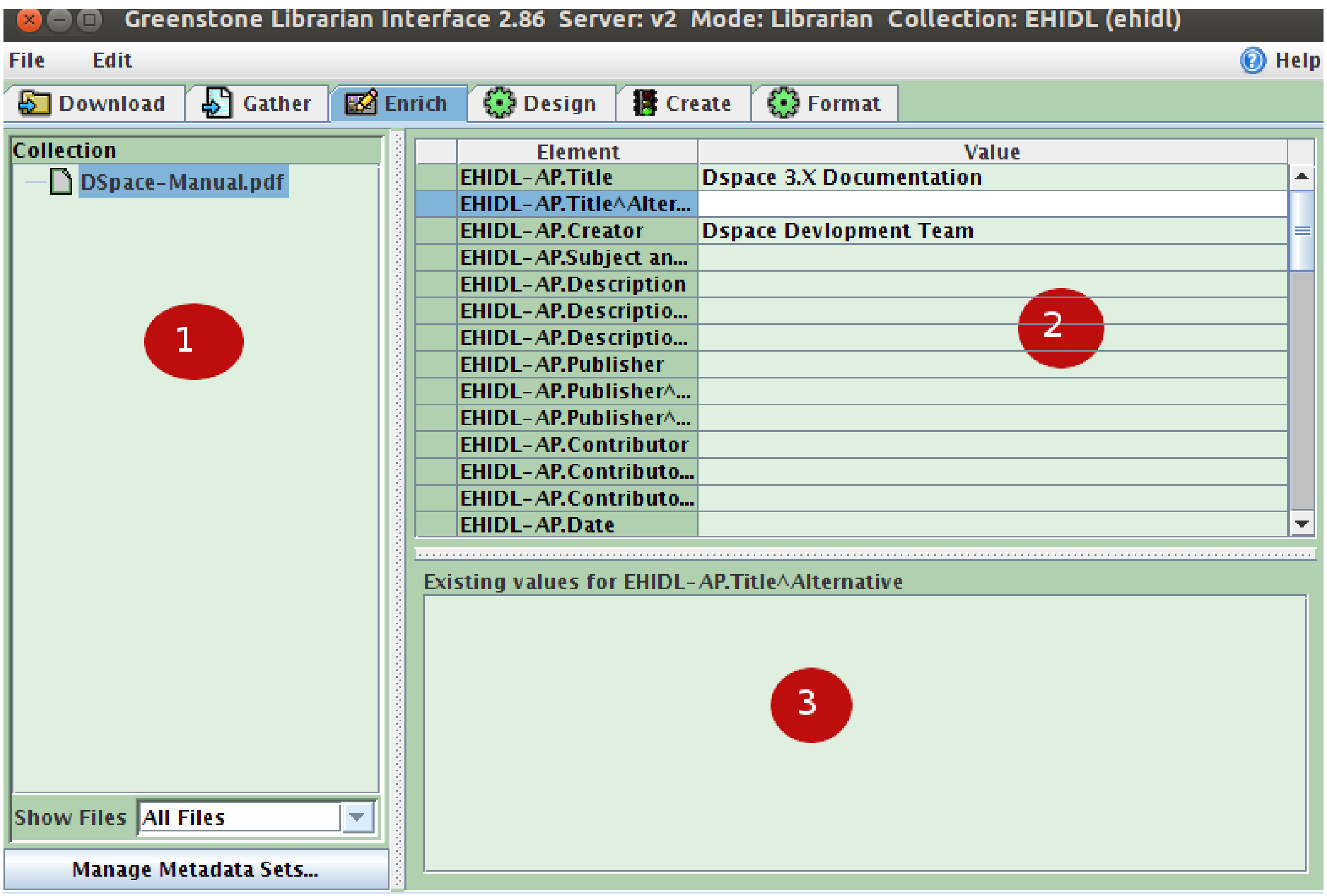


Figure 10 GSDL Enerich Panel

Behind the scenes, the Librarian interface stores any metadata manually assigned to documents in automatically created XML files called *metadata.xml*. In fact, if you create your own metadata files in this format and simply place them in the folder along with your documents, the Librarian interface will automatically pick up the metadata when you drag the documents in.

6.2 OAI-PMH Data Provider

Those metadata created accordance of EHIDL-AP by using DSpace and GSDL metadata editing tools to be maximizing items describing and discovering for cross archive level those institutions needs to implement OAI-PMH data provider. The OAI-PMH defines various roles in an architecture built on metadata. A *data provider* makes its metadata available for use in one or more description formats. From the common finding aids, records are made available so as to match a formal XML schema (Foulonneau and Dawson, 2003)

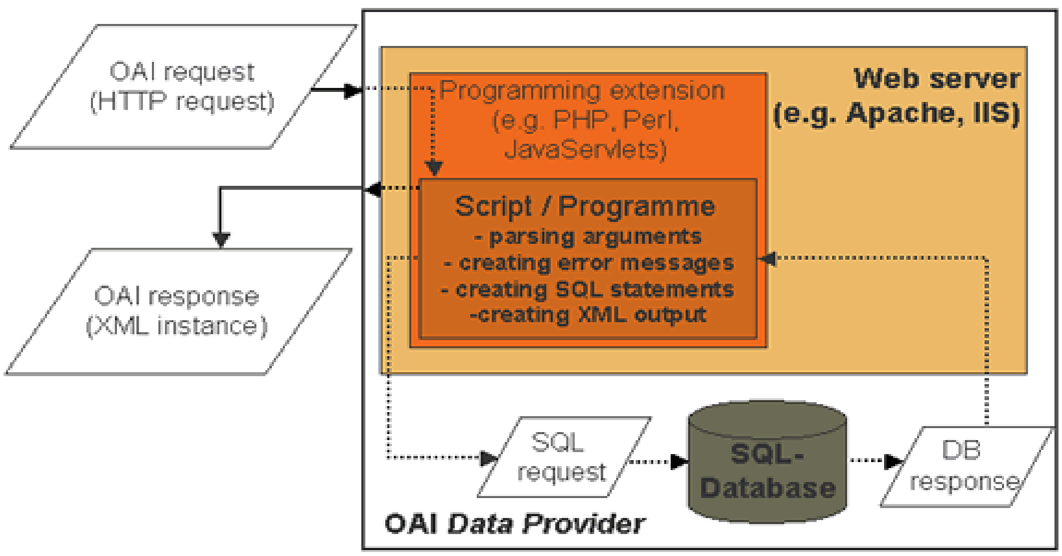


Figure 11 OAI-PMH Data Provider Architecture

Figure 11 illustrate OAI-PMH data provider implementation components: the SQL Databases based on the metadata schema they want to deliver store metadata records. The web server implements OAI response (xml format) for OAI seven verb http request with valid response as well as error messages for different kinds of error verb requests.

Both DSpace & Greenstone digital library tools have pre-built OAI-PMH data provider implementation. General configurations & implementation with this study mandatory EHIDL-AP metadata employment in both systems is discussed in below sections.

6.2.1 <u>DSpace OAI-PMH Data Provider</u>

The DSpace platform supports the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) version 2.0 as a data provider. This is accomplished using the XOAI OAI-PMH Java Toolkit. DSpace exposes the Dublin Core metadata for items that are publicly (anonymously) accessible. Additionally, the collection structure is also exposed via the OAI protocol's 'sets' mechanism. OCLC's open source OAICat framework is used to provide this functionality. DSpace's OAI service supports the exposing of deletion information for withdrawn items, but not for items that are expunged. DSpace also supports OAI-PMH resumption tokens. The OAI service can also be configured to use of any crosswalk plugin to offer additional metadata format.

OAI 2.0 of DSpace has a configurable data source, by default it will not query the DSpace SQL database which will decreases performance significantly at the time of the OAI-PMH request. Instead, it keeps the required metadata in its Solr index. Therefore by using command line to import the metadata and/or clean the cache from the DSpace system to OAI 2.0 of DSpace ***([DSpace]/bin/DSpace oai <action> [parameters])***, However we recommend to scheduler the import activities by using crontab.

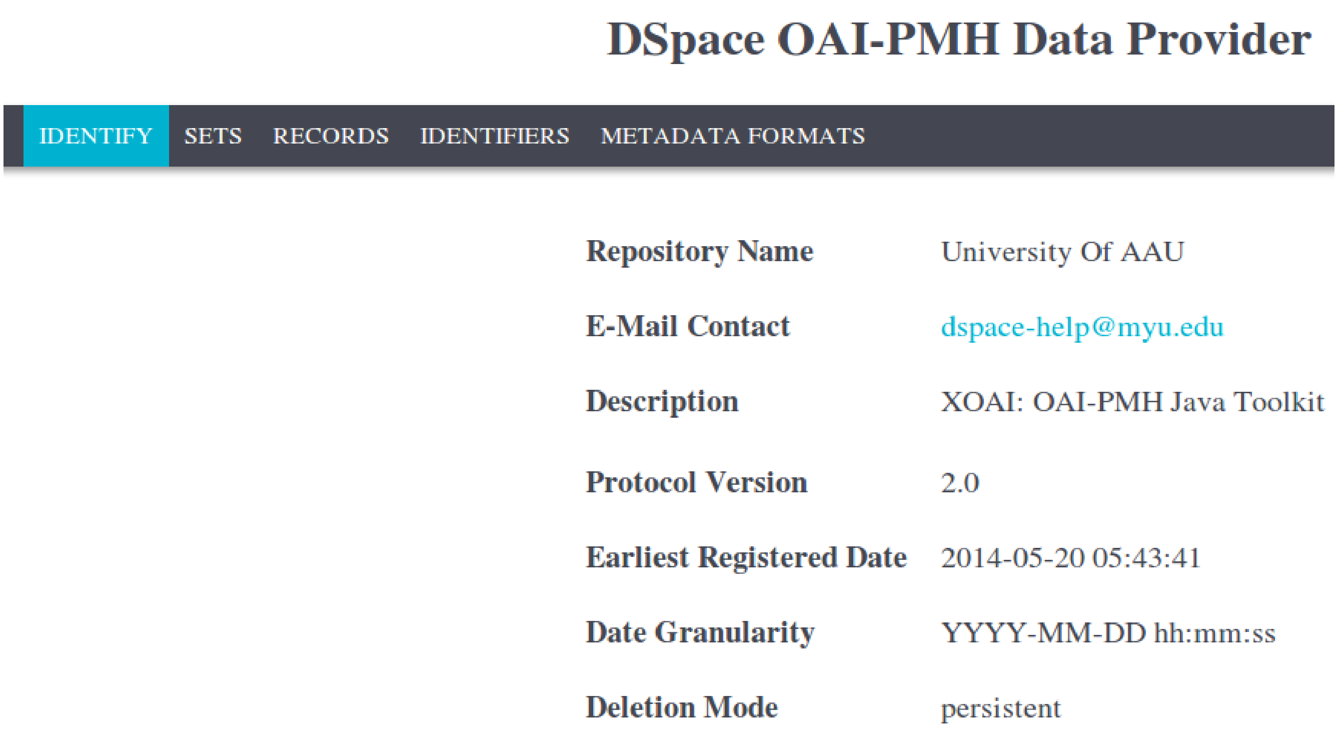


Figure 12 DSpace Data Provider Identify result

Figure 6.11 shows a screenshot of DSpace OAI-PMH data provider Identify verb response (ofcourse all OAI-PMH verb response are xml in fact DSpace using XSLT they change to human readable format).

OAI-PMH have seven request verbs the http://myDSpace.org/oai/request/verb?identify request in DSpace shows with repository name; email address of the system administrator; descriptions of Data provider toolkit; OAI-PMH have two versions the latest version 2.0 is implemented here; Earliest registered date of the system; date granularity method used includes seconds; deletion of the record is persistent.

OAI-PMH allows repositories to expose an hierarchy of sets in which records may be placed. A record can be in zero or more sets. DSpace exposes collections and communities as sets. Each community and collection has a corresponding OAI set, discoverable by harvesters via the ListSets verb. The setSpec is based on the community/collection handle, with the "/" converted to underscore to form a legal setSpec. The setSpec is prefixed by "com_" or "col_" for communities and collections. Because of Institutions structure are varying from one to other this study will not force them to use single set structure, thus institutions have the right to use their structure.

Unique Identifier

Every item in OAI-PMH data repository must have a unique identifier, which must conform to the URI syntax. As of DSpace 1.2, Handles are not used; this is because in OAI-PMH, the OAI identifier identifies the *metadata record* associated with the *resource*. The *resource* is the DSpace item, whose *resource identifier* is the Handle. In practical terms, using the Handle for the OAI identifier may cause problems in the future if DSpace instances share items with the same Handles; the OAI metadata record identifiers should be different as the different DSpace instances would need to be harvested separately and may have different metadata for the item. The OAI identifiers that DSpace uses are of the form: ***oai:PREFIX:handle.*** If you wish to use a different scheme, this can easily be changed by editing the value of identifier.prefix at [DSpace]/config/modules/oai.cfg file.

Access control

OAI provides no authentication/authorization details, although these could be implemented using standard HTTP methods. It is assumed that all access will be anonymous for the time being.

A question is, "is all metadata public?" Presently the answer to this is yes; all metadata is exposed via

OAI-PMH, even if the item has restricted access policies. The reasoning behind this is that people who do actually have permission to read a restricted item should still be able to use OAI-based services to discover the content. But, exposed data could be changed by changing the XSLT defined at *[DSpace]/config/crosswalks/oai/metadataFormats.*

Modification Date (OAI Date Stamp)

OAI-PMH harvesters need to know when a record has been created, changed or deleted. DSpace keeps track of a "last modified" date for each item in the system, and this date is used for the OAI-PMH date stamp. This means that any changes to the metadata (e.g. admins correcting a field, or a withdrawal) will be exposed to harvesters.

"About" Information

As part of each record given out to a harvester, there is an optional, repeatable "about" section which can be filled out in any (XML-schema conformant) way. Common uses are for provenance and rights information, and there are schemas in use by OAI communities for this. Presently DSpace does not provide any of this information, but XOAI core library allows its definition. This requires to dive into code and perform some changes.

Deletions

DSpace keeps track of deletions (withdrawals). These are exposed via OAI, which has a specific mechansim for dealing with this. Since DSpace keeps a permanent record of withdrawn items, in the OAI-PMH sense DSpace supports deletions "persistently". This is as opposed to "transient" deletion support, which would mean that deleted records are forgotten after a time.

Once an item has been withdrawn, OAI-PMH harvests of the date range in which the withdrawal occurred will find the "deleted" record header. Harvests of a date range prior to the withdrawal will *not* find the record, despite the fact that the record did exist at that time.

Flow Control (Resumption Tokens)

An OAI data provider can prevent any performance impact caused by harvesting by forcing a harvester to receive data in time-separated chunks. If the data provider receives a request for a lot of data, it can send part of the data with a resumption token. The harvester can then return later with the resumption token and continue.

DSpace supports resumption tokens for "ListRecords", "ListIdentifiers" and "ListSets" OAI-PMH requests. Each OAI-PMH ListRecords request will return at most 100 records (by default) but it could be configured in the [*DSpace]/config/crosswalks/oai/xoai.xml* file. Because of current OHS cannot handle resumption token we should avoid it in here.

DSpace by default expose 12 metadata schema's over OAI 2.0 request page EHIDL-AP schema to be available first EHIDL-AP.xsl must be created and put into *[DSpace]/config/crosswalks/oai/MetadatFormat* and then at *[DSpace]/config/crosswalks/oai/xoai.xml* Under *<Context baseurl="request"> <Format refid="EHIDL-AP"/>* and also include ***refid*** the ***prefix*** of metadata schema, the path the ***XSLT*** located, the ***namespace*** and ***schemaLocation*** must be described below under <Formats>

<Format id="EHIDL-AP">

<Prefix>EHIDL-AP</Prefix>

<XSLT>metadataFormats/EHIDL-AP.xsl</XSLT>

<Namespace>http://www.moe.edu.et</Namespace>

<SchemaLocation>http://www.moe.edu.et/ehidl_ap.xsd</SchemaLocation>

</Format>

6.2.2 Greenstone OAI-PMH data Provider

Greenstone incorporates a server that can serve any collection over the OAI Protocol for Metadata Harvesting. (However, it does not operate on the Windows Local Library server.) The address of the server is the same as the address of the digital library installation, but using the program *oaiserver.cgi* instead of the program *library.cgi*. For gsdl home page *http://mygsdl/greenstone/cgi-bin/library.cgi the oai will held on* ***http://mygsdl/greenstone/cgi-bin/oaiserver.cgi?verb=Identify***.

OAI 2.0 Request Results

Identify | ListRecords (oai_dc) | ListSets | ListMetadataFormats | ListIdentifiers (oai_dc)

You are viewing an HTML version of the XML OAI response. To see the underlying XML use your web browsers view source the page.

Datestamp of response	2014-05-30T14:37:00Z
Request URL	http://localhost:80/greenstone/cgi-bin/oaiserver.cgi

Request was of type Identify.

Repository Name	ASTU
Base URL	http://localhost:80/greenstone/cgi-bin/oaiserver.cgi
Protocol Version	2.0
Earliest Datestamp	2013-10-31
Deleted Record Policy	no
Granularity	YYYY-MM-DD
Admin Email	greenstone@cs.waikato.ac.nz

Figure 13 Greenstone OAI-PMH response for verb Identify

Configuration of the server is done via the oai.cfg file in the Greenstone etc directory. Basic of *repositoryName* and *repositoryId* fields, and also If the standard Apache setup that comes with Greenstone not used, set *oaiserverPath*, *libraryPath*, *docRootPath*. Optionally, you can set *baseServerURL* to use a domain name instead of IP address in URLs.

Greenstone OAI-PMH By default, collections are not accessible. Thus to enable the collection on ***[gsdl]/etc/oai.cfg*** edit (uncomment) *oaicollection* [*collectionname*]. And *oaisetname collname* "Name of Collection" *oaisetdescription collname* "Collection description".

Greenstone's OAI server currently supports Dublin Core, Qualified Dublin Core, RFC1807 metadata. For collections that use other metadata sets, including extracted metadata, metadata mapping rules should be provided to map the existing metadata to Dublin Core. Hence to use EHIDL-AP metadata schema we followed Appendix D.

Generally the overall OAI-PMH data provider is discussed in chapter five data provider section and DSpace data provider section, However in here for gsdl users how to configure OAI-PMH data provider is described therefore the next sections will elaborate a system which will harvest those metadata and present value add services.

6.3 EHIDLs OHS

OAI-PMH is basically an implementation of REST-based Web services protocols. REST architecture consists of a server and a client. REST client in the OAI-PMH uses GET and POST operations to retrieve metadata collections that are stored by the REST server. Data is sent from the server to the client in the form of XML documents. And OAI-PMH *service provider* launches a programme called *harvester* to visit a data providers and collect metadata in the format it requires, if this is available, at least in unqualified Dublin Core. The service provider processes the metadata gathered and offers a service based on those metadata. Digital library tools and standard implement service provider based on the value service add they want to include.

PKP Harvester2 is an open source metadata harvester and aggregator that have been developed by the Public Knowledge Project (Public Knowledge project, 2013) which aimed at expanding and improve access to global research. Harvester2 (Public knowledge project, 2010) has designed as a flexible tool for fetching, storing, indexing and searching data from different types of information sources (Liu, Kurt, Zubair, & Nelson, 2001). Harvester2 supports multiple harvesting protocols versions (OAI version 1 and 2), metadata standards (Dublin core, MODS, MARC), and languages with an emphasis on performance and simplicity of use (Public knowledge project, 2010). Among different types of harvesters, PKP Harvester2 provides easy management and installation of the base system. It can be further customized by designing new plugins, patches to the basesystem (Public knowledge project, 2010).

6.3.1 Architecture of OHS

PKP Harvester2 is written in object oriented PHP using the Smarty template system for user interface abstraction. Data is stored in a MySQL database, with database calls abstracted via the ADODB Database Abstraction library.

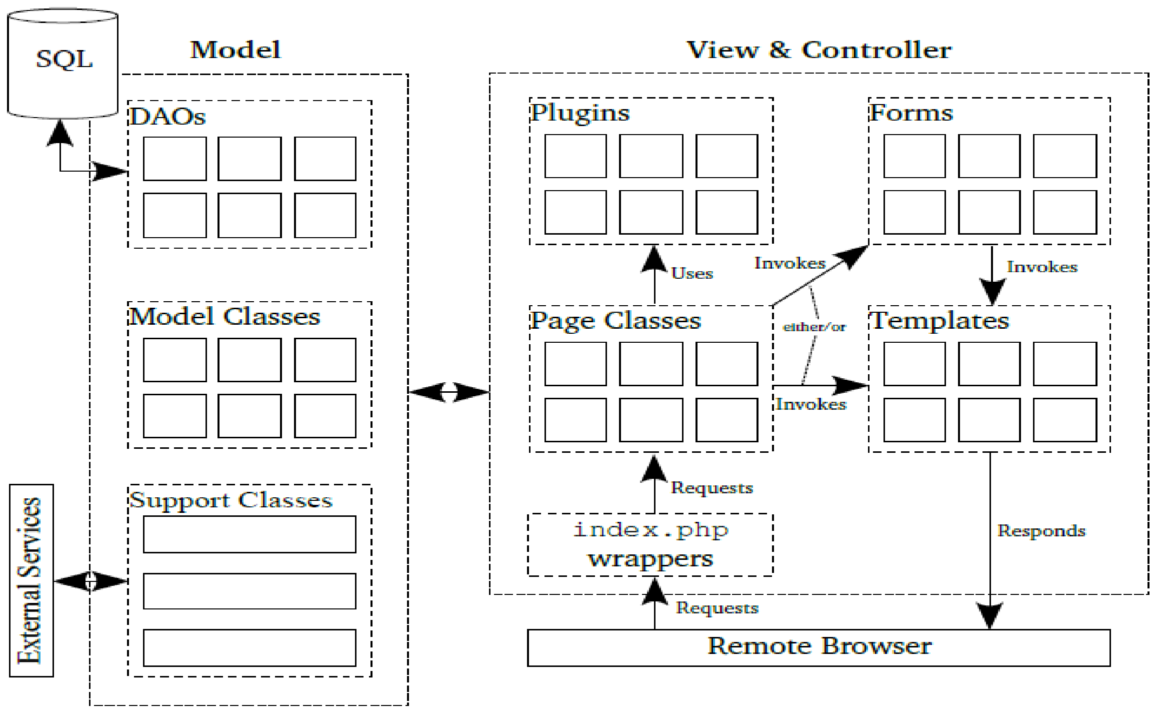


Figure 14 OHS architecture

The OHS architecture is quite similar with Sun's Enterprise Java Beans technology or the ModelViewController (MVC) pattern. As in a MVC structure, data storage and representation, user interface presentation, and control are separated into different layers.

OHS's **Model classes**, which implement PHP objects representing the system's various entities, such as Archives and Records; For example, the archives table stores archive information in the database; there is a corresponding Model class called Archive and DAO class called ArchiveDAO. Methods provided by Model classes are largely get/set methods to retrieve and store information, such as the getTitle() and setTitle($title) methods of the Archive class. Model classes are not responsible for database storage or updates; this is accomplished by the associated DAO class. **Data Access Objects** are used to retrieve data from the database in the form of Model classes, to update the database given a modified Model class, or to delete rows from the database. **Support classes**, which provide core functionalities, miscellaneous common classes and functions, more specifically on managing composing, addressing and sending email on the system as well as local configuration of language.

The **Forms class** and its various subclasses which are used by a Site Administrator to modify an Archive, centralize the implementation of common tasks related to form processing such as validation and error handling.

Pages classes receive requests from users' web browsers, delegate any required processing to various other classes, and call up the appropriate Smarty template to generate a response (if necessary). All page classes are located in the pages directory, and each of them must extend the Handler class. Additionally, page classes are responsible for ensuring that user requests are valid and any authentication requirements are met.

A **plugin** is a self-contained collection of code and resources that implements an extension of or modification to Harvester2. When placed in the appropriate directory within the codebase, it is loaded and called automatically depending on the **category** it is part of.

The **User Interface** is implemented as a large set of Smarty templates, which are called from the various Page classes. And those templates are responsible for the HTML markup of each page; however, all content is provided either by template variables (such as archive titles) or through Localespecific translations using a custom Smarty function.

6.3.2 Installation and Basic Configuration of EHIDL OHS

EHIDL OHS installation is tested on Ubuntu 14.04 LTS linux kerenel OS, with Mysql version 5.5.38 PHP 5.5.9-1ubuntu4.3 Apache2.47 Web server.

Downloading the current release (ohs 2.3.2) from http://pkp.sfu.ca/harvester2 and unpack it into a path on the web server of /var/www. The home page of http://www.localhost:8383/ohs some basic installation and configurations are also done.

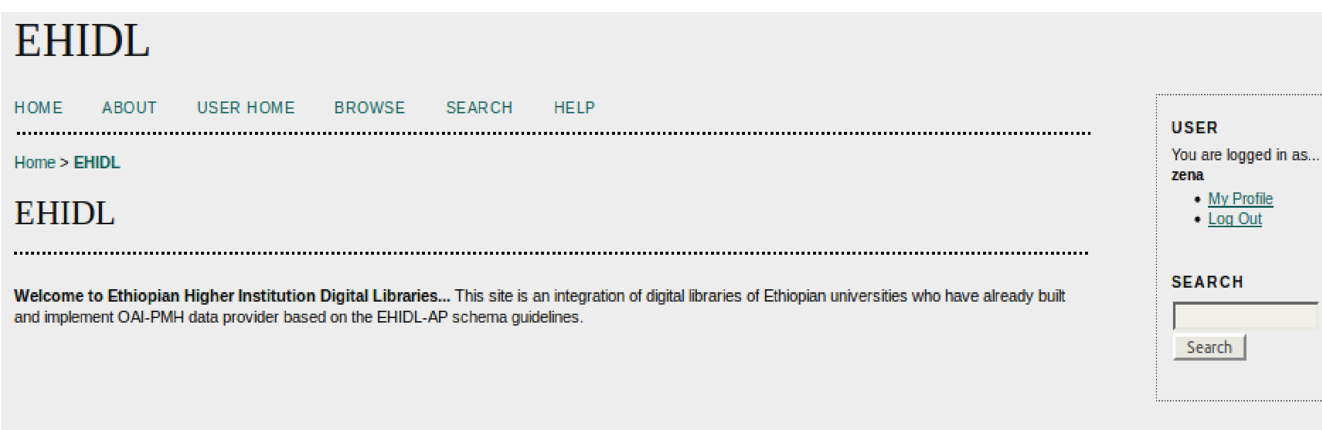


Figure 15 Home page of EHIDL Federated system

As you can see in Figure 15 the EHIDL OHS pages are classified with Home, About, User Home, Browse, Search and Help pages of main bar. Apart the User Home all pages are visible to the whole users, this pages only visible for the authenticate users.

And thus, for the administration page Under Site management categories of site setting link the site title, description, custom logo as well as name of the administrator including the administrator email is customized. Layout link configure how system wide the pages feels looks like as such on the right division of all pages the user profile and mysql search block added. For now the default language English is visible in language page. The Reading Tools provide relevant links to external resources like search engines and dictionaries. Many of the core features of the Harvester2 are implemented using plugins, which can be used to implement several additional functions such as filtering harvested data and extending metadata handling. Crosswalks are used to provide searching of and sorting by equivalent fields across multiple schemas. The About and Help pages which describes about PKP harvester2 and giving support how to use the System functionalities respectively no need of customization for this study so that the default is used. Because of the depth of the configuration or also the necessity of the customization of other functions they are described as sections.

6.3.3 Schema Plugins

An OHS **plugin** is a selfcontained collection of code and resources that implements an extension of or modification to Harvester2. When placed in the appropriate directory within the codebase, it is loaded and called automatically depending on the **category** it is part of.

Each plugin belongs to a single **category**, which defines its behavior. We implement EHIDL-AP schema Plugins in the schemas category which implement functions specific to a metadata schema loaded whenever a schema specific function is used. EHIDL-AP schema plugin interpret the defined OAI-PMH data provider can handle this metadata format and if it support harvest the metadata content based on this metadata format; as well as for display purpose when the record is browse or fetch from search result it classfied the element content by using date; string and language encoding syntax.

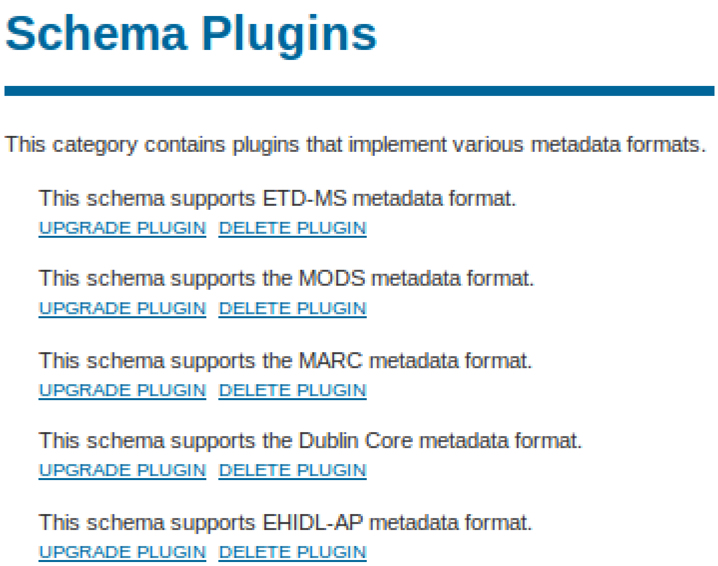


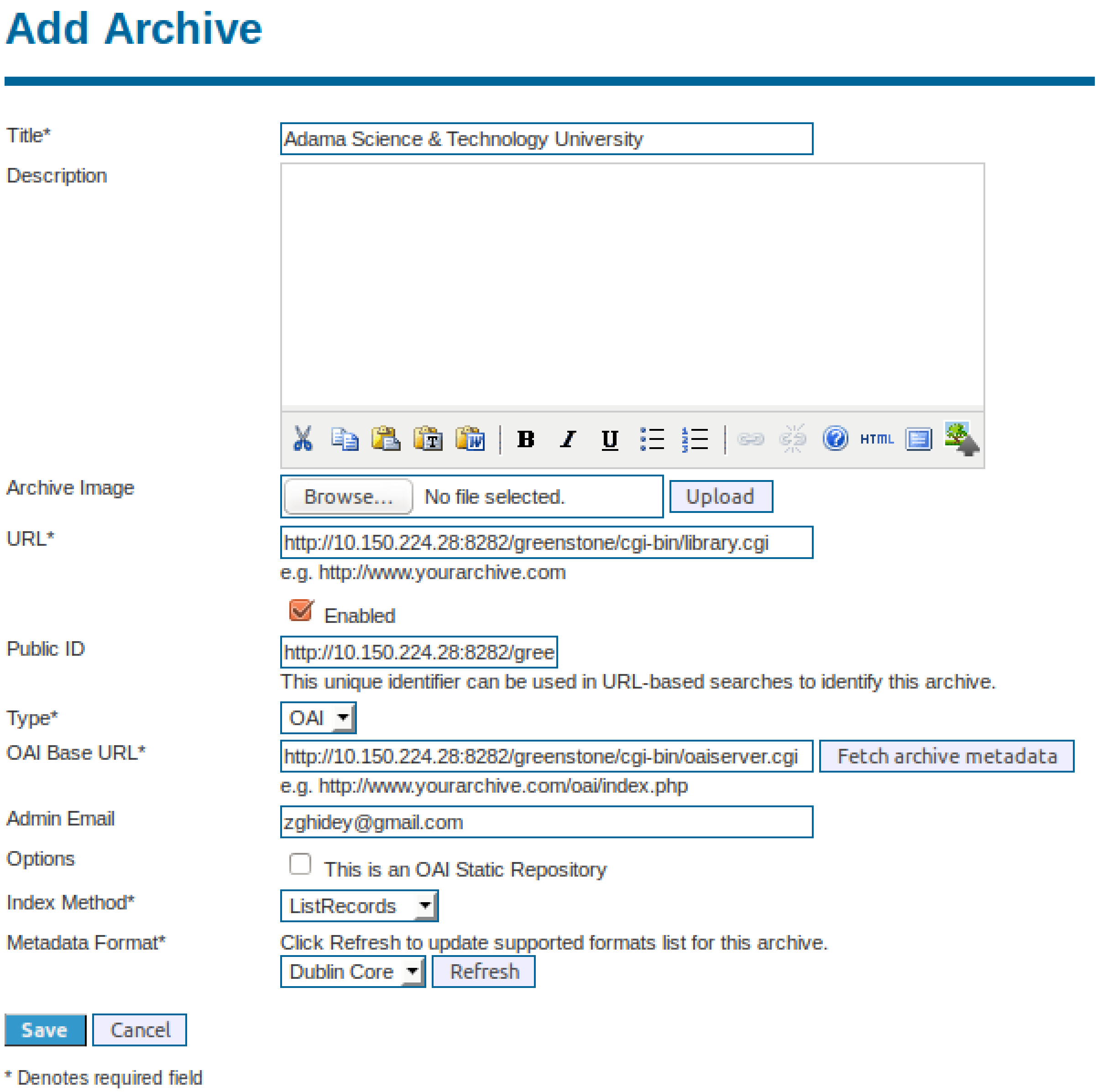
Figure 16 OHS Schema Plugin page

Figure 16 shows under schema plugins page which metadata formats are supported including upgrade plugin button which helps to upgrade from tar file and we can remove the metadata format using Delete Plugin button.

Generally EHIDL-AP schema plugins helps to interoperable the metadata schema the data provider format with the way OHS metadata schema support so that harvesting the metadata content with the each appropriate elements which resides.

6.3.4 Adding Archives

Those EHIDL digital libraries which implements OAI-PMH data provider must be registered by using the add archive form to OHS. As the figure 6.17 illustrated The Title, description and administration fields of add archive form are automatically extracted from the data provider. However the basic URL (the digital Library main page); and the OAI Base URL which it is the main OAI-PMH data provider implementations URL excluding the verb must be encoded correctly with they implements. We need first to harverst all records by using ListRecord verb refreshing the metadata format EHIDL-AP will be showed and save the archive. By using managing archives link we can also update metadata index and flush metadata after the archives are added.



Add Archive

Title* Adama Science & Technology University

Description

Archive Image Browse... No file selected. Upload

URL* http://10.150.224.28:8282/greenstone/cgi-bin/library.cgi

e.g. http://www.yourarchive.com

Enabled

Public ID http://10.150.224.28:8282/gree

This unique identifier can be used in URL-based searches to identify this archive.

Type* OAI

OAI Base URL* http://10.150.224.28:8282/greenstone/cgi-bin/oaiserver.cgi Fetch archive metadata

e.g. http://www.yourarchive.com/oai/index.php

Admin Email zghidey@gmail.com

Options This is an OAI Static Repository

Index Method* ListRecords

Metadata Format* Click Refresh to update supported formats list for this archive.

Dublin Core Refresh

Save Cancel

* Denotes required field

Figure 17 OHS Add Archive (Data Provider) Form

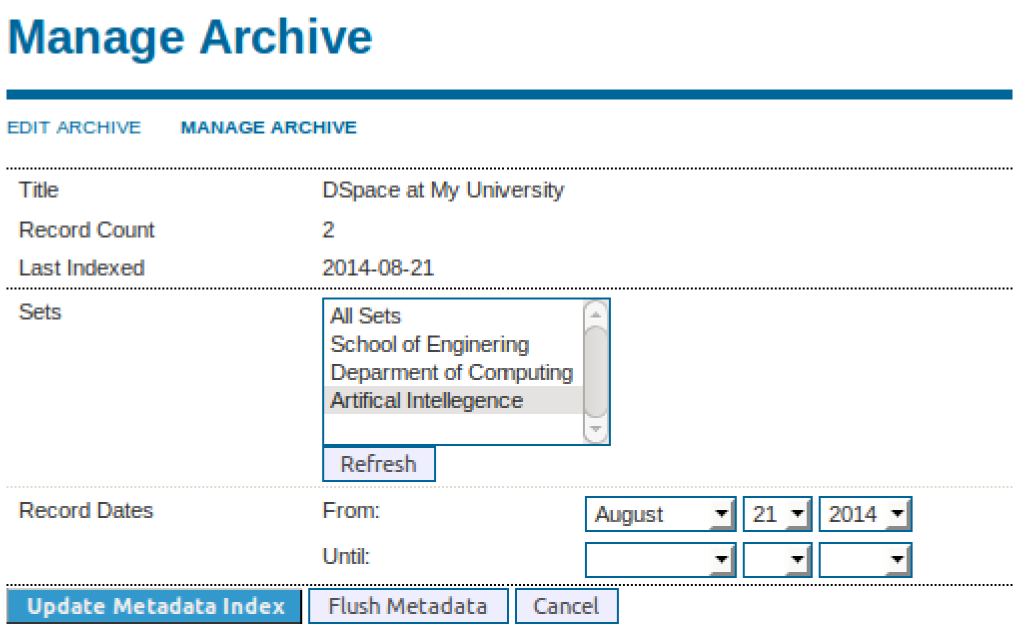


Figure 18 Mange Archive Form

Figure 18 After the data provider registered with the harvesting the first time, using manage archive shows how much record already harvest with date of last index; shows sets and we can manually update the metadata index.

6.3.5 Harvesting Metadata

EHIDL OHS harvest metadata from added archives. When harvesting a large archive using the webbased interface, it is possible for a timeout situation to occur; in this case, the archive will only be partially harvested. Because of that for large harvests we used command line harvesting mechanisms. Under the /var/www/ohs/tools path some commands we used are

$php harvest.php list # Lists archives shows

```
root@zena-HP-ProBook-6570b:/var/www/ohs-2.3.2/tools# php harvest.php list
Archive List
------------
6: DSpace at My University (2 records)
7: Adama Science and Technology University (13 records)
```

$php harvest.php [archiveId] usage # Display usage for the specified archive

$php harvest.php [archiveId] flush # Flush & harvest if all is used instead of archiveid all archives metadata will be harvested. Additional options are available, depending on the harvester plugin used, to implement features such as selective harvesting; use the above syntax to display usage for a full list of options. For example, the OAI harvester allows harvesting by sets and timestamps, just as it does in the web administration interface.

One useful strategy for keeping the metadata in Harvester2 up to date is to set up a cron job to reharvest archives every day or week. Thus we configured the crontabs to update the metadata index each day at mid night time. Even though OHS support full content harvesting but we only harvest the metadata of the content with valid URL of the resources residence. Thus the user if they want to access the resources the OHS will redirect to the resource URL using Identifier element.

6.3.6 Searching

Because of EHIDL OHS only harvest the metadata apart the content (resources), thus the system only index the metadata which make searching full text is not applicable. Searching resources can be done by inserting the query text and hit the search button which available on the right divisions of search forms; However for advanced search options using search main bar link including the simple search options as well as the search refinement can be done by narrow for specified archive (institutions digital library resources) and or also narrowing to specified metadata element. When searching with a specific archive or several archives of the same metadata format, all field names for that schema are allowed. Even though we recommend for EHIDL data providers to use EHIDL-AP schema for semantic level of interoperability crosswalk EHIDL-AP elements with other metadata schema elements also possible. Based on the metadata content value defined in EHIDL-AP schema plugins for date and language values the users can selective values otherwise they user must insert text query.

Search terms are case-insensitive and common words are ignored. It can generate search quires using simple Boolean operators (AND, OR, NOT) and more complex combination of quarries using parentheses e.g., archive ((journal OR conference) NOT theses). In addition to this harvester can search;

- for an exact phrase by putting it in quotes; e.g., "open access publishing"
- Exclude a word by prefixing it with - or NOT; e.g. online -politics or online NOT
- Use * in a term as a wildcard to match any sequence of characters; e.g., soci* morality would match documents containing "sociological" or "societal"

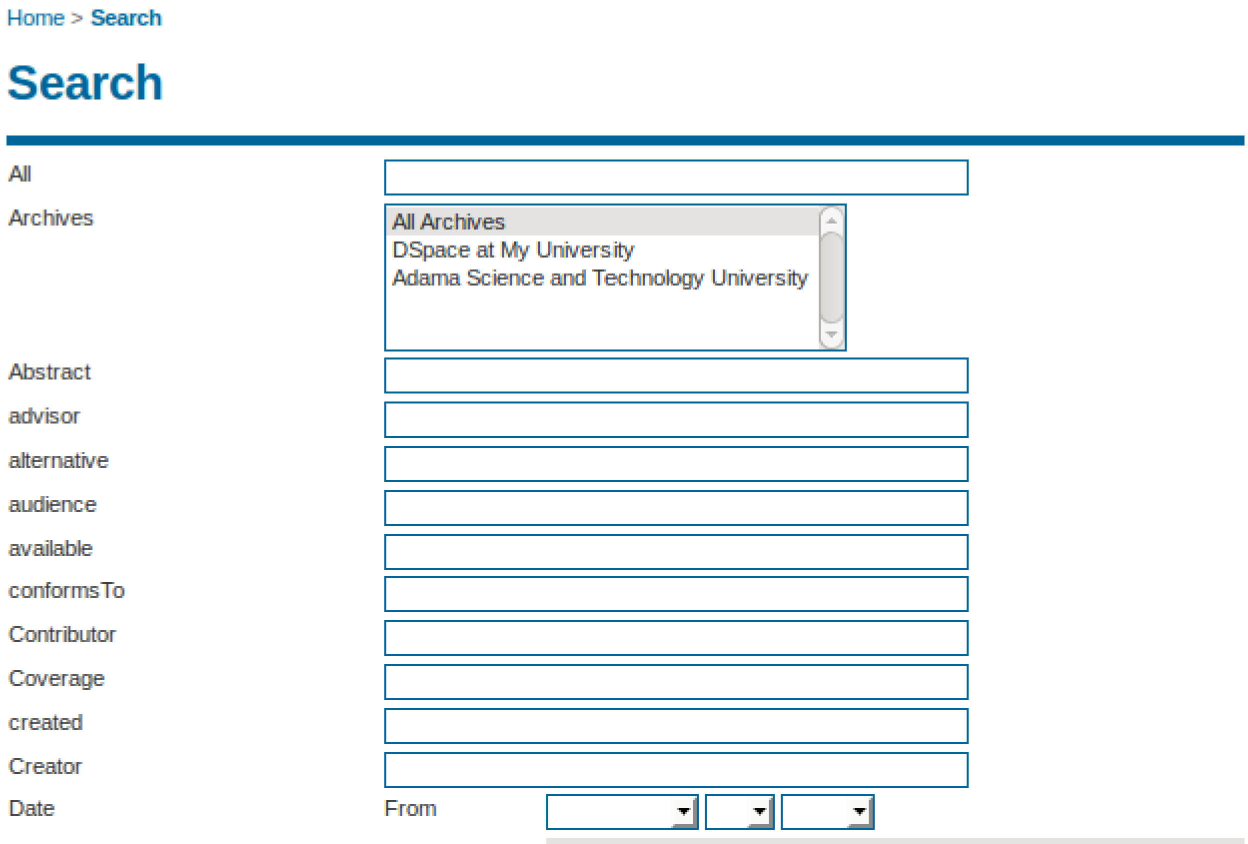


Figure 19 EHIDL OHS advance search form

As Figue 19 shows advance search form if the query hit in the all field it will search the query from whole element but the user can refine with the institutions. All left field describe the element where as the syntax of the query is specified if the archives or languages selected the user only chooses from the archive; if they want date they will select from and until date or they will insert the query text with the appropriate fields.

Figure 20 OHS Search Result

Fig 20 shows based on Enhancing query term searched the system will display the search result with highlighting match term with query term.

6.3.7 Browsing

One advantages of using metadata is support users to browse resources based on metadata elements such as title, subject and author etc or any other metadata elements. Therefore this project support users to browse for all archives or single archive based on Title, Author, Subject and Resource Type.

Figure 21 EHIDL Browsing records

Figure 21 shows for a single archive named DSpace at my university browsing records available the title and with new line author(s). View record will show the whole descriptions of the resources (all metadata elements filled) and view original is a link which will redirect to the resource beside to the digital library preserves.

6.3.8 Testing

To test a system, it is necessary to choose the set of test cases which will be used for the test. In particular it is very important to have a good coverage of each part of the fact bases. Therefore, the sets of test cases must be built very carefully. As in the case of any system after development of the system Testing and Evaluation is required to check for validity and verification. Therefore, the system is tested by the domain expert and the researcher whether it has achieved the desired output or not. Because of copyright issues and implementing Data provider for DLs takes a series time for testing we install & configure DSpace using Debian 7.03 and GSDL using Ubuntu 12.04 LTS. With the assumptions of each DL as single institution digital library then after from ASTU 5 instructors; 10 undergraduate students and 4 postgraduate students' total of 19 users are chosen to evaluate the EHIDL OHS.

		Criterions					
N$_{\underline{O}}$	Questions	Poor(1)	Fa(2)	Go(3)	V.G(4)	Exc(5)	Average
1	The easiness of the system to use and interact with it is?	0	0	0	7	12	
2	Attractiveness of the system is?	0	6	5	4	1	3
3	The efficiency of the system in time is?	0	0	0	6	13	
4	The accuracy of the system to get the e resources the user needs	0	0	4	6	9	
5	The completeness of advance search options	0	2	8	7	2	
6	The help and support of browsing service added to the users	0	0	3	9	7	
7	The sufficiency of the knowledge does the system incorporate all the users queries?	0	0	5	6	8	
8	How do you rate the significance of the system in the domain Area?	0	0	1	8	10	
Total Average							

Table 5 EHIDLs OHS user performance evaluation results

The evaluation and testing procedures help to address the question of user acceptance and accuracy of the implemented federated system. Visual interaction and questionnaire methods are used to assess user's acceptance issues and applicability of the federated system. Based on the evaluation results obtained from visual interaction with closed ended questions none of the evaluators respond as a poor. On the other hand evaluators reply fair 8 times (xxx%), good 26 times (xxx%), very good 53 times (xxx%), excellent 62 times (xxx%) and average 3 times (xxx%) the following table summarizes the results obtained on close ended questions.

Respondents who respond as	Poor (1)	Fair (2)	Good(3)	V.good(4)	Excellent(5)	Average
Total number	0	8	26	53	62	3
Percentage (100%)	0	5.26%	17.10%	34.86%	40.79%	1.98%

Table 6 Users evaluation result summary on closed ended questions

6.3.9 Evaluation

As one of this study objective is to design user friendly federated system user evaluated the proposed system was vital; therefore user evaluations for the new system shows that more than 75% users are evaluated very good and excellent to the federated system. We believe this federated system will helpful for EHILDs community still works to make more attractive are essential and we recommend users also involve on designing the interface.

Finally this study is a pilot projects to shows how Ethiopian higher institutions share their resources with simple implementations of OAI-PMH to interoperable with other digital library managements. The summary or the outcome of this study and the future works which researchers, librarians as well as the responsible persons should consider about also described in next chapter.

CHAPTER SEVEN

CONCLUSIONS AND FUTURE WORKS

This chapter presents the outcomes and future related works of this study specifically for Ethiopian higher institutions users' community as well as issues for digital library interoperability aspects. And both the conclusions part as well as future related works are also described based on the types of users will help and the discipline will addressed respectively.

7.1 Conclusions

The primary purpose of this study was to design and implement an interoperability system for Ethiopian higher institutions using the best suited protocols available for digital libraries by considering the institutions digital library management functions, so that in this study we tried to addressed.

- Describing Greenstone digital library tool and DSpace institutional repository management system which are used by those institutions who implement digital library to manage their electronic/digital resources for the provision of their particular institutions.

- Describe digital library standards are used, especially on metadata schema standards. Even though all institutions currently used Dublin core metadata schema with no guidelines how to use metadata schema so that this study propose to use an application profile called EHIDL-AP with metadata elements from different metadata schema; guidelines how to use each metadata elements and to ensure semantic interoperability controlled vocabularies also used.

- And EHIDL-AP implementations for both digital library tools are demonstrated with the hope librarians to enrich their resources metadata using it.

- Analyzing digital library interoperability approach and protocols among them metadata harvesting approach using OAI-PMH protocol is suited for EHIDL. Demonstration how Institutions who have digital library release their resources metadata using EHIDL-AP is also presented.

- Demonstrating EHIDL OHS federated system how archive (data providers) are added and managed; adding EHIDL-AP schema; harvesting metadata from archives; searching and browsing also presented.

Generally this study design and implement EHIDL federated system by evaluating the current state of digital library management based on metadata level of interoperability.

7.2 Future Work

Even though we tried to design and implement federated system for EHIDL community we believe there are so many jobs are still should be done. In addition most of effective digital library projects are done by collaboration with community from diverse disciplines across the globe so that evaluations of this study outcomes also vital.

- Evaluating EHIDL-AP schema elements, encoding syntax, controlled vocabularies and if necessity extending the elements as well as the elements contents.
- Creating data validator tools to content and metadata repository institutions to prevent data errors.
- Creating aggregator to join the same resources which could harvest from various institutions repository as well as expanding the semantic level of interoperability issues.
- Enhancing the searching and browsing functions of federated system and expanding value added services too.

<u>**APPENDIX A**</u>

Ethiopian Higher Institutions Digital Library – Application Profile (EHIDL-AP)

This part briefly describe the whole element sets, metadata content values used whether by restricted using from controlled vocabularies or encoding syntax.

A. EHIDL-AP Element Sets

EHIDLs-AP has element set with descriptions of **Element Name, Attribute**, **Definition**, **Comment**, **Refinement**, **Obligation** and **cardinality**, **metadata content**.

Element Name: - is a uniquely token assigned to the element term by the originator with prefix of the namespace. For example for term creator which is derived from Dublin core metadata intiative of Dublin core (dc), it will be write as ***dc:creator.***

Attribute: -some elements are still very broad in describing the resource therefore refinement using an attribute adds clarifications.

Definition:-A definition of the term given by the originator or EHIDL-AP. Term definitions by EHIDL-AP will not have different semantic meaning, where as it might have narrow or broad meaning to accommodate the semantic interoperability for this project.

Metadata content: - describe how the value of the element is present. Some element content must use controlled value from defined listed, and as well as the content syntax.

Obligation, Term could be ***Mandatory*** (the value must be present) or ***recommended*** (for the better of reaching representation) or ***optional***

Comment:- For metadata implementers general comments about the term Platform **Refinement**:-Some terms is very broad so they will have refinement. So the refined term will have the above description under the refined by term.

Cardinality: -the minimum and maximum occurrence of the term value.

1. <u>**dc:audience**</u>

It is a class of entity for whom the resource is intended or useful. Audience of EHIDL's is students, teachers, researchers as well as the community. The element content value must be from controlled vocabulary of EHIDL:audienceType. It is optional and repeatable.

2. <u>**dc:contributor**</u>

It is an entity responsible for making contributions to the content of the resource. Even though the primary responsible body for the content is author or creator, however sometimes there might be other types of significant contribution for the content may happen. Because of there are different kinds of contribution, this element has controlled values of EHIDL:contributorType**.** The element is optional and could be repeatable.

3. <u>dc:coverage</u>

The extent or scope of the content of the resource including geographical span and time of period it addresses. The element is refined by sub element of dc:spatial and dc:temporal.

3.1 *dc:spatial*

Spatial characteristics of the intellectual content of the resource. This element is very helpful especially for research works, in fact it could be applied to others works too. Some works are designed, evaluate or applied knowledge for specific places. The content value of the element is restricted to LCSH: Geographic values. It is optional and repeatable.

3.2 *dc:temporal*

Temporal characteristics of the intellectual content of the resource. For works designed, evaluate or applied knowledge at specific time. The content value of the element is restricted to LCSH:Chronological values. It is optional and repeatable

4. <u>dc:creator</u>

An entity primarily responsible for making the content of the resource. This includes the authors, painters, developer or researchers etc. It is highly recommended to fill the element content. And if in case there is more than one creator, repeat the element But follow the sequence of creator as they present. In Ethiopia Personals names sequence is different from western country therefore first use So use the western format.

5. <u>dc:date</u>

A date associated with an event in the life cycle of the resource. Date includes the time span associated the resources created, submitted and accepted by the responsible body as well as the availability and modification of the resources in the digital repository. The content value to be filled by librarian and digital repository software identify respectively.

7.3 ***dc:created***

Date of creation of the resource. This element is valuable for the resources content is time-based works. Date, month and year enough. It is optional and the element must only have one content.

7.4 ***dc:modified***

Date on which resource was changed. After the resources available, there may be modification made in the resources. It is mandatory if modification take place and the element must only have one content.

7.5 dc:available

Date (often a range) that the resource will become or did become available. Actually, the digital library software ***notify*** the date and time of source available during importing phase. The element content is mandatory and must have only one content.

7.6 ***dc:dateSubmitted***

Date of submission of the resource. One of EHIDLs resources are thesis or articles done by students, instructors or researchers with the specific time to performed. However this element content value is the date of the resource submitted to the responsible body for further step. The element content is optional and must only have one content.

7.7 dc:dateAccepted

Date of acceptance of the resource by the responsible body (e.g. of thesis by university department/institution, of article by journal, etc.).The element content is optional and must only have one content.

8 <u>dc:description</u>

An account of the content of the resource. Description includes abstract of the document as well as table of content of the documents.

8.1 dc:Abstract

The general descriptions of the work. Most of the time the creator of the resource put the abstract before the content chapters. The element content is recommended and must only have one content.

8.2 dc:TableOfContent

A list of subunits of the content of the resource. Which may either be recorded explicitly or provided as alink to such a list. The element content is recommended and have unbounded content.

9 <u>dc:format</u>

The physical or digital manifestation of the resource. The content must be MIMI values Still EHIDLs software's identify the resources format during building process. The element content is mandatory and must only have one content.

10 <u>dc:identifier</u>

An unambiguous reference to the resource within a given context. Identifier has attribute of DOI, ISBN, ISSN of the resource.

Any resource has given unique address or location in the repository called DOI. There fore DOI address must be valid of URI and The element content is mandatory and must only have one content.

International Standard Book number is also unique number given for manuscript type materials. Use 13 elements for ISBN, however for serial use serial number, issues number and journal name. The element content is recommended and must only have one content.

11 <u>dc:language</u>

The language of the intellectual content of the resource.RFC:4646 encoding system for international languages (use the first three letter of the language). The element content is recommended and must only have one content.

12 <u>dc:publisher</u>

An entity responsible for making the resource available. This element has one attribute EHIDL:place the place of the publisher. It is Optional and repatable.

13 <u>dc:title</u>

A proper name given to the resource. The element content is mandatory and must only have one content. Some resources might have an alternative title.

13.1 ***dc:alternative***

It is any form of the title used as a substitute or alternative to the formal title of the resource. Some resource differentiates main title and alternative using punctuation marks.

14 <u>dc:relation</u>

A reference to a related resource.this element describe relationship between resource to other resource(s). And there are so many relation types we want to work out.

14.1 dc:IsVersionOf

The described resource is a version, edition, or adaptation of the referenced resource. A change in version implies substantive changes in content rather than differences in format. The element content value must be string or numbers. The element content is recommended if applicable and must only have one content.

14.2 dc:IsFormatOf

The described resource is the same intellectual content of the referenced resource, but presented in another format.Don't use MIMI format for this element content value because the dc:format element content already existed use DOI of the referenced resource. The element content is recommended if applicable and must only have one content.

14.3 dc:IsPartOf

The described resource is a physical or logical part of the referenced resource. Use DOI of the referenced resource for the element content value.The element content is recommended if applicable and must only have one content.

15 dc:rights

Information about rights held in and over the resource. Rights element will contain a rights management statement for the resource, or reference a service providing such information. Rights information often encompasses Intellectual Property Rights (IPR), Copyright, and various Property Rights. If the Rights element is absent, no assumptions can be made about the status of these and other rights with respect to the resource. for each resources the element values is choice of A, {abstract} O, organization statement F, full access right.

16 dc:subject

The topic of the content of the resource. One way of intellectual materials classification is based on the subject they thought. Different subject classification schemas are existed, however we only used Library of Congress Subject Heading for general resources and Medical Subject Heading for medical resources. The element content is highly recommended and must only have one content.

Attribute dcs:LCSH with controlled values of LCSH:Topical values.

17 dc:type

The nature or genre of the content of the resource. Type includes terms describing general categories, functions, genres, or aggregation levels for content. The element content value must be controlled vocabulary of dc:dcmitype lists. The element content is highly recommended and must only have one content.

18 etd:degree

General descriptions of research thesis, articles. One of the resources content types managed by EHIDLs are thesis, research and thesis projectwhich might be done in their institutions or in the community by their students for final fulfillment their discipline.Thisresources will be refined by

18.1 *etd:degree.name*

Name of the degree associated with the work as it appears within the work. Highly recommend to fill the content using the department name which the paper is doing.

18.2 etd:degree level

level of education associated to the document. The element content value is restricted to EHIDL:degreeLevelType.

18.3 etd:degree discipline

Area of study of the intellectual content of the document. Use broad classification of EHIDL:discplinetype

18.4 ***etd:degree grantor***

Institution granting the degree associated with the work.

18.5 EHIDL:thesistype

the content type of the thesis, project, thesis and research

B. Controlled vocabularies

The element content value can be either from controlled vocabulary or any value based on encoding schema. Controlled vocabulary is list of value to be used for the specified element content. For element content that use controlled vocabulary is restricted to use only them. So this topic discuss the controlled vocabularies this project use with descriptions.

1. EHIDL:audeinceType

Is listed value of the resource targetaudience. Audience can be one Or more listed values.

2. EHIDL:contributorType

Contribute to the resource content may be in different kind. Therefore this
controlled vocabulary help to choose kind. Advisors, editors, conductor etc

3. LCSH:geographical

Lists of country code defined by Library congress

4. LCSH: chronological

Time based chronological aspects defined by Library congress

5. MIME

   Multipurpose Internet Mail Extensions (MIME) is an Internet standard that extends the format of email to support.

6. RFC:4646

   Structured format for language tags.

7. dc:dcmitype

   Cross-domain list of approved terms that may be used as values for the Type element to identify the genre of a resource by DCMI.

APPENDIX B

A questionnaire

Digital library interoperability becoming a paramount issue as the internet unites digital library systems of differing types, run by separate organizations which are geographically distributed all over the world. Federated digital library systems, in the form of cooperating autonomous systems are emerging in a bid to make distributed collections of heterogeneous resources appear to be a single, virtually integrated collection. Digital library interoperability unites isolated system (each universities digital resource) to share their digital resources among national level. Thus the major purpose of this study is to identify the best array of interoperability protocols for Ethiopian higher education digital libraries. And there are so many interoperability protocols are currently available and choosing among them could be appropriate by considering each universities digital libraries standards they used as well as service they provide. Therefore this questionnaire deals with to identify the digital libraries/ repositories software, metadata standards as well as technologies are they deploying.

Note: This study limited with identifying best suited interoperability protocols for the improvement of user interaction among Ethiopian higher education's digital libraries. Thus the questionnaire as well as the study as whole doesn't aim to evaluate or compare Ethiopian universities digital libraries.

Fill the blank with appropriate answers where as when the question is choice fill the red color to the appropriate choices.

a. Personal Information

1. Name of the institution ______________________________
2. Position

   Library Director

   Digital Librarian /System Administrator
3. Sex

   Male Female
4. Level of Education

   Diploma

   Degree

   Masters

   Phd and above

   Others please specify

b. Interoperability issue

1. Which Digital library softwares with versions are you using? (if you use more than one digital library software you can choose as much as you use)

    EPrints

    Greenstone DL

    Dspace

    Fedora

    If others (please specify)

2. In which operating System is the DL is running on and please specifies the version of your operating system.

    Microsoft Windows

    Linux

    Mac

    Others

3. Which services are your institutions digital library are providing.

    A search of a library's collection

    Reference and Question-answering Services

    Filtering and Selective Dissemination of Information

    Instructional Services

4. Which metadata are you using?

    Dublin core (Dc)

    Text Encoding Initiatives (TEI)

    Metadata Encoding and Transmission Standards (METS)

    Metadata Object Description Schema (MODS)

    Encoded Archive Description (EAD)

    Others please specify

5. Why you choose the above metadata…____.

6. Do you have any idea about below digital library interoperability protocols?

a. Z39.50

b. Open Archive Initiatives – Protocols for Metadata harvesting (OAI-PMH)

c. Open Archive Initiatives – Object Reuse and Exchange (OAI-ORE)

d. ResourceSync

e. Simple Web-service Offering Repository Deposit (SWORD)

f. Search Retrieve by Web-service/ URL (SRW/U)

g. Dinest Protocols

h. Simple Digital Library Interoperability Protocol (SDLIP)

If others please specify

6.1 From the above protocols which one are you currently using and also for what purposes are using.____

____

____.

6.2 What is the strongest side and weakness of the protocol are you using.__________

____

________________________________.

7. Finally could you please specify which features do you really think this study should be included____

____

____.

Thank you for your cooperation!!!

APPENDIX C

Interview

The main purpose of this interview is to identify the stakeholders, Developers and end users of the anticipated federated system.

a. Stakeholders Issues

1. How did you get your digital library collections? i.e from university budget, federal or regional governmental ministries, or other national and/or international organizations.
2. Could you describe your library relationships with the above stakeholders?
3. Who is responsible for the resources you mange on legal matters (copy right issues), and how it is handled.
4. Do you have any contract or subscribe digital information resources in your digital repositories to be used for your institutions only.

b. Developers Issues

1. How could you describe the structure of your digital library teams? Based on level of educations, expertise level, technical and professional skills.
2. How did they perform their task, do they have policy or standards to be followed.
3. Did they get regular and continuous training, workshops and seminars?
4. Did they complain about technical skill problem on their task? If they did which kinds of problem.
5. Do you have cooperation with any institutions/organization located outside your institutions on the issues of digital library projects?

c. End Users Issues

1. Who really are your end users, (i.e on information or/and computer literacy, cognitive styles, cultural (background), skill levels).
2. What are the rules or obligations must be fulfill to be digital library patrons.
3. How could you describe your library relationship with end users?
4. Do you have a team for orientation on "how to use the digital library", help and support for your end users.

APPENDIX D

Adding EHIDL-AP for GSDL

After EHIDL-AP metadata set created by Using GEMS Greenstone define schema's to use oai server there are different steps needs to follow

at ***GSDLHOME/runtime-src/src/oaiservr:***

Create a metaformat class for the metadata set. ehidl.h/cpp,

edit Makefile.in, Makefile SOURCES = \

ehidl.cpp \

OBJECTS = \

ehidl.o \

and win32.mak

SOURCES = \

ehidl.cpp \

OBJECTS = \

ehidl.o \

to use the new files

Edit recordaction.cpp to include the new header file and instantiate the new class (in recordaction()) if (metadataset.count("ehidl_ap") > 0) {

fptr.set_class(new dublin_core());

this->formatMap[fptr.get_class()->formatName()] = fptr;

}

Tell the server to use the new set: edit etc/oai.cfg and add the set name to the oaimetadata line.

oaimetadata oai_dc gsdl_qdc ehidl_ap

You may also need to add **oaimapping** information.

Recompile and test.

REFERENCES

- (Greenstone Wiki, 2014). Metadata in Greenstone. Available at: http://wiki.greenstone.org/doku.php?id=en:user:metadata (accesed at January 2013)
- (The DSpace Developer Team, 2013). DSpace 3.x Documentation. Accessed at https://wiki.duraspace.org/display/DSDOC3x
- African Digital Libraries Support Network (ADLSN n.d). Available at:http://adlsn.org/ethiopia (accessed December 10, 2012)
- Arms, W.Y (2002). Digital Libraries. MIT Press, available at: www.cs.cornell.edu/wya/DigLib/MS1999/Chapter7.html (accessed, December, 10, 2012).
- Blandford, A, Keith, S., Lodder, T. & Fields, B.(2002). Claims Analysis 'in the wild': a case study on Digital Libraries. Available at: www.uclic.ucl.ac.uk/annb/DLUsability/DLtools.html (Accessed Decmeber 15,2012)
- Blandford, A., Keith, S., Connell, I. & Edwards, H. (2004) Analytical usability evaluation for Digital Libraries: a case study. In *Proc. ACM/IEEE Joint Conference on Digital Libraries. 27-36.*
- Brueni, D.J., Cross. B.T., Fox, E.A., Heath,L.S.,Hix, D.,Nowell,L.T., &Wake,W. (1993). What if there were desktop access to the computer science literature?. *Proceedings of 21s Annual ACM Computer Science Conference: CSC '93(pp.15-22).Indianapolis, IN. NewYork: ACMPress.*
- David Bainbridge et al (2004). GREENSTONE DIGITAL LIBRARY DEVELOPER'S GUIDE. New Zealand digital library. New Zealand.
- David Bainbridge et al. Greenstone: A platform for distributed digital library applications.
- David Kellogg (2004). Open Source OAI Metadata Harvesting Tools. Accessed at https://diva.cdlib.org/projects/crawling harvesting/oai harvesting.pdf
- DELOS, (2005).Semantic Interoperability in Digital Library Systems. UKOLN, University of Bath.
- Digital Library Technology & Methodology Cookbook, (2011). An Interoperability Framework, Best Practices & Solutions. *D-Lib Magazine. 9 (1). ISSN 1082-987.*

- Ethiopian higher education proclamation (2009). proclamation no 650/2009. AddisAbeba. Federal Negarit Gazeta No.64.
- EU-NSF Digital Library Working Group (200). Interoperability between Digital Libraries Position Paper. Accessed at http://www.ics.forth.gr/isl/publications/paperlink/interop.htm
- EUROPEAN COMMITTEE FOR STANDARDIZATION (2003). Dublin Core Application Profile Guidelines. Management Centre: rue de Stassart. *CWA 14855:2003.*
- Feinstein, O. (2002). Use of evaluations and the evaluation of their use. Evaluation, 8, 433-439.
- Fukuda, R., & Bubb, H, (2003). Eye tracking study on Web use: Comparison between younger and elderly users in case of search task with electronic time table service. *PsychNology Journal, Volume1, Number3, 202-228.*
- Gail Hodge (2005). Metadata for electronic information resources: From variety to interoperability. *Information Services & Use .25. 35–45*.
- Gelaw, D and Kelly, S (2006). Developing the ICT infrastructure for Africa: Overview of barriers to harnessing the full power of the Internet. Journal of Education for Library and Information Science*, 47(1),pp.4–16.*
- Getachew, B., &Gojeh, L, (2007). Library and Information Science Education in Ethiopia. *Ethiopia Journal Education & Science, 4 (1)* .
- Goutam Biswas and Dibyendu Paul (2010). An evaluative study on the open source digital library softwares for institutional repository: Special reference to Dspace and greenstone digital library. *International Journal of Library and Information Science. 2(1) pp. 001-010.*
- Gowda & Bhandi, (2006). Metadata Harvesting and The Open Archives Initiative. *4th International Convention CALIBER-2006, Gulbarga. INFLIBNET Centre, Ahmedabad.*
- Gradmann (2009). Interoperability challenges on the digital libraries. Humboldt Universitätzu Berlin.
- Hevner & Chatterjee (2012). Design Research in Information Systems: Theory and Practice. NewYork: Springer Science + Business Media,.
- Hussien Suleman and Edward Fox (2001). The Open Archives Initiative: Realizing Simple and Effective Digital Library Interoperability. Accessed at http://shaun.dlib.vt.edu/projects/OAI/reports/jla_2001_article_oai.pdf

- Ian H. Witten and Stefan Boddie (2004). GREENSTONE DIGITAL LIBRARY INSTALLER'S GUIDE. New Zealand digital library. New Zealand.
- Ian Witten & David Bainbridge & David Nichols (2009). How to Build a Digital Library. Morgan Kaufmann. USA.
- Joshua Tauberer (2008). What is RDF and what is it good for?
- Kani-Zabihi, E., Ghinea, G. and Chen, S.Y. (2006), "Digital libraries: what do users want?", *Online Information Review, 30(4), pp. 395-412.*
- Kim, K. (2002), A model of digital library information-seeking process (DLISP model) as a frame for classifying usability problems, PhD dissertation, Rutgers University, New Brunswick, NJ.
- Kyrillidou, M,(2000). Research Library Trends: ARL Statistics. *Journal of Academic Librarianship, 26 (6) 427-436*. Accessed at: http://www.arl.org/stats/arlstat/jal99.html
- Lee,A.S.(2001)Editorial,MISQuarterly25(1),pp.iii–vii.
- MacKenzie Smith (2002). DSpace: An Institutional Repository from the MIT Libraries and Hewlett Packard Laboratories. Springer-Verlag.
- MacKenzie Smith et al (2003). DSpace An Open Source Dynamic Digital Repository.
- Mary S. Woodley (2008). Crosswalks, Metadata Harvesting, Federated Searching, Metasearching: Using Metadata to Connect Users and Information.
- Masanès, J.(2006)Collecting the hidden web. NewYork: Springer.
- Matera, M., Rizzo, F., & Carughi, G, (2002). Web Usability: Principles and Evaluation Methods. Italy. Politecnicodi Milano Piazza Leonardoda Vinci.
- Mathison, S.(2001).What's it like when the participatory evaluator is a "genuine" stakeholder? *American Journal of Evaluation, 22, 29-35.*
- Mehdi Alipour-Hafezi et al, (2010). Interoperability models in digital libraries: an overview. *The Electronic Library. 28 (3), pp. 438-452.*
- Muriel Foulonneau and David Dawson (2003). Open Archives Initiative – Protocol For Metadata Harvesting: Practices of cultural heritage actors.

- Nadjla Hariri, & Yaghoub Norouzi, (2011), "Determining evaluation criteria for digital libraries' user interface: are view", *The Electronic Library, 29 Iss: 5pp. 698 – 722.*
- NISO, (2004).Understanding Metadata. Bethesda. Nisopress.
- OECD, (2009).Understanding the Relationship between Economic Growth, Employment and Poverty Reduction. Accessed at: www.**oecd**.org/dac/povertyreduction/43280288.pdf
- Paepcke, A., Chang, C. K., Winograd, T., & García- Molina, H. (1998). Interoperability for Digital Libraries Worldwide. *Communications of the ACM, 41, pp 33-42.*
- Paihama, J., Williams, K., & Suleman, H. (2012). Assessing the Design of Web Interoperability Protocols. South Africa. University of Cape Town.
- Rao, R., Card, S. K., Jellinek, H. D., Mackinlay, J. D., and Robertson, G.G. (1995). Rich Interaction in the Digital Library. *ACM Press, 38 (4).*
- Robert Tansley et al (2003). The DSpace Institutional Digital Repository System: Current Functionality. *Joint Conference on Digital Libraries (JCDL'03). 0-7695-1939-3.*
- Saint, William (2004). Higher education in Ethiopia: the vision and its challenges, *Journal of Higher Education in Africa 2(3)pp 83-113*
- Sastry, G., & Reddy, C, (2009). User Interface Design Principles for Digital Libraries. *International Journal of Web Applications 1 (2).*
- Shiri, Ali (2003). Digital library research: current developments and trends. *Library Review 52 (2) pp. 198-202.*
- Suleman, Hussien (2002). Open Digital Libraries. PhD dissertation. Virginia Polytechnic Institute and State University. Blacksburg, Virginia Tech.
- Taylor, Arlene. (2004) The Organization of Information. 2nded. West port, CN: Libraries Unlimited.
- The Simple Digital Library Interoperability Protocol (SDLIP-Core). Accessed at http://diglib.stanford.edu:8091/~testbed/doc2/SDLIP/
- The International Standards Organization (1994, 10). ISO 9241-11 Guidance on Usability. Switzerland.

- Toms, E.G. (2002). Information interaction: providing a framework for information architecture. *Journal of the American Society for Information Science and Technology, 53 (10), pp 855-862.*
- UNESCO/OECD, (2003). Financing Education- Investments and Returns. UNESCO Institute for Statistics.
- University World News, (2009). ETHIOPIA: Electronic library system for universities. *Issue74. Acessed at:* http://www.universityworldnews.com/article.php?story=2009043019440541
- USAID/Ethiopia, (2012). Country Development Cooperation Strategy: 2011-2015. Public Version, Accessed at: **ethiopia**.**usaid**.gov/sites/default/files/images/CDCS-**Ethiopia**.**pdf**
- VanEimeren, B.,& Gerhard, H. (2000). ARD/ZDF Online Studie 2000: *Gebrauch swertent scheidet Media, 8/2000, 338-349.*
- Varatharajan, N & Chandrasekhara, M(2006).Digital library and interoperability: a general perspective.*4th International Convention CALIBER-2006, Gulbarga, 2-4. INFLIBNET Centre, Ahmedabad.*
- Wilson, T.D (1999). Models in information behavior research. *The Journal of Documentation 55 (3).*
- Witten et al (2005). StoneD: A Bridge between Greenstone and DSpace. *D-Lib Magazine. 11(9).*
- World Bank. (2003) Public expenditure review: Public spending in the social sectors – Ethiopia. Washington, D.C.: The World Bank.
- Xie, H. (2006). Evaluation of digital libraries: criteria and problems from users' perspectives. *Library & Information Science Research, 28, pp. 433-52.*
- Yizengaw, Teshome (2004). The Status and Challenges of Ethiopian Higher Education System and its Contribution to Development. *The Ethiopian Journal of Higher Education 1(1)*