Customization of DSpace for managing Multimedia Institutional Repository: The case of United Nations Economic Commission for Africa (UNECA)

By

YONAS AMSALU

A THESIS SUBMITTED TO THE COLLEGE OF NATURAL SCIENCES OF ADDIS ABABA UNIVERSITY IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF SCIENCE IN INFORMATION SCIENCE.

May, 2016
Customization of DSpace for managing Multimedia Institutional Repository: The case of United Nations Economic Commission for Africa (UNECA)

By

YONAS AMSALU

Name and Signature of members of Examination Board:

1. Dr. Million Meshesha, Advisor  
2. Dr. Dereje Teferi  
3. Dr. Tibebe Beshah

May, 2016
Declaration

I, the undersigned, hereby declare that the work on this thesis is based is my original work (except where acknowledgements indicate otherwise) and that neither the whole work nor any part of it has been, is being, or is to be submitted for another degree in Addis Ababa University or any other university.

I empower the university to reproduce for the purpose of research either the whole or any portion of the contents in any manner whatsoever.

Declared by:

Name: ______________________

Signature: ____________________

Date: _______________________

Confirmed by advisor:

Name: ______________________

Signature: ____________________

Date: _______________________

# Table of contents

Acronyms, and abbreviations .................................................................................................................. x
Abstract .................................................................................................................................................... xii

## CHAPTER ONE .................................................................................................................................... 1
Introduction................................................................................................................................................ 1
1.1 Background ......................................................................................................................................... 1
1.2 Statement of the problem .................................................................................................................. 3
1.3 Objective of the study ....................................................................................................................... 5
  1.3.1 General Objective ....................................................................................................................... 5
  1.3.2 Specific Objectives ..................................................................................................................... 5
1.4 Scope and Limitation of the Study .................................................................................................. 6
1.5 Significance of the Study ................................................................................................................. 7
1.6 Organization of the thesis ................................................................................................................ 7

## CHAPTER TWO .................................................................................................................................... 8
Literature Review ....................................................................................................................................... 8
2.1 Overview of multimedia .................................................................................................................... 8
  2.1.1 Digital Images ............................................................................................................................ 9
    Image file formats ............................................................................................................................ 9
    Image size – Resolution .................................................................................................................. 10
    Image size – file size ...................................................................................................................... 11
  2.1.2 Digital Audio ............................................................................................................................. 11
    Sound file formats .......................................................................................................................... 11
  2.1.3 Digital Video .............................................................................................................................. 14
    Video File Types ........................................................................................................................... 15
2.2 Institutional Repository ...................................................................................................................... 17
  2.2.1 Tools ........................................................................................................................................... 18
  Table 2.4 - Institutional Repository development tools ........................................................................ 21
IR software Comparison ......................................................................................................................... 21
Metadata .................................................................................................................................................. 22
Resource Discovery .................................................................................................................................. 23
Organizing Electronic Resources ........................................................................................................... 23
Interoperability ....................................................................................................................................... 23
CHAPTER FIVE

2.3 Challenges .................................................................................................................. 27
2.4 Related Work ............................................................................................................... 28

CHAPTER THREE ........................................................................................................... 29

Methodology ..................................................................................................................... 29
4.1 Literature Review ....................................................................................................... 30
4.2 Dataset Preparation ................................................................................................... 30
4.2.1.1 Inventory .......................................................................................................... 30
4.2.1.2 Reviewing and selection .................................................................................... 31
4.2.1.3 Digitization ....................................................................................................... 32
4.3 Implementation Tools for Customization of DSpace for managing MIR .............. 32
4.4 Evaluation Procedure ............................................................................................... 34

CHAPTER FOUR .............................................................................................................. 35

Metadata for Multimedia Management ........................................................................... 35

CHAPTER FIVE ................................................................................................................ 39

Customization of Multimedia Institutional Repository ................................................ 39
5.1 Data Preparation ........................................................................................................ 39
5.1.1 Multimedia item inventory .................................................................................... 39
5.1.2 Selection criteria .................................................................................................... 41
5.2.1 DSpace Installation ............................................................................................... 43
5.2.1.1 Oracle Java JDK 7 or OpenJDK 7 .................................................................... 43
5.2.1.2 Apache Maven 3.0.5+ (Java build tool) ............................................................... 43
5.2.1.3 Apache Ant 1.8 or later (Java build tool) Ant ..................................................... 43
5.2.1.4 Relational Database: (PostgreSQL or Oracle) Database ................................. 43
5.2.1.5 Servlet Engine (Apache Tomcat 7 or later, Jetty, Caucho Resin or equivalent) .... 44
5.2.2 User Interface Design ............................................................................................ 44
5.2.3 Customization ........................................................................................................ 48
5.2.3.1 Multimedia metadata implementation ............................................................... 48
5.2.3.2 A customization/implementation to handle multimedia items ......................... 48

Implementation of community, sub community, collection and items for MIR ............... 50
5.2.4 Testing .................................................................................................................... 50
5.2.4.1 Evaluation criteria ........................................................................................................... 50
5.2.4.2 User Acceptance Testing ............................................................................................... 51

CHAPTER SIX ...................................................................................................................................... 53
Conclusion and Recommendation .................................................................................................. 53

6.1 Conclusion .................................................................................................................................... 53
6.2 Recommendation .......................................................................................................................... 54

References ......................................................................................................................................... 55
Annex 1: Technology requirement and availability for video items ......................................................... 58
Annex 2: Technology requirement and availability for audio items .......................................................... 60
Annex 3: Technology requirement and availability for image items ......................................................... 61
Annex 4: User Acceptance Testing Questionnaire .................................................................................. 62
Lists of Tables

Table 2.1 - Common image file type format................................................................. 9
Table 2.2 - Lossy compressed audio format............................................................... 13
Table 2.3 - common video file type........................................................................... 15
Table 2.4 - Institutional Repository development tools............................................. 18
Table 2.5 - Institutional Repository software comparison.......................................... 21
Table 4.1 - Identified multimedia metadata standard for UNECA multimedia resources...........36
Table 5.1 - Inventory of multimedia items .................................................................40
Table 5.2 - Non digital items that needs technology to convert.....................................40
Table 5.3 - Non digital audio items that needs technology to convert.............................41
Table 5.3 – User acceptance testing questionnaire analysis.......................................51
Lists of Figures

Figure 5.1 - User interface of MIR prototype................................................................. 46
Figure 5.2 - A metadata code for Issue Date field.........................................................48
Figure 5.3 - A sample source code to handle .jpg image file type............................... 49
Figure 5.4 - Community of MIR (Macroeconomic Policy, Social Development, Regional Integration & Trade, Natural Resource, Innovation and Technology, Gender and Governance)...........50
Acknowledgement

First of all I would like to thank God for giving me the strength to finish my research work. I want to thank Dr. Million Meshesha my amazing advisor for giving me his guidance, support and advice throughout the course of this thesis's work.

This thesis owes its existence to the help, support, and inspiration of many people. It is a pleasant aspect that I have, now, the opportunity to express my gratitude for all of them.

I acknowledge UNECA for being a major data source for this research and the staff members for their collaboration and valuable assistance in the research.

My sincere thanks go to Lideya Mekonene for providing me invaluable insight and keeping me focus on doing this thesis. I would also like to thank Ismael Abdella for reviewing my thesis and offering me indispensable feedbacks.

Finally I also like to thank my beloved mother for her prayers and special wishes for successful completion of my MSc study and I owe special gratitude to my family & friends for their continuous and unconditional love, help and encouragement throughout my study.
**Acronyms, and abbreviations**

**AIIF**: Audio Interchange File Format

**ASF**: Apache Software Foundation.

**AVI**: Audio Video Interleave

**BMP**: BitMaP

**CAD**: Computer Aided Design

**CAM**: Computer Aided Manufacturing

**CSCs**: Common Service Centers

**CSS**: Cascading Style Sheet

**CSDGM**: Content Standard for Digital Geospatial Metadata

**DCMI**: Dublin core metadata initiative

**ECAMRS**: External Communications and Media Relations Section

**FGDC**: Federal Geographic Data Committee

**FLV**: Flash Live Video

**GIF**: Graphics Interchange Format

**HTML**: Hyper Text Markup Language

**JPG**: Joint Photographic Graphics

**PNG**: Portable Network Graphics

**MKV**: Matroska
MP3: MPEG Layer-3 Sound file

MP4: Motion Picture 4

METS: Metadata Encoding and Transmission Standard

MIR: Multimedia Institutional Repository

OAI-PMH: Open Archives Initiative – Protocol for Metadata Harvesting

TEI: The Text Encoding Initiative

W3C: World Wide Web consortium

WMA: Window Media Audio

WMV: Window Media Video

XML: Extensible Markup Language
Abstract

This research “Customization of DSpace for managing Multimedia Institutional Repository: The case of UNECA” focuses on identifying a proper metadata standard for multimedia items representation and a customization work on open source Institutional Repository software called DSpace for proper handling of multimedia items. The identified metadata standard helps for proper preservation and ease of access of multimedia items. This multimedia metadata standard is an extension of Dublin Core metadata standard, but it suitably customized to multimedia item management.

The core DSpace module can accept the multimedia items, but the representation of multimedia items on DSpace are the same that of pdf or any other word processing documents. It uploads the items and creates a link for downloading option. Any users on the repository can see the link to download the items on the repository. Due to the nature of the file, multimedia items should not be treated as a text file document, all uploaded items on the repository should be handled based on the item type. One of the main accomplishments of this research is customizing DSpace for the proper handling of multimedia items. The customized MIR application handles the multimedia items based on the item file type. If the item is image, it displays the image (the default DSpace option is creating a link to download image). If the item is audio/video, the system displays an audio/video player and the user can listen/view live without downloading the audio/video.

Based on the conducted user acceptance testing the users strongly agreed that MIR is beneficial to them and the customized prototype application is user friendly so that they can navigate on the application and access multimedia items easily.

The identified multimedia metadata enables the users to have a good idea of the item before actually downloading it. This proves useful in situations where accessing the item is costly. For example, let the item that we want to access be a video of several gigabytes, and after downloading it from the Internet we learn that it is not the video that we wanted. Here, we wasted a considerable amount of bandwidth for downloading the wrong content. But, if we first consulted the metadata of the video, we could have learnt that it is not what we wanted, with a much lesser wastage of bandwidth.

But on MIR user can get a detail information about each multimedia items on the metadata description and have a preview of the item without downloading it.
A detailed research is done on the area of multimedia, metadata standards, and institutional repository. The customized prototype of MIR is done by customizing DSpace. The customization is mainly done by implementing the identified metadata standard and programmatically changing the DSpace item displaying method to handle multimedia item.

**Keywords:** Multimedia, DSpace, Institutional Repository, MIR, Metadata standards
CHAPTER ONE

Introduction

1.1 Background

The rapid expansion of the field Information Technology led to explosive growth in exchange of information and services. Along with the advancement of communication and access to information, various applications, that take advantage of the communication infrastructure, have come to existence [1].

Due to this, historical and operational resources are being built up in the digital form day by day. But these digital contents are, at the same time, disappearing on a daily basis. On the other hand, the technology provides opportunities for recording, archiving and disseminating digitalized information on historical and operational resources, creating space for interoperation dialogue and new operational expressions. In addition, large digital libraries become available and huge amounts of materials are accessible over the Internet with the increase of communications bandwidth and storage capacities [1].

The accumulating vast and growing amount of data created holds a vast quantity of information and knowledge in a variety of formats, including printed and electronic. These represent the corporate memory, providing historical evidence of its actions and decisions. These information resources include multimedia materials which are outlined as text, image, audio and video. These are typically the elements for the building blocks for generalized multimedia environments, platforms, or integrating tools [1].

Texts can be stored vary greatly. In addition to ASCII based files, text is typically stored in processor files, spreadsheets, databases and annotations on more general multimedia objects. With availability and proliferation of GUIs, text fonts, complex special effects (color, shades) making the storing of text is complex.

There is great variance in the quality and size of storage for still images. Digitalized images are sequence of pixels that represents a region in the user's graphical display. The space overhead for still
images varies on the basis of resolution, size, complexity, and compression scheme used to store image. The popular image formats are JPG, PNG, BMP, GIF, and TIFF.

An increasingly popular data type being integrated in most of applications is Audio. It’s quite space intensive. One minute of sound can take up to 2-3 Mbs of space. Several techniques are used to compress it in suitable format. The popular audio media types are: MP3, and WMA.

Video is also another most space consuming multimedia data type. The digitalized videos are stored as sequence of frames. Depending upon its resolution and size a single frame can consume up to 1 MB. Also to have realistic video playback, the transmission, compression, and decompression of digitalized require continuous transfer rate. The popular video media types are: MP4, AVI, MKV, WMV, and FLV.

Variants of image are graphic objects that are consisting of special data structures used to define 2D and 3D shapes based on which we can define multimedia objects. These include various formats used by image, video editing applications. Examples are CAD / CAM objects.

Information Technology has evolved to store and retrieve continuous media data types. These systems are expected to play a major role in information systems regarding the handling of multimedia items for preserving and making it available in well-organized manner. A challenging task when implementing these systems is to support continuous display of multimedia objects. This is because of the difference in media types and the technological issues related to the multimedia objects.

Digital multimedia materials are increasingly available for public accesses. However, the rate at which access to this key element of multimedia computing is growing, but unparalleled with the tools and techniques that have been developed for retrieving and browsing it.

Multimedia Institutional Repository (MIR) concerns the preservation and dissemination of the institutions multimedia resources. For the customization of Multimedia Institutional Repository the first task is to collect multimedia resources such as images, audios, and videos and make them available into digital format. The second task is to find a new automatic process to create metadata and annotation of these digital contents so as to improve access to multimedia information, retrieval and interaction with such information. The third task is to ensure interoperability to maintain the distributed multimedia repository system. The fourth task is to develop suitable interfaces to the Multimedia Institutional Repository such that various devices (PC, tablets, and smaller mobile phones)
handle the presentation of multimedia information and adapt themselves to fit users’ preferences, capabilities, and needs? [1]

Institutional Repository was first developed as an online solution for collecting, preserving, and disseminating the scholarship of universities, colleges and other research institutions. With a variety of platforms available, an institution looking to start a repository program or move to the state of the art platform has many options and features to compare. Locally hosted software offers customizations unique to the institution, but requires repository programming and IT teams to build and maintain. The cloud-based, hosted platforms offer a turnkey solution with consistent platform versions, upgrades, and customer support that will assist in developing a successful repository.

1.2 Statement of the problem

The easy availability of digital cameras, smart phones and other advanced devices become the reason for the rapid increasing of multimedia data; however the technologies or tools that enable applications to use these digital medias effectively and efficiently are still in their infancy. One of the facets of problems that surround such systems is finding efficient ways to summarize the huge amount of data involved. This is because multimedia material is sheer volume, complex, unstructured format and content rich, due to this content-based access to a multimedia data is often impeded. Therefore, efficient and effective methodologies for organizing and manipulating the vast volume of multimedia data are highly needed.

The United Nations Economic Commission for Africa (UNECA or ECA) was established by the Economic and Social Council (ECOSOC) of the United Nations (UN) in 1958 as one of the UN's five regional commissions, ECA's mandate is to promote the economic and social development of its member states, foster intra-regional integration, and promote international cooperation for Africa's development. Made up of 54 member states, and playing a dual role as a regional arm of the UN and as a key component of the African institutional landscape, ECA is well positioned to make unique contributions to address the Continent’s development challenges [8].

In its more than 50 years of existence, UNECA has created and holds information and knowledge in a variety of formats, including printed and electronic format. Currently UNECA implemented an Institutional Repository project that handles the institutions electronic material on online web based application. But due to lack of deep research & different technological issues regarding the multimedia data management, MIR is not yet implemented in UNECA. Due to this fact, priceless historical
multimedia items (images, audios, and videos) are scattered into different divisions and the safety and proper reservation of the items are in question. Moreover those items are not easily accessible and not described properly. The ultimate goal of this research is to customize a Multimedia Institutional Repository web application by creating an online solution for collecting, preserving and disseminating multimedia items [9].

In the current developmental stage of information and knowledge industry which is creating a number of platforms of information exchanges that facilitate the adequate accessibility of information for the various purposes. For the matter of that, information and information resources must be manipulated, evaluated, stored and finally retrievable in adequate manner. These Information resources might be publications, reports, articles or the information can be in the form of multimedia products.

UNECA multimedia resources are initiated to establish a well standardized multimedia digital archive preservation system. However, UNECA multimedia resources are not well organized to provide standardized information retrieval services. As a result, all of the images and audio visual services are not well provided which frustrates, misleads, and lags behind the customer’s information seeking behavior.

Most of speeches and press releases are already available on UNECA website and the UNECA External Communications and Media Relations Section (ECAMRS) tries to collect not all but most of the multimedia products in their studio. These multimedia products are in different formats. Most of the old multimedia materials are not in a digital format, to convert those resources to their digital form, some of technologies or equipment’s are unavailable in UNECA compounds. The other issue is that the old multimedia materials need the preventive maintenance so that it doesn’t lose its valuable information. Moreover the UNECA Multimedia resources have no metadata records of the publications and an incomplete collection of all UNECA Multimedia resources.

The aim of this study is to explore optimal ways to create a web-based Multimedia Institutional Repository for UNECA. To this end, the research attempts to investigate and answer the following research questions:

- How the dataset preparation of MIR is conducted? What are the steps included?
- What is the suitable multimedia metadata standard identified to clearly describe the multimedia items?
• What are the issues identified on the customization of DSpace for managing Multimedia Institutional Repository?
  o How is the identified multimedia metadata standard implemented on a web-based MIR?
  o How the customization/development is done?
  o How to make the multimedia items (images, audio, and video) accessible on the web-based MIR?
  o How to make MIR friendly with mobile devices?
• What kind of evaluation procedure followed to test the performance of the developed Multimedia Institutional Repository?

1.3 Objective of the study

1.3.1 General Objective

The main objective of this study is to customize DSpace for managing Multimedia Institutional Repository (MIR) system for United Nation Economic Commission for Africa (UNECA) so as to make the multimedia materials properly preserved and easily accessible to information seekers.

1.3.2 Specific Objectives

To achieve the general objective of the current study, the following specific objectives are identified:

• To review literatures and analyze methods and approaches for Multimedia Institutional Repository (MIR).
• To collect and prepare a dataset for representing the multimedia objects through in deep study on the digital multimedia material format,
• To identify a metadata standard to represent multimedia items on Multimedia Institutional Repository (MIR)
• To customize DSpace to handle a Multimedia Institutional Repository (MIR) system that handle multimedia items (text, images, audio and video) and accessible via a mobile devices.
• To evaluate the performance of the proposed prototype.
1.4 Scope and Limitation of the Study

The focus of the research is on the customization of DSpace for a web-based Multimedia Institutional Repository for UNECA. To manage the content of a multimedia material and its metadata under an Institutional Repository environment, a suitable repository model is required. To this end, a multimedia data repository model is defined for text, image, audio and video and its constituent units are proposed. It is defined in such a way that the perceptual features (i.e. visual and temporal features) and high-level descriptions of a multimedia data are captured so as to provide multi-criteria query support.

Specifically the followings are covered on this research:

- Review different literatures on the area of multimedia items type and Institutional Repository.
- Preparing the dataset for the customization of MIR by doing multimedia item inventory in UNECA and identify the selection criteria.
- Identification of metadata standard for UNECA multimedia materials, this is the customization of a well-known metadata standard called Dublin Core for a better representation of multimedia materials in UNECA.
- Customization of DSpace for managing a web-based Multimedia Institutional Repository.
  - Review different development tools and identify the most appropriate one.
  - Implementing the identified multimedia metadata into the system so that each multimedia item can be represented based on the identified metadata.
  - Customization of the system to handle multimedia items.
  - Develop mobile-device ready MIR that can make the multimedia items searchable and accessible in an easy way.

This research is limited to customization of DSpace for managing a Multimedia Institutional Repository for United Nations Economic Commission for Africa (UNECA). But it is highly customizable and the methodologies & techniques identified on this research can be used for other institutes too.

There are some tasks that are defined on the scope of this research but not done:

- Development of Album Viewer for image items. For a better viewing/representing of images on MIR, a group of image needs to be displayed in Album format so that the user can view it in
attractive way, but in developed prototype a group of images are listed not displayed in album viewer.

- Customization of DSpace for multimedia items on this research is mainly focused on the making available of multimedia items in MIR web-based application. The aesthetic values have got fewer priorities.

- Development of the mobile-device ready MIR is focused on the responsiveness of the web-based application pages based on the different device (PC, tablets, and mobile phones) accessing it. Not developing a mobile app for Apple, Android, or Microsoft devices.

1.5 Significance of the Study

This research helps UNECA for proper handling of multimedia items and making it available to information seekers. In current situation UNECA holds multimedia items for the past 50 years, due to the lake of proper multimedia items preservation and technological solution, some priceless old multimedia items are in danger and also it is not accessible to user.

The research output not only preserves the multimedia items, it also identifies a proper multimedia metadata and makes it easily accessible to information seeker with the customization of DSpace for managing MIR.

1.6 Organization of the thesis

This thesis document is organized as follows: Chapter 2 presents a detailed discussion on different core concepts of this research like multimedia items and its different types, the concept related to the Institutional Repository with respect to tools and technologies, and metadata standards for the proper representation of items. Chapter 3 is dedicated for metadata for multimedia management; it is identified a metadata standard for proper representation of multimedia items. Chapter 4 discusses the methodology used for this research. Chapter 5 discusses the customization of DSpace for managing of Multimedia Institutional Repository. Finally, a conclusion of the thesis work and suggestions for future work are outlined in Chapter 5.
CHAPTER TWO

Literature Review

The Internet is maturing, bandwidths are expanding, new software is being created, and hard drive capacities are on the rise. It has never been easier to create or distribute multimedia. For instance video, one of the most common multimedia formats, has become so widely available that a popular website YouTube.com regularly handles over 100 million video downloads per day [9].

The rapid emergence of multimedia based (including Web-based) information systems in recent years presents a serious challenge for the development skills of information systems professionals. Until recently, many such systems were simple, stand-alone applications like electronic brochures. Some have now become extensive, complex, e-commerce applications. As Web-based systems integrate with organizational activities such as customer support, sales and marketing, and distribution and technical support, Web system developers will encounter similar development issues as conventional and multimedia systems development. Therefore, there is a need to follow the more sophisticated and disciplined approach toward systems development, which includes resolving various problems that accompany large scale multimedia systems development.

2.1 Overview of multimedia

Multimedia has become an inevitable part of any presentation. It has found a variety of applications and uses multiple forms of information content and information processing (e.g. text, audio, graphics, animation, and video) for targeted use.

Multimedia is usually recorded and played, displayed, or accessed by information content processing devices, such as computerized and electronic devices, but can also be part of a live performance. Multimedia devices are electronic media devices used to store and experience multimedia content.
2.1.1 Digital Images
Pictures are the most common and convenient means of conveying or transmitting information. Pictures concisely convey information about positions, sizes and inter-relationships between objects. They portray spatial information that we can recognize as objects. Human beings are good at deriving information from such images, because of our innate visual and mental abilities. About 75% of the information received by human is in pictorial form [11].

Digital imaging has become more than just a popular pastime in contemporary culture. Personal computers, in addition to a plethora of hand-held electronic devices, have become the preferred mode of communication for increasingly large portions of the population. Digital cameras and camera phones have made taking, processing, and sharing photos almost instantaneous, making digital images a common component of hi-tech communications.

Image file formats
It is also important to understand the common image file formats of digital images, how these file formats differ, and what their recommended use is. As shown in table 2.1 TIFF (.tif), JPG (.jpg, .jpeg), GIF (.gif) and PNG (.png) are file formats (and their respective file extensions) that you are likely to encounter. Other image file formats are used to a lesser extent; these formats are often proprietary, such as Adobe Photoshop .psd files [13].

<table>
<thead>
<tr>
<th>Name</th>
<th>File Format</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tagged Image File Format</td>
<td>TIFF</td>
<td>A lossless file format that can be compressed. This format is widely supported across operating systems. TIFF is the best file format for archiving high quality images.</td>
</tr>
<tr>
<td>Joint Photographic Graphics</td>
<td>JPG or JPEG</td>
<td>A format that was specifically created for photographs, and can contain millions of colors. JPGs are automatically compressed (Can choose the level of compression to match the desired image quality), resulting in a relatively small file size while still retaining quality. For this reason, JPGs are ideal for email and Web use. JPGs are lossy, discarding information each time that they are compressed.</td>
</tr>
</tbody>
</table>
Graphics Interchange Format
GIF
The lossless and compressed file format that is preferred for graphics, because it keeps edges and lines sharp. GIFs are limited to 256 or fewer colors, and are not recommended for photographs, but rather for images with flat fields of color, such as clip art. GIFs can be static or animated.

Portable Network Graphics
PNG
Portable Network Graphics format, an open source substitute for GIFs. PNGs provide a higher lossless compression rate than GIFs, and help to reduce cross-platform differences in image display quality, among other technical advantages. PNG provides a useful format for the storage of images during intermediate stages of editing.

<table>
<thead>
<tr>
<th>Table 2.1- Common image file type format</th>
</tr>
</thead>
</table>

**Image size – Resolution**
Resolution refers to the sharpness and clarity of image. Actually, the sharpness and clarity vary depending on the context of the image. When scanning images, it usually prompted to choose what resolution would like the images scanned at. The intended use or display of the images will help to make this choice [13].

Images intended for high-quality printing should be scanned in at no less than 150dpi, preferably 300dpi (dpi is “dots per inch”), and the more dots per inch, the sharper and crisper the images will look on the printed page but requires more memory space.

**Computer display:** Images intended primarily for display on a computer monitor (such as email attachments or images on a Web page) really only need a 72ppi to 96ppi resolution, as that is what computer monitors are capable of displaying. A higher resolution will not make image appear any better on the screen. In fact, if we don’t expect to print the images, these images should be “optimized”, which means making them as small and compressed a file as possible (while still retaining the appropriate visual clarity), so that they load quickly, and do not take up too much space on the hard drive [13].
Image size – file size
File size is proportional to the pixel dimensions of an image, is partially determined by the file format, and it specifically refers to how much disk space the file occupies, either on the hard drive, CD, or other storage medium. Image files normally range in sizes measured in Kilobytes (KBs) and Megabytes (MBs); extremely large files may be measured in Gigabytes (GBs). Images for print will have larger file sizes, since higher resolution means more data stored in the file. Images for Web or computer display will have smaller file sizes. As a general rule (and there are appropriate times and places to break it), Web images should fall between 10KB and 200KB; images with larger file sizes will load more slowly, and generally be too cumbersome for effective computer distribution and display. Saving images in .JPG or .GIF format will automatically compress and reduce the file sizes significantly, so that it can fit more files onto a disk or attach more to an email message, without exceeding the disk storage or file size limits [13].

2.1.2 Digital Audio
Digital audio is a technology that can be used to record, store, generate, manipulate, and reproduce sound using audio signals encoded in digital form.

Digital sound first reached the general public in 1982 by means of the computer disc (CD) format, a 12-cm optical disc read by laser. The CD format was developed jointly by the Philips and Sony corporations after years of development. It was a tremendous commercial success, selling over 1.35 million players and tens of millions of discs within two years [14].

Sound file formats
An audio file format is a file format for storing digital audio data on a computer system. The bit layout of the audio data (excluding metadata) is called the audio coding format and can be uncompressed, or compressed to reduce the file size, often using lossy compression. The data can be a raw bit stream in an audio coding format, but it is usually embedded in a container format or an audio data format with defined storage layer.

An amazingly varied number of sound file formats exist within audio and multimedia production. Here are the most commonly used audio production formats that don’t use data compression [14]:

- Wave (.wav) - The Microsoft Windows format supports both mono and stereo files at a variety of resolutions and sample rates. WAV files contain PCM coded audio (uncompressed Pulse Code Modulation formatted data) that follows the Resource Information File Format (RIFF)
specifically, which allows extra user information to be embedded and saved within the file itself.

- Broadcast wave (.wav) - In terms of audio content, broadcast wave files are the same as regular wave files; however, text strings for supplying additional information can be embedded in the file according to a standardized data format.

- Wave64 (.w64) - This proprietary format was developed by Sonic Foundry, Inc. (now operating under the Sony name). In terms of audio quality, Wave64 files are identical to wave files, except that their file headers use 64-bit values (instead of Wave’s 32-bit values). As a result, Wave64 files can be considerably larger than standard wave files, and this format is a good choice for long recordings (e.g., surround files and file sizes over 2 GB).

- Apple AIFF (.aif or .snd) - this standard sound file format from Apple supports mono or stereo, 8-bit or 16-bit audio at a wide range of sample rates. Like broadcast wave files, AIFF files can contain embedded text strings.

**Lossless compressed audio format**

A lossless compressed format stores data in less space without losing any information. The original, uncompressed data can be recreated from the compressed version.

Uncompressed audio formats encode both sound and silence with the same number of bits per unit of time. Encoding an uncompressed minute of absolute silence produces a file of the same size as encoding an uncompressed minute of music. In a lossless compressed format, however, the music would occupy a smaller file than an uncompressed format and the silence would take up almost no space at all.

**Lossy compressed audio format**

Lossy compression enables even greater reductions in file size by removing some of the audio information and simplifying the data. This of course results in a reduction in audio quality, but a variety of techniques are used, mainly by exploiting psychoacoustics, to remove the parts of the sound that have the least effect on perceived quality, and to minimize the amount of audible noise added during the process. The popular MP3 format is probably the best-known example, but the AAC format found on the iTunes Music Store is also common. Most formats (see table 2.2 – below) offer a
range of degrees of compression, generally measured in bit rate. The lower the rate, the smaller the file and the more significant the quality loss.

<table>
<thead>
<tr>
<th>Name</th>
<th>File Extension</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>3gp</td>
<td>.3gp</td>
<td>multimedia container format can contain proprietary formats as AMR, AMR-WB or AMR-WB+, but also some open formats</td>
</tr>
<tr>
<td>Act</td>
<td>.act</td>
<td>ACT is a lossy ADPCM 8 kbit/s compressed audio format recorded by most Chinese MP3 and MP4 players with a recording function, and voice recorders</td>
</tr>
<tr>
<td>Audio Interchange</td>
<td>.aiff</td>
<td>Standard audio file format used by Apple. It could be considered the Apple equivalent of wav.</td>
</tr>
<tr>
<td>File Format</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Audio</td>
<td>.au</td>
<td>The standard audio file format used by Sun, Unix and Java. The audio in .au files can be PCM or compressed with the μ-law, a-law or G729 codecs.</td>
</tr>
<tr>
<td>MPEG 4 (audio)</td>
<td>.m4a</td>
<td>An audio-only MPEG-4 file, used by Apple for unprotected music downloaded from their iTunes Music Store. Audio within the m4a file is typically encoded with AAC, although lossless ALAC may also be used.</td>
</tr>
<tr>
<td>Protected</td>
<td>.m4p</td>
<td>A version of AAC with proprietary Digital Rights Management developed by Apple for use in music downloaded from their iTunes Music Store.</td>
</tr>
<tr>
<td>Layer-3 Sound file</td>
<td>.mp3</td>
<td>MPEG Layer III Audio. Is the most common sound file format used today</td>
</tr>
<tr>
<td>Ogg, Oga</td>
<td>.ogg, .oga</td>
<td>A free, open source container format supporting a variety of formats, the most popular of which is the audio format Vorbis. Vorbis offers</td>
</tr>
</tbody>
</table>
A raw file can contain audio in any format but is usually used with PCM audio data. It is rarely used except for technical tests.

Standard audio file container format used mainly in Windows PCs. Commonly used for storing uncompressed (PCM), CD-quality sound files, which means that they can be large in size—around 10 MB per minute. Wave files can also contain data encoded with a variety of (lossy) codecs to reduce the file size (for example the GSM or MP3 formats). Wav files use a RIFF structure.

Windows Media Audio format, created by Microsoft. Designed with Digital Rights Management (DRM) abilities for copy protection.

Royalty-free format created for HTML5 video.

<table>
<thead>
<tr>
<th>Raw image format</th>
<th>.raw</th>
<th>A raw file can contain audio in any format but is usually used with PCM audio data. It is rarely used except for technical tests.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Window Media Video</td>
<td>.wav</td>
<td>Standard audio file container format used mainly in Windows PCs. Commonly used for storing uncompressed (PCM), CD-quality sound files, which means that they can be large in size—around 10 MB per minute. Wave files can also contain data encoded with a variety of (lossy) codecs to reduce the file size (for example the GSM or MP3 formats). Wav files use a RIFF structure.</td>
</tr>
<tr>
<td>Window Media Audio</td>
<td>.wma</td>
<td>Windows Media Audio format, created by Microsoft. Designed with Digital Rights Management (DRM) abilities for copy protection.</td>
</tr>
<tr>
<td>WebM</td>
<td>.webm</td>
<td>Royalty-free format created for HTML5 video.</td>
</tr>
</tbody>
</table>

Table 2.2 – Lossy compressed audio format

2.1.3 Digital Video

Digital video refers to the capturing, manipulation, and storage of moving images that can be displaced on computer screens [11]. This requires that the moving images be digitally handled by the computer.

Due to bandwidth constraints, use of video should currently be minimized on the web. Eventually, video will be used more widely, but for the next few years most videos will be short and will use very small viewing areas. Under these constraints, video has to serve as a supplement to text and images more often than it will provide the main content of a website.

Digital video via Internet

Due to the size of some large video files and the limited bandwidth, transmission of video and audio data via the Internet is only possible using streaming technology. The current big players in the streaming media business are Real Networks and Microsoft. But, it was big communication companies such as CNN who first offered this technology. As noted by Fisher & Schoreder Now Real Network’s Real Audio and Real video streaming formats are the pseudo standards [10].
Integrating digital video into a Web Page

There are some pros and cons in putting a video on a web page. For one, it could be very compelling and could really enhance the interest level of the site, thus attract people to view the page. On the other hand, poorly designed video (or animation), can distract viewers. If the video is too large, it will take a long time to load, and the viewer will grow impatient waiting on a blank screen and will move on to another website. A good rule of thumb is to not let video be the only way to access information. Offer alternatives: preview clips with lower resolution and shorter duration, a still picture with some captions, exported sound files, movies of smaller size windows. For example, if we have four videos on a page, by showing a still image of each with a sentence or two describing the video we would allow the viewer to decide which they wanted to take the time to watch and would not need to start streaming the ones in which they are not interested.

Because of the limited bandwidth and the traffic load on the net, it is important to compress a digital video file as much as feasible in order to deliver good quality video. However, do not take a movie that’s already compressed for CD ROM and re-compress it for the web. Compress the originally captured files with light or no compression and reedit and compress those, to ensure quality. Another point is if the video to use on a page uses a smaller window than the original, reduce the window before compress it to get a smaller file. As far as color, if the compressor lets to choose the number of colors, choose the greatest number. Lowering the number of colors generally doesn’t lower the final file size enough, but it might lose the quality of colors and lighting in the final product. Nevertheless, if we know that the video uses 256 colors, there is no need to go beyond that.

Video File Types

The Video Files category includes a wide range of video formats, which use different codecs to encode and compress video data. This category also includes video project files and video information files, which may not contain video data.

Table 2.4 presents the list of most common video file types

<table>
<thead>
<tr>
<th>Name</th>
<th>Extension(s)</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>WebM</td>
<td>.webm</td>
<td>Free and libre format created for HTML5 video.</td>
</tr>
<tr>
<td>Matroska</td>
<td>.mkv</td>
<td></td>
</tr>
<tr>
<td>Format</td>
<td>Extension</td>
<td>Description</td>
</tr>
<tr>
<td>---------------------</td>
<td>-----------</td>
<td>-------------</td>
</tr>
<tr>
<td>Flash Video (FLV)</td>
<td>.flv</td>
<td>Use of the H.264 and AAC compression formats in the FLV file format has some limitations and authors of Flash Player strongly encourage everyone to embrace the new standard F4V file format.[10] De facto standard for web-based streaming video (over RTMP).</td>
</tr>
<tr>
<td>F4V</td>
<td>.flv</td>
<td>Replacement for FLV.</td>
</tr>
<tr>
<td>Vob</td>
<td>.vob</td>
<td>Files in VOB format have .vob filename extension and are typically stored in the VIDEO_TS folder at the root of a DVD.[13] The VOB format is based on the MPEG program stream format.</td>
</tr>
<tr>
<td>AVI</td>
<td>.avi</td>
<td>Uses resource interchange file format.</td>
</tr>
<tr>
<td>QuickTime File Format</td>
<td>.mov, .qt</td>
<td></td>
</tr>
<tr>
<td>Windows Media Video</td>
<td>.wmv</td>
<td></td>
</tr>
<tr>
<td>MPEG-4 Part 14 (MP4)</td>
<td>.mp4, .m4p (with DRM), .m4v</td>
<td></td>
</tr>
<tr>
<td>MPEG-1 Video</td>
<td>.mpg, .mp2, .mpeg, .mpe, .mpv</td>
<td>Old, but very widely used due to installed base.</td>
</tr>
<tr>
<td>MPEG-2 Video</td>
<td>.mpg, .mpeg, .m2v</td>
<td></td>
</tr>
<tr>
<td>M4V</td>
<td>.m4v</td>
<td>Developed by Apple, used in iTunes. Very similar to MP4 format, but may optionally have DRM.</td>
</tr>
<tr>
<td>3GPP</td>
<td>.3gp</td>
<td>Common video format for cell phones</td>
</tr>
<tr>
<td>3GPP2</td>
<td>.3g2</td>
<td>Common video format for cell phones</td>
</tr>
</tbody>
</table>
RealMedia (RM) .rm Made for RealPlayer

| Ogg Video | .ogv, .ogg | Open source |

Table 2.3 – Common video file type

Response time

Many multimedia elements are big and take a long time to download with the horribly low bandwidth available to most users. It is recommended that the file format and size are indicated in parentheses after the link whenever we point to a file that would take more than 15 seconds to download with the bandwidth available to most of the users. If we don't know what bandwidth the users are using we should do a survey to find out since this information is important for many other page design issues. At this time, most home users have at most 28.8 Kb, meaning that files longer than 50 KB need a size warning. Business users often have higher bandwidth, but we should probably still mark files larger than about 200 KB.

Finally, digital video are becoming more popular and accessible through the various media technology advances which enable users to capture, manipulate and store video data in efficient and inexpensive ways. With the increasingly efficient compression formats and easiness of integrating videos in web pages, more people are able to enjoy producing and publishing movies in the digital world.

### 2.2 Institutional Repository

The easy availability of digital cameras, smart phones and other advanced devices become the reason for the rapid increasing of multimedia data; however the technologies or tools that enable applications to use these digital medias effectively and efficiently are still in their infancy. One of the facets of problems that surround such systems is finding efficient ways to summarize the huge amount of data involved. This is because multimedia material is sheer volume, complex, unstructured format and content rich, due to this content-based access to a multimedia data is often impeded. Therefore, efficient and effective methodologies for organizing and manipulating the vast volume of multimedia data are highly needed.

An Institutional Repository is an online locus for collecting, preserving, and disseminating, in digital form, the intellectual output of an institution.
There are different international tools (software packages) and standards used to build Institutional Repository.

**Tools**

There are open-source software packages highly used for building open source repositories that are complaint with the Open Archives Initiative (OAI). Sample Open-source software packages for running a repository includes DSpace, Eprint, Fedora, CDSware, and Bepress [15].

<table>
<thead>
<tr>
<th>Tool</th>
<th>Description</th>
<th>Technology</th>
</tr>
</thead>
</table>
| DSpace  | DSpace is an open source software package that provides the tools for management of digital assets, and is also commonly used as the basis for an institutional repository. It supports a wide variety of data, including books, theses, 3D digital scans of objects, photographs, film, video, research data sets, and other forms of content. The data is arranged as community collections of items, which bundle bit streams together. | **Programming language:** Java and JSP, using the Java Servlet API  
**Database:** PostgreSQL and Oracle.  
**Web Server:** Any  
**Search Engine:** Lucene  
**Standard specification:** OAI-PMH v2.0 and is capable of exporting METS (Metadata Encoding and Transmission Standard) packages.  
**Operating system:** Unix/MACOSX/Windows |
<table>
<thead>
<tr>
<th><strong>EPrints</strong></th>
<th>Eprints is an open source software package for building open access repositories that are compliant with the Open Archives Initiative Protocol for Metadata Harvesting. It shares many of the features commonly seen in Document Management systems, but is primarily used for institutional repositories and scientific journals. EPrints has been developed at the University of Southampton School of Electronics and Computer Science and released under a GPL license (GNU General Public License).</th>
</tr>
</thead>
</table>
| **Fedora** | Flexible Extensible Digital Object Repository Architecture) The Fedora Repository Project and the Fedora Commons community forum are under the stewardship of the DuraSpace not-for-profit organization.  
Fedora was originally developed by researchers at Cornell University as an architecture for storing, managing, and accessing digital content in the form of digital objects inspired by the Kahn and Wilensky Framework. Fedora defines a set of abstractions for expressing digital objects, asserting relationships among digital objects, and linking "behaviors" (i.e., services) to digital objects. |

**Programming Language:** 
Perl  
**Database:** MySQL  
**Web Server:** Apache 1.3  
**Search Engine:** N/A  
**Standard specification:** OAI-PMH v2.0  
**Operating system:** successfully run under Linux, Solaris, and Mac OS X

**Programming Language:** 
Java  
**Database:** MySQL/MCKoi/Oracle  
**Web Server:** Tomcat 4.1  
**Search Engine:** Database  
**Standard specification:** OAI-PMH v2.0  
Fedora provides RDF support and the repository software is integrated with semantic triple store technology, including the Mulgara RDF database.
| CERN Document Server Software (CDSware) | CERN Document Server Software (CDSware) is a suite of applications which provides the framework and tools for building and managing an autonomous digital library server. CERN as the international European Organization for Nuclear Research has been involved since its early beginnings with the open dissemination of scientific results. The dissemination started by free paper distribution of preprints by CERN Library and continued electronically via FTP bulletin boards, the World Wide Web to the current OAI-compliant CERN Document Server. |
| Bepress (the Berkeley Electronic Press) | It is licensed by the Berkeley Electronic Press (Bepress is taken as its abbreviation). It is used by associations, consortia, universities and colleges to preserve and showcase their scholarly output. Digital Commons is one of their products. Institutions can add their content to their repository through batch uploads, by linking to external sites, or via a submit form. Digital Commons allows for a variety of publication types and auto-converts Word, WordPerfect, and RTF documents to PDF. A unique web page is generated automatically for each article that includes title, author, abstract, and citation information. All pages maintain a persistent URL and meet web accessibility standards. Digital Commons supports data harvesting and feeding. Content is optimized for fast and accurate indexing by Google and Google Scholar and is OAI compliant. Digital Commons provides user notification tools. This includes RSS feeds and automatic email notification for reports of newly published content, Mailing list manager to announce new research, and the "Tell a colleague" email |

**Programming language:** Python/PHP  
**Database:** MySQL  
**Web Server:** Apache/PHP/Python  
**Search Engine:** cdsware  
**Standard specification:** OAI-PMH v2.0
Digital Commons also provides individual readership statistics.

Table 2.4 - Institutional Repository development tools

**IR software Comparison**

Institutional Repositories were first developed as an online solution for collecting, preserving, and disseminating the scholarship of universities, colleges, and other research institutions. The repository quickly evolved into a platform for libraries to publish and showcase the entire breadth of an institution's scholarship including articles, books, theses, dissertations, and journals. Added support for images, video, audio, and other media has brought greater depth to repository collections. Since 2000, a number of repository platforms have been developed, each with their own set of benefits and technical criteria as presented in table 5.4 [23].

<table>
<thead>
<tr>
<th><strong>Infrastructure</strong></th>
<th><strong>Digital Commons</strong></th>
<th><strong>DSpace</strong></th>
<th><strong>EPrint</strong></th>
<th><strong>Fedora</strong></th>
<th><strong>Islandora</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Hosted Solution</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Locally Installed Software Solution</td>
<td>-</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Customer Support/Community Support</td>
<td>Customer support: Email, phone, resource and community support</td>
<td>Community support</td>
<td>Community support</td>
<td>Community support</td>
<td>Community support</td>
</tr>
<tr>
<td>Flexible Repository Structure</td>
<td>Yes</td>
<td>Limited</td>
<td>Yes</td>
<td>Limited</td>
<td>Yes</td>
</tr>
<tr>
<td>Simple and Qualified Dublin Core Metadata</td>
<td>Yes</td>
<td>Yes</td>
<td>Simple Dublin Core only</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Customized Metadata</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>
For this research the selected repository development software is DSpace. The reason for choosing DSpace is:

- DSpace is fully open source software and it is free to customize and use it.
- DSpace has got a great online community for support
- The researcher has got experience on DSpace customization

**Metadata**

Metadata is structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage an information resource [16]. In short, metadata is often referred to as data about data or information about information. Such descriptive information enables us to have a good idea of the object before actually accessing it. This proves useful in situations where accessing the object is costly. For instance, let the object that we want to access be a video of several megabytes or gigabytes, and after downloading it from the Internet we learn that it is not the video that we wanted. Here, we wasted a considerable amount of bandwidth for downloading the wrong content. But, if we first consulted the metadata of the video (either from its original location or by downloading it to our machine), we could have learnt that it is not what we wanted, with a much lesser wastage of bandwidth [17].

An important reason for creating metadata is to facilitate discovery of relevant information. In addition to resource discovery, metadata can help organize electronic resources, facilitate interoperability and legacy resource integration; it provides digital identification, and support archiving and preservation.
Resource Discovery
Metadata serves the same functions in resource discovery as good cataloging does by:

- Allowing resources to be found by relevant criteria
- Enables easy resources identifying
- Bringing similar resources together
- Distinguishing dissimilar resources and
- Giving location information.

Organizing Electronic Resources
As the number of Web-based resources grows exponentially, aggregate sites or portals are increasingly useful in organizing links to resources based on audience or topic. Such lists can be built as static webpages, with the names and locations of the resources “hardcoded” in the HTML. However, it is more efficient and increasingly more common to build these pages dynamically from metadata stored in databases. Various software tools can be used to automatically extract and reformat the information for web applications.

Interoperability
Describing a resource with metadata allows it to be understood by both humans and machines in ways that promote interoperability. Interoperability is the ability of multiple systems with different hardware and software platforms, data structures, and interfaces to exchange data with minimal loss of content and functionality. Using defined metadata schemes, shared transfer protocols, and crosswalks between schemes, resources across the network can be searched more seamlessly.

Digital Identification
Most metadata schemes include elements such as standard numbers to uniquely identify the work or object to which the metadata refers. The location of a digital object may also be given using a file name, URL (Uniform Resource Locator), or some more persistent identifier.

Archiving and Preservation
Most current metadata efforts center on the discovery of recently created resources. However, there is a growing concern that digital resources will not survive in usable form into the future. 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. Format migration and
perhaps emulation of current hardware and software behavior in future hardware and software platforms are strategies for overcoming these challenges.

Metadata is key to ensuring that resources will survive and continue to be accessible into the future. Archiving and preservation require special elements to track the lineage of a digital object (where it came from and how it has changed over time), to detail its physical characteristics, and to document its behavior in order to emulate it on future technologies.

Furthermore, due to its nature, exact-match querying of audio-visual content is difficult, if not impossible. Hence, the descriptive information in the metadata can prove helpful in querying audio-visual content. For instance, textual descriptions of digital images can be included in the metadata for query/retrieval purposes so that the images can be searched based on the description found in their metadata.

Metadata has been in use in traditional libraries for a long time as a means of indexing books and other documents in the form of catalogs. The information in the catalogs is used to easily search and locate items in the library, which are properly placed in shelves based on some order. In digital libraries as well, metadata is used to facilitate indexing, searching, retrieval, and management of digital documents. The use of metadata can also be extended to museums for description of the various items found in the museum.

**Metadata Standards**

Standards are crucial for interoperability and having a common understanding about a certain phenomenon. This is especially true for multimedia systems where it is impossible to avoid having heterogeneous systems. Metadata is created by and used by various domains and different levels within a domain, including computer science. And, most importantly, in multimedia systems the metadata will be inevitably used by various types of devices and systems. Therefore, to create a common understanding, to achieve interoperability and meaningful communication among professionals and among computing systems, metadata standards are needed. The following are some of the widely used standards. [17]

There are different metadata standards used to build Institutional Repositories that describe, explain, locate, or otherwise make it easier to retrieve, use or manage an information resource. There are some protocols to harvest metadata in XML-formatted metadata from repositories and necessary
requirements for registering an institutional repository as a data provider so that the providers can expose their standardized metadata to be used by other institutional repositories [15].

- **Metadata**, commonly defined as “data about data” or “information about data”, is a structured set of information which describes data (including both digital and non-digital datasets) stored in administrative systems. Metadata may provide a short summary about the content, purpose, quality, location of the data as well as information related to its creation.

- Metadata standards provide data producers with the format and content for properly describing their data, allowing users to evaluate the usefulness of the data in addressing their specific needs. The standards provide a documented, common set of terms and definitions that are presented in a structured format. Standardized metadata support users in effectively and efficiently accessing data by using a common set of terminology and metadata elements that allow for a quick means of data discovery and retrieval from metadata clearinghouses. The metadata based on standards ensure information consistency and quality and avoid that important parts of data knowledge are lost.

Some common metadata standards used to describe information resources of an institution includes the following:

- **Dublin core**: The Dublin Core Metadata Initiative (DCMI) is an open organization, incorporated in Singapore as a public, not-for-profit Company limited by Guarantee (registration number 200823602C), supporting innovation in metadata design and best practices across a broad range of purposes and business models.

- **Agris (International System for Agricultural Science and Technology)** AGRIS is a global public domain Database with 2,830,342 structured bibliographical records on agricultural science and technology. 79.78% of records are citations from scientific journals. The bibliographic references contain either links to the full text of the publication or additional information retrieved from related Internet resources.

- **RDF (Resource Description Framework)** The Resource Description Framework (RDF) is a language for representing information about resources in the World Wide Web. This Primer is designed to provide the reader with the basic knowledge required to effectively use RDF. It introduces the basic concepts of RDF and describes its XML syntax. It describes how to define RDF vocabularies using the RDF Vocabulary Description Language, and gives an overview of some deployed RDF applications. It also describes the content and purpose of other RDF specification documents.
- **RIF (Rule Interchange Format)** The Rule Interchange Format (RIF) is a W3C Recommendation. RIF is part of the infrastructure for the semantic web, along with (principally) SPARQL, RDF and OWL. Although originally envisioned by many as a "rules layer" for the semantic web, in reality the design of RIF is based on the observation that there are many "rules languages" in existence, and what is needed is to exchange rules between them. RIF includes three dialects, a Core dialect which is extended into a Basic Logic Dialect (BLD) and Production Rule Dialect (PRD).

- **RDFa (Resource Description Framework – in – attributes)** is a W3C Recommendation that adds a set of attribute-level extensions to XHTML for embedding rich metadata within Web documents. The RDF data-model mapping enables its use for embedding RDFsubject-predicate-object expressions within XHTML documents; it also enables the extraction of RDF model triples by compliant user agents.

- Metadata harvesters are programs or scripts that capture metadata from other metadata sources. They are slightly different from metadata generators in that they access sets of metadata that are already created. The most well-known example of a metadata harvester is the Open Archives Initiative-Protocol for Metadata Harvesting (OAI-PMH). This script which is available as open source software is designed to identify repositories of OAI-compliant metadata. Using the OAI-PMH organizations can create aggregated repositories of metadata for distributed digital objects. This metadata can be merged with the organization’s own metadata. Aggregations can be built around subjects, library consortia or regional collaborations. The minimum requirement is the exposure of the unqualified Dublin Core Metadata element set. A similar approach is used by OCLC’s Connexion system. Based on the principal of shared cataloguing; this system allows libraries to share the Dublin Core-based metadata records that have been created by other libraries. As with other shared cataloguing activities, the basic metadata structure is well controlled but content standards must be agreed upon and consistently implemented or levels of quality will vary. “
2.3 Challenges

The customization involves different high level task, which are briefly discussed on Customization of DSpace for managing MIR section of this research. Some of the challenges faced during the research of Customization of DSpace for managing MIR are discussed below:

- Difficulty in finding a related research on the area of Customization of DSpace and Multimedia Institutional Repository. There are research works on the area of multimedia and Institutional Repository separately, but it is difficult for the researcher to find a combined research of both areas.

- Since the researcher uses UNECA as a research case study, which exists for the past 50 years and holds a huge set of multimedia resources, it is difficult to select the dataset for the research. The multimedia items are scattered in different division and identification of selection criteria and inventory of the multimedia items are the challenge.

- Most of the collected multimedia items from UNECA are big sized files, especially for the audio and video. Due to the computer space constraint, from the selected multimedia items only some of the items are randomly selected and submitted to the prototype Multimedia Institutional Repository system.

- A constraint of the Institutional Repository software in handling multimedia items. Most Institutional Repository software are designed to handle text based items like publications, books, and articles. For the proper handling of multimedia items, it needs a high level of customization. For these research customizing the core DSpace module to handle the multimedia item took more time and challenging.
2.4 Related Work

Different research works have been done for the area of multimedia and institutional repository separately. One of the researches reviewed is "Multimedia Content's Metadata Management for Pervasive Environment" by Fitsum Meshesha [19], this thesis work addresses the issue of content metadata management in a pervasive environment in relation to content adaptation and it proposes an architecture that consists of components for storage, retrieval, update, and removal of metadata in the system.

Multimedia International Maze and Wheat Improvement Center (CIMMYT)'s Institutional Repository is another MIR work reviewed for this research [17]. The research is done to guard the historical multimedia items in the proper conditions. In order to do this, several actions have been implemented.

The safeguarding of the physical and original images have been implemented through a stabilizing process and digitized in a specific resolution of preservation. A total amount of 1,011 images are collected and individuals on the images are identified in each individual image.

Videos were sorted for digitization by priority based upon their physical status. 16mm and 8mm are being digitized first due to their historical importance [17].

Besides the similarity of working on the area of MIR, there are differences between this research and CIMMYT’s research. CIMMYT mainly conducted on identifying a stabilizing process and digitizing the multimedia items to safeguarding the physical and original multimedia items.

This research is also on proper reservation and easy access of multimedia on institutional repository software. The main focus is by identifying a multimedia metadata standard and implement to DSpace to represent each multimedia items. Customization of DSpace to create a proper displaying mechanism for multimedia item based on the item type.
CHAPTER THREE

Methodology

The study follows experimental research for customization of DSpace for managing the proposed web-based Multimedia Institutional Repository.

Experimental research is a collection of research designs which use manipulation and controlled testing to understand causal processes. Generally, one or more datasets are manipulated for images, audio and video. Then based on the formulated methodology customized a prototype MIR to demonstrate and evaluate the system. The research follows a research method for customization of DSpace for the proposed computer based Institutional Repository web application. The research involves three basic steps; first there is a need to prepare a multimedia datasets of images, audios and videos. This is followed by the customization of Institutional Repository software called DSpace. Finally, the usability of the prototype has been demonstrated and evaluated to determine its performance.

Web based applications enable users to access information from anywhere and using a variety of devices. The devices vary from small portable devices like mobile phones and tablets to powerful personal computers. The communication system also varies from slower dial up internet connection and intermittent wireless links to high speed 4G wired/wireless links. As a result, the storage, processing, displaying and communication capabilities of these devices also vary greatly. Hence, content, especially multimedia content, cannot be equally accessed by all of these devices unless it is customized appropriately to suit each device. Furthermore web browsers open the door to the display of various kinds of multimedia objects and now able to display many different kinds of information including text, image, video, audio and even virtual reality. Web browsers are able to support different multimedia formats through the use of plug-ins and add-ons that tell the browser how to display the different kinds of standard files that may be sent from web servers.
4.1 Literature Review

Related literatures from different sources (books, the internet, journals, conference proceedings, and articles) are reviewed to understand Multimedia Institutional Repository (MIR) issues. The major activities done include:

- Studying major researches in the area of Institutional Repository
- Investigating the application of digitizing multimedia items.
- Investigating the metadata standards and defining a metadata standard for the multimedia items
- Investigating the application of digitizing multimedia items, and understanding the development tools and customization.

4.2 Dataset Preparation

UNECA hosts different regional and international events and each of these activities generates huge amount of multimedia data (image, audio and video). For instance if we take one year event of African Development Forum (ADF), it has got around 310 images, 89 audios and 5 video items. One of the main tasks for the customization of DSpace for managing MIR is dataset preparation. Datasets are the major criteria for evaluating the customized prototype of the research output. To achieve the general objective of this study, that is Customization of DSpace for managing Multimedia Institutional Repository, the dataset preparation task is divided into the following phases:

- Understanding UNECA multimedia item inventory.
- Identification of multimedia item selection criteria.
- Identification of items for each collection. Identification is done by purposive sampling based on the quality of multimedia items, multimedia items with a higher quality selected first because it doesn’t need any other conversion/technology to use on MIR.
- Pre-processing on the multimedia items. Preprocessing in this stage is converting to digital format, multimedia item compression, and resizing the item.

4.2.1.1 Inventory

The main objective of the inventory phase involves gathering all multimedia outputs of UNECA and organizes them according to their type. In this process register all the
collected multimedia resources to be used in the next phases. During this task, all the items are identified based on their media type.

To successfully collect all multimedia products, we have to consider the main factors or key areas which can highly influenced the inventory phase. One of the major factors is availability of multimedia products. Most of these products are scattered in different divisions such as on personal computers and stores. Therefore, identifying the whereabouts of these products is one of the key factor. The second key factor is getting permission to have the product.

The first step to collect the items are identifying where the item with the multimedia content are found. These items include images, audio, video, speeches and press releases which are intellectual outputs of UNECA. For this study the collection starts from External Communication and Media Relation Section because most of multimedia items exists on that division and then will continue to other section.

The inventory of the products is according to products type; then each product is organized by UNECA main thematic areas. UNECA thematic areas are: Macroeconomic Policy, Social Development, Regional Integration & Trade, Natural Resource, Innovation and Technology, Gender and Governance.

4.2.1.2 Reviewing and selection

Reviewing and selection phase involves identification of products which are going to be relevant for repository purposes. The contents of this product must be clear, meaningful and must contain important information about the events.

An attempt made to identify which items are available in the repository. One of the challenge encountered is that all collected multimedia products might not be available in the institutional repository due to some reasons. The first reason is storage space, the storage space required for uploading multimedia products is not similar to the space required for document if we upload all of the products, and the performance of the system might be decreases. The second factor is ability of identifying the content of each product (who, where, when, what) and the last factor
is availability of technology to review the multimedia products (what technology does we have and what we don’t have).

During reviewing and selection of multimedia items, we follow the following steps:

- Removal of duplication
- Give priorities for each product by product content, quality
- Selection (Organizing) products based on year then by event and thematic areas
- Review available technology in UNECA for the conversion process
- Identify the way to find unavailable technology

### 4.2.1.3 Digitization

Digitalization process involves conversion of printed image, documents and analog videos and audios in to digital format. The main aim of digitization is a long-term preservation of multimedia resources to ensure that the digital materials to be available and accessible in the long term and to be available to the customized multimedia repository software. Text and images can be digitized similarly: a scanner captures an image (which may be an image of text) and converts it to an image file. Audio and video digitization uses one of many analog-to-digital conversion processes in which a continuously variable (analog) signal is changed. In this phase each converted file will have standard file format which is easy for the web services (cyber services) publication by using interactive multimedia software for instance, digital formats like mp3, TXT, DOC, PDF, JPEG, MPEG and TIFF and etc.

### 4.2.1.4 Indexing and Metadata selection criteria

Metadata used to properly describe the multimedia items. For multimedia metadata representation a Qualified Dublin Core is used.

### 4.3 Implementation Tools for Customization of DSpace for managing MIR

For this research an Open-source software packages highly used for building web-based Multimedia Institutional Repository (MIR). The technologies implemented for the sample tools are also described.
DSpace is one of the open source software package that provides the tools for management of digital assets, and is also commonly used as the basis for an institutional repository. It supports a wide variety of data, including books, theses, and 3D digital scans of objects, photographs, film, video, research data sets, and other forms of content. The data is arranged as community collections of items, which bundle bit streams together.

There are different technologies used on DSpace and implemented environment for this research:

- Programming language used for customization and development of MIR is Java and JSP, using the Java Servlet API
- Database used is PostgreSQL. A database name dspace is created with the role name dspace with full privilege.
- To run MIR the web Server is Apache Tomcat 7.
- Search Engine within MIR is Elastic Search
- Standard specification: OAI-PMH v2.0 and is capable of exporting METS (Metadata Encoding and Transmission Standard) packages.
- Prototype is deployed on Windows 10 operating system for experimentation and performance evaluation.

The prototype of MIR is mainly customization of the core DSpace package for multimedia item handling module for images, audio and video. There is different technological option for the proper handling of the multimedia items. For this research different web-based application development technologies are used and some of them are listed below:

- Extensible Markup Language (XML) is used to implement the identified metadata standard to MIR.
- Hyper Text Markup Language (HTML 5) used for front end design of MIR
- Bootstrap JavaScript Library is used for easy accessible and performance
- Cascading Style Sheet (CSS 3) for the responsiveness and overall layout of MIR
4.4 Evaluation Procedure

There are some general criteria that can be used to evaluate. The following are the high level evaluation criteria:

- Application functionality testing
  - To check the customized system handles the multimedia items properly.
  - Displaying the multimedia items on the web based interface and easily accessible on the well-known internet browsers. Testing based on application functionality regarding making available multimedia materials and performance testing based on the response time
  - To check the metadata standard proposed properly represent the multimedia items.
  - Making sure the multimedia items are easily searchable with the metadata field by indexing all resources of the repository and using elastic searching functionality of DSpace
  - To check the system has a mobile device friendly user interface on different mobile apparatus and evaluate the rendering of the repository contents.

- Application performance testing
  - Based on user acceptance testing questionnaire.

Multimedia resources are the main information sources included in Institutional Repository platform and make accessible to the potential users. Organizing multimedia resources add a value for a sustainable and successful dissemination of information to achieve organizational or institutional objectives. In summary, a multimedia institutional repository has an effort to bring a long-term preservation of ECA multimedia digital resources which can encounter with the historical heritages, documentary films, current events and valuable information of the institution.
CHAPTER FOUR

Metadata for Multimedia Management

Metadata content are providing rich interlinked searchable metadata about that data. Richer metadata contains detailed and meaningful names, subject keywords and full descriptions are enables the researcher to get to the data. It also has the capability to support more sophisticated ways of managing and distributing information that are often given such labels as knowledge management and information resource management.

In the Metadata description below, each element has a descriptive label for human consumption and a unique token (“name”) for use in machine processing.

In accordance with the DCMI (Dublin core metadata initiative) Namespace Policy, the “name” of an element is appended to a DCMI namespace URI to construct a Uniform Resource Identifier as a globally unique identifier for that element. Table 4.1 below presents the structure of metadata identified for customizing Multimedia Institutional Repository

The identified metadata standard is the extended version of the Dublin Core metadata initiative.
<table>
<thead>
<tr>
<th>No</th>
<th>Element</th>
<th>Metadata Standard</th>
<th>Definition</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Title</td>
<td>dc. Title</td>
<td>The name given to the resource.</td>
<td>Typically, a Title will be a name by which the resource is formally known.</td>
</tr>
<tr>
<td>2</td>
<td>Creator</td>
<td>dc. Creator</td>
<td>An entity primarily Responsible for making the content of the resource.</td>
<td>Examples include a person, an organization, or a Service. The name of the Creator should be used to indicate the entity.</td>
</tr>
<tr>
<td>3</td>
<td>Subject</td>
<td>dc. Subject</td>
<td>The topic of the content of the resource.</td>
<td>A Subject will be expressed as keywords or key phrases or classification codes that describe the topic of the Resource. Recommended best practice is to select a value from a controlled vocabulary or formal Classification scheme.</td>
</tr>
</tbody>
</table>
| 4  | Description | dc. Description  | Extent, Note  
Definition: *Extent* is the total number of sequences associated with the multimedia object. *Note* is a textual statement about the multimedia sequence. | Description may include but not limited to; an abstract, a table of contents, a graphical representation, or a free-text account of the resource. |
<p>| 5  | Publisher | dc. Publisher     | The entity responsible for making the resource available. | The name of a publisher should be used to indicate the entity. Examples of a publisher include a person, an organization, or a service. Typically, |</p>
<table>
<thead>
<tr>
<th>Column</th>
<th>DC Element</th>
<th>Description</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>Date</td>
<td>Date will be associated with the creation or availability of the resource. Recommended best practice for encoding the date value is defined in a profile of ISO 8601 [Date and Time Formats, W3C Note, <a href="http://www.w3.org/TR/NOTE-datetime">http://www.w3.org/TR/NOTE-datetime</a>] and follows the YYYY-MM-DD format.</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Type</td>
<td>Recommended best practice is to use a controlled vocabulary. To describe the file format, physical medium, or dimension of the resources, use the format element.</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Format</td>
<td>Use the attributes minimum and maximum to record the minimum and maximum numerical value for the BitRate in cases of variable bit rate; For example 64, 128, 256, etc.</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>Identifier</td>
<td>Recommended best practice is to identify the resource by means of a string or number conforming to a formal identification system. Examples of formal identification systems include the Uniform Resource Identifier (URI) (including the Uniform Resource Locator (URL), the Digital Object Identifier (DOI))</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Source</td>
<td>The statement may include a description of any changes successive custodians made to the resource.</td>
<td></td>
</tr>
</tbody>
</table>
|   |   | significant for its authenticity, integrity and interpretation. | Use the three-letter ISO 639-3 code for the representation of the name of the language(s).
In cases where more than one language is spoken in video, use a separate language field for each language. |
|   | Language | dc. Language | A language of the resource. |
|   |   | dc.language.iso |   |
|   | Coverage | dc. Coverage | The spatial or temporal topic of the resource, the special applicability of the resource, or the jurisdiction under which the resource is relevant. |
|   |   | dc.coverage.spatial |   |
|   | Rights | dc. Right | Information about rights held in and over the resource. |
|   |   | dc.right.uri |   |
|   |   | dc.right.issued |   |
|   |   | dc.right.general |   |

Table 4.1 - Identified multimedia metadata standard for UNECA multimedia resources.
CHAPTER FIVE

Customization of Multimedia Institutional Repository

5.1 Data Preparation

5.1.1 Multimedia item inventory

In this phase it is tried to collect different multimedia items from different section in UNECA. This item includes images, audio, and videos. The images include meetings in Africa Hall, UNECA building pictures, UNECA anniversary pictures, UN head quarter pictures and Project pictures. It is categorized by year and events and UNECA thematic areas.

In addition, the video resources include Betacam, VHS, Filmstrip, U-matic, MiniDV, DV, DVcam, CD and Floppy. Further the collections of Audio resources are Real tape, Cassette, CD and Minidisc. All the multimedia resources collected from different offices. Most of the resources are collect from the ECAMRS studios’. All video and audio resources are categorized and identified by unique ID, classified based on their types.

The video and audio resources start from 1969 and the images are from 1945 up to now. For instance the video resources contain the filmstrip on the title Committee of 24 in Africa – May 1969 and Human right Concert in 1974. In addition to this on audio format the interviewee Dr. Robert K.A Gardiner, retiring Executive Secretary of the Economic Commission for Africa (ECA) in 1975 and ECA meeting on 26 February in 1975 have a real tape records.

Based on those tasks the inventory, reviewing and processing 173,635 printed and digital photos, 1,350 Image slides, 64,359 digital files and 4,081 analogues pieces of videos and also 892 digital files and 1,099 analogues pieces of audios documents.

Table 5.1 summarizes the details of the total multimedia inventory collected by ECAMRS.
Digital multimedia

<table>
<thead>
<tr>
<th>Type</th>
<th>Status</th>
<th>Total Quantity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Printed</td>
<td>Digital</td>
</tr>
<tr>
<td></td>
<td>Quantity</td>
<td>size</td>
</tr>
<tr>
<td>Images</td>
<td>8,409</td>
<td>165,226</td>
</tr>
<tr>
<td></td>
<td>645.44GB</td>
<td></td>
</tr>
<tr>
<td>Image slide</td>
<td>1,350</td>
<td></td>
</tr>
<tr>
<td>Video</td>
<td>64,359</td>
<td>745.5GB</td>
</tr>
<tr>
<td>Audio</td>
<td>892</td>
<td>29.02GB</td>
</tr>
</tbody>
</table>

Table 5.1: Inventory of multimedia items

Non Digital multimedia items that needs technology

<table>
<thead>
<tr>
<th>Video Items</th>
</tr>
</thead>
<tbody>
<tr>
<td>No</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>5</td>
</tr>
<tr>
<td>6</td>
</tr>
<tr>
<td>7</td>
</tr>
<tr>
<td>Total</td>
</tr>
</tbody>
</table>

Table 5.2 - Non digital items that needs technology to convert
### Audio Items

<table>
<thead>
<tr>
<th>No</th>
<th>Title</th>
<th>Number</th>
<th>Remark</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Real Tape</td>
<td>589</td>
<td>Need technology</td>
</tr>
<tr>
<td>2</td>
<td>Cassette</td>
<td>448</td>
<td>Need technology</td>
</tr>
<tr>
<td>3</td>
<td>Minidisc</td>
<td>16</td>
<td>Need technology</td>
</tr>
<tr>
<td></td>
<td><strong>Total</strong></td>
<td></td>
<td><strong>1,053</strong></td>
</tr>
</tbody>
</table>

Table 5.3 - Non digital audio items that needs technology to convert

#### 5.1.2 Selection criteria

Reviewing and selection phase involves identification of items which are going to be relevant for the repository purposes. The contents of this item must be clear, meaningful and must contain important information about the events.

Review and selection phase has the following steps:

- Removal of duplication
- Give priorities for each item by its content (who, where, when, what)
- Rank the items based on their quality
- Selection (Organizing) items based on year then by event and UNECA thematic area
- Review available technology (see annex 1, 2,3 ) in UNECA for the conversion process
- Identify the way to find unavailable technology

Uploading all multimedia resources in to institutional repository is a time taking activity and it requires a huge space. Prior to these, all multimedia items may not have equal quality and may not describe the content or the subject clearly. In addition to these items might give us information from different angle.
5.2 Customization and evaluation of Multimedia Institutional Repository

The ultimate goal of this research is creating a digital space for UNECA multimedia materials that can be accessed easily for the information seeker. To this end web based Multimedia Institutional Repository software is customized and system can be accessed in both networked environment and World Wide Web.

The customization of DSpace for managing MIR starts by doing a research on different Institutional Repository software. There are different kinds of tools available with its own features. For this research the prototype system is customized on the single laptop. In production environment implementation it can have a dedicated server for both the web application and database. The development language used for the customization and handling of multimedia items on this research is XML and, the Institutional Repository software used for the customization is DSpace.

The XML (eXtensible Markup Language) is a meta-markup language developed by the World Wide Web consortium (W3C) to deal with a number of the shortcomings of Hypertext Markup Language (HTML) for web based document sharing. The more functionality was added to HTML to account for the diverse needs of users of the Web, the language began to grow increasingly complex. The semantics and syntax of tags are fixed in HTML, but in XML the author of the document is free to create tags whose syntax and semantics are specific to the target application. Originally the purpose of XML was to markup content, but it became clear that XML also enables to describe structured data thus making it necessary as a data storage and interchange format [21]. XML provides a simple, universal and widely accepted standard for representing, publishing and exchanging data on the web. Traditionally, data is stored and managed in relational databases and is transformed into XML whenever necessary. However, there is increasing number of applications using native XML documents because the requirements of modern information systems are often too complex to be implemented by relational tables [20]. If XML is used as a first class data model, then it requires effective means for the management of persistent XML documents as databases. In particular, the management of XML data has to cope with issues like consistency, reliability, recovery, integration, efficient access and distribution.
5.2.1 DSpace Installation

DSpace installation is a combination of installing some prerequisite software and the configuration of the DSpace for the specific use. Here below describes the third-party components and tools to run a DSpace server. Since DSpace is built on open source, standards-based tools, there are numerous other possibilities and setups.

5.2.1.1 Oracle Java JDK 7 or OpenJDK 7

The web server Tomcat uses Java 1.6 to compile JSPs by default and DSpace requires the full JDK (Java Development Kit) be installed, rather than just the JRE (Java Runtime Environment).

OpenJDK and Oracle's Java with its installation instructions can be freely downloadable and operating systems also provide an easy path to install OpenJDK or Oracle’s Java.

5.2.1.2 Apache Maven 3.0.5+ (Java build tool)

Maven is necessary in the first stage of the build process to assemble the installation package for the DSpace instance. It gives the flexibility to customize DSpace using the existing Maven projects found in the \[dspace-source/dspace/module\] directory or by adding in its own Maven project to build the installation for DSpace, and apply any custom interface "overlay" changes. Maven is a free downloadable package.

5.2.1.3 Apache Ant 1.8 or later (Java build tool) Ant

Apache Ant is required for the second stage of the build process. It is used once the installation package has been constructed in and still uses some of the familiar ant build \[dspace-source/dspace/target/dspace-installer\] targets found in the 1.4.x build process. Ant also a freely downloadable package.

5.2.1.4 Relational Database: (PostgreSQL or Oracle) Database

PostgreSQL or Oracle is required as a DBMS to install DSpace. For this research we used PostgreSQL. PostgreSQL a freely downloadable DBMS. Unicode (specifically UTF-8) support must be enabled (but this is enabled by default). Once installed, we need to enable TCP/IP connections (DSpace uses JDBC).
5.2.1.5 Servlet Engine (Apache Tomcat 7 or later, Jetty, Caucho Resin or equivalent)

Apache Tomcat is an open-source web server and servlet container developed by the Apache Software Foundation (ASF). Tomcat implements several Java EE specifications including Java Servlet, JavaServer Pages (JSP), Java EL, and WebSocket, and provides a "pure Java" HTTP web server environment for Java code to run in.

5.2.2 User Interface Design

A good interface design determines the popularity of the system and therefore becomes a competitive advantage in the field that it is designed for. For MIR, one of the key requirements is to build a system that is efficient and is easy to use; therefore the interface design serves as a vital ingredient in the success of this study.

Design rules in interface design are in the form of standards and guidelines that provide direction for design, regarding issues of presentation, behavior and interaction for interface elements and controls although the use of standards and guidelines supports decisions in design choices but these need to be adjusted according to the needs of the interface. Thus, all design choices, while designing the interface for MIR is made according to the standard and the task for which the system made for.

Further to the interface design, it is also essential to consider the usability of the system, which is considered a key factor in the design of any software application. According to ISO 9241, usability is defined as “the effectiveness, efficiency and satisfaction with which specified users achieve specified goals in particular environment” [19]. In simple words, it denotes how well the users can make use of the functionalities of the system. One of the most accepted measures of such usability comes in the form of Nielsen’s Usability Heuristics [19], which are given below:

- Simple and natural dialogue
- Speak the user’s language
- Minimize the user’s memory load
- Consistency
- Feedback
- Help and documentation
- Clearly marked exits
- Prevent errors
• Descriptive error messages

For this research the user interface designing uses a mobile first theming package for the ease access of the system. One of technology used are for user interface design described below:

**Mirage 2**

Mirage is the default XML User Interface in DSpace. It was originally developed by Art Lowel and contributed to DSpace 1.7. Mirage 2 has been created according to responsive user interface design principles. As part of this process, the theme has been tested on devices with varying screen dimensions. Instead of forcing every single link or button you would see on a large screen onto smaller screen size displays, the creation of a better user experience was the main focus. The theme brings technologies like:

• Bootstrap mobile first front-end framework for faster and easier web development for making MIR responsive to different devices (desktop, tablets, mobile devices)
• Grunt the JavaScript Task Runner, this is a feature integrated on Mirage 2 theme for automation, performing repetitive tasks like minification, compilation of MIR.
• Compass CSS Authoring framework, Sass: Syntactically Awesome Style Sheets CSS extension language, HTML 5 and CSS 3 are for the overall styling and look and feel of MIR.
Main Areas Of Works

- Macroeconomic Policy
- Social Development
- Regional Integration and Trade
- Natural Resource Management
- Innovation and Technology
- Gender
- Governance
Dr. Carlos Lopes remarks on LAAP prize

Addis Ababa, 10 October, 2014 (ECA) - Economic Commission for Africa's Executive Secretary, Carlos Lopes has been awarded the prestigious Lifetime Africa Achievement Prize (LAAP) by the Millennium Excellence Foundation for his contributions to overturn Africa's economic fortunes, in the category Action for Africa. The award ceremony took place in Lagos, Nigeria on October 10.

Citation

URI
http://hdl.handle.net/123456789/77

Collections
Lifetime Africa Achievement Prize (LAAP) 2014

Subject

Date
2014-10-10

Share
Facebook
Twitter
LinkedIn

Figure 5.1 User interface of MIR prototype
5.2.3 Customization

The core DSpace module has the metadata standards and accepts the multimedia items to the repository. One of the main drawbacks is the metadata standards by default incorporated on DSpace is specially designed to electronic materials like documents, for multimedia items the metadata needs to be customized to meet the user need. The other issues is DSpace creates a link for every uploaded materials, but for multimedia items besides the downloading option it needs to display/play the multimedia items based on item type without downloading it.

The main customization tasks on this research categorized into five parts:

5.2.3.1 Multimedia metadata implementation

As discussed on Chapter Four: Multimedia Metadata Management, a metadata standard is identified to handle the multimedia items in UNECA.

In order to apply the new multimedia metadata standard, we need to customize the input-forms.xml file under the configuration section of the installed DSpace.

```
<field>
    <dc-schema>dc</dc-schema>
    <dc-element>date</dc-element>
    <dc-qualifier>issued</dc-qualifier>
    <repeatable>false</repeatable>
    <label>Date of Issue</label>
    <input-type>date</input-type>
    <hint>Please give the date of previous publication or public distribution below. You can leave out the day and/or month if they aren't applicable.</hint>
    <required>You must enter at least the year.</required>
</field>
```

Figure 5.2 - A metadata code for Issue Date Field

Customized Multimedia Metadata used in this research is discussed in detail in Annex 4.

5.2.3.2 A customization/implementation to handle multimedia items

DSpace allow us to submit multimedia items on the repository, just like we upload pdf or word documents; we can upload multimedia items (images, audio and video) on the repository.

The default item accessing method on the DSpace is downloading the item; it automatically generates a link to download the item and when the user clicks it starts downloading. The default item accessing method seems okay for the non-multimedia items (documents, files, etc...), but for the big sized multimedia items the downloading mechanism is not a good solution.
What is done on this research is instead of downloading the item, based on the multimedia item file type we developed a multimedia item viewing mechanism. If the item is image, it displays the image. If the item is audio or video, it displays audio or video player so that the user can listen or view it without downloading.

To develop the multimedia item viewing, it needs to customize the default DSpace item viewing source code. The code first needs to evaluate the item based on its file type so that it generates the viewing mechanism as described below in figure 5.1.

```xml
<xsl:when test="contains(mets:FLocat[@LOCTYPE='URL']/@xlink:href, 'jpg')">
  <div>
    <a class="image-link">
      <xsl:attribute name="href">
        <xsl:value-of select="mets:FLocat[@LOCTYPE='URL']/@xlink:href"/>
      </xsl:attribute>
      <img style="border-radius:10px; width:100%; height:130px;" class="img-zoom">
      <xsl:attribute name="src">
        <xsl:value-of select="mets:FLocat[@LOCTYPE='URL']/@xlink:href"/>
      </xsl:attribute>
    </a>
  </div>
</xsl:when>
```

Figure 5.3 - A sample source code to handle .jpg image file type.
Implementation of community, sub community, collection and items for MIR

The design and structure of MIR is designed based on the UNECA thematic areas. Multimedia items are also organized based on this. Here we used is the thematic areas presented in figure 5.1 as a community of the repository.

![Image of thematic areas]

Figure 5.4: Community of MIR (Macroeconomic Policy, Social Development, Regional Integration & Trade, Natural Resource, Innovation and Technology, Gender and Governance).

For all thematic area community, the Sub-Communities are **Image, Audio** and **Video**. Collections are categorized into Event (conference, Meeting, Forum, Summit, Press briefing, press, and speech, Exhibition, Farewell, Ceremony), Documentary, Interview and Training (workshop, Seminar)

### 5.2.4 Testing

#### 5.2.4.1 Evaluation criteria

The Multimedia Institutional Repository web application serves as an ideal solution for representing UNECA multimedia resources (images, audio and video) easily accessible form; beside that the system must also satisfy the following:

- **Naturalness of multimedia items**: multimedia items needs to be treated based on its item type. The way we represent multimedia items should align with the nature of the item. Instead of treating multimedia items as pdf or a file documents and creates a link for downloading option, it better to display the item without losing the naturalness of the multimedia.

- **Usability**: Usability is an important consideration for web applications. Therefore, the MIR should be easily navigable. Also, the interface of MIR will be designed while keeping in mind Nielsen’s Usability Heuristics [18].
5.2.4.2 User Acceptance Testing

A user acceptance testing questionnaire is distributed to 12 UNECA staff member user to perform a testing on the prototype MIR application (Please refer Annex 5 to see the user acceptance testing questionnaire distributed), based on their response to the question all of the users strongly agreed that MIR is beneficial to them and the prototype application is user-friendly so that they can easily navigate on the application.

<table>
<thead>
<tr>
<th>Question</th>
<th>Strongly Disagree</th>
<th>Disagree</th>
<th>Neutral</th>
<th>Agree</th>
<th>Strongly Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1  Do you agree that using MIR would definitely beneficial to you?</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>11</td>
</tr>
<tr>
<td>2  Do you agree MIR prototype application user-friendly?</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>10</td>
</tr>
<tr>
<td>3  Does multimedia items easily accessible on MIR?</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>10</td>
<td>2</td>
</tr>
<tr>
<td>4  Does each multimedia items properly described with the metadata standard?</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>9</td>
</tr>
<tr>
<td>5  How do you see the performance of the prototype?</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>6  Does prototype of MIR mobile device ready application?</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>10</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 5.4: User acceptance testing questionnaire analysis
Based on the responses of the user the average sentiment among the respondents agreed that using MIR would definitely beneficial to the user and prototype application is user-friendly.

From the conducted user acceptances testing the users also agreed on the customized prototype application multimedia items are easily accessible and the multimedia items are properly described by the metadata standard.

With regard to the performance of MIR the users agreed MIR has a good performance, but the performance is slow with audio/video items.

Overall, all the testing conducted to test the functionality of the application led to a single conclusion that MIR is compliant with the requirements detailed in the identifying the methodology stage. It was also felt that MIR would perform substantially better as compared to its performance during the testing process mainly because of the following reason:

- The machine on which MIR was hosted during testing cannot achieve the same performance level as a high-end web server because the server would have a faster processor and a lot more memory. Furthermore, during performance testing the load testing application was also being run on the same machine as MIR, which means an increased burden on the machine resources as compared to normal circumstances.
- For a production repository application development, the database and the webservice are separated, but for this research both the database and application hosted on the same machine. If the database and the web hosted in two different locations, it will minimize the load on the server and increase the performance.
CHAPTER SIX

Conclusion and Recommendation

6.1 Conclusion

In this study we have tried to address the issue of Multimedia Institutional Repository for the case of UNECA by customizing DSpace for managing a web based MIR application that facilitates the management of multimedia items (image, audio and video). Many valuable lessons were learnt during the course of this research on both technical and human level. Perhaps the most important of them being related to the extent of planning required for an experimental based research.

The customization of DSpace for managing MIR is the key step of this research, first we compare different Institutional Repository software and select open source software called DSpace, and then we developed a customized code to handle multimedia items (images, audio and video) on the top of the core DSpace. We implemented the identified Multimedia Metadata on the customized application and design a mobile ready user interface for the application.

Even though most of the issues regarding the customization of DSpace for managing Multimedia Institutional Repository were addressed in this study, there are areas which were not covered on the research. One of these areas involves the outcome of the MIR with regard to the increase of multimedia items on the performance of the system; we implement the customization of web application and the database on standalone computer and the performance not tested with a large amount of multimedia items.

The strong side of this research and the major contributions are given as follows:

- Research is conducted on the area of Multimedia (Digital Image, Digital Audio and Digital Video), and Institutional Repository.
- A basic Multimedia Metadata Standard for UNECA is identified for the proper representation of multimedia items.
- Basic steps for the customization of DSpace for managing MIR are identified on this study.
- Inventory of multimedia items in UNECA and identification of selection criteria are proposed and implemented on the prototype.
- Customization of DSpace for managing Multimedia Institutional Repository.
  - DSpace installation
  - Implementation of identified Multimedia Metadata standard the deployed DSpace
- Development of multimedia handling module; this is done based on the multimedia item file format types; it displays a proper mechanism to view the item.

- Develop a mobile device ready system by integrating HTML 5 and CSS 3 page responsive feature and now MIR is easily accessible on mobile devices.

6.2 Recommendation

This research is an open architecture for integrating additional features and enhancements for future development. In light with this, the following recommendations are made for further research and improvements:

- Adding a bilingual feature to handle multiple language (for UNECA at least English and French)

- Customize the core DSpace to handle flexible repository structure

- Adding Google Analytical integration for performance analysis, page hit report, and overall reports of the repository.

- Development of Album Viewer for images with JQuery or Bootstrap popup viewing features
References


[10]. Derek Alexander Muller. (n.d). Designing Effective Multimedia for Physics Education. School of Physics University of Sydney Australia.


[13]. Melanie Cofield. (2005). Digital Imaging Basics. School of Information, the University of Texas at Austin


# Annex 1: Technology requirement and availability for video items

<table>
<thead>
<tr>
<th>No</th>
<th>Physical media</th>
<th>Technology</th>
<th>Technology available in ECA</th>
<th>Remark</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Beta cam</td>
<td>Beta cam</td>
<td>Video Cassette Recorder UVW-1800p (status: it works but it needs maintenance and will be measured the capacity of how may resources will converted)</td>
<td>If Beta cam Player Machine standard used to play medium and large size cassette, that Machine may also used to play the VHS resources.</td>
</tr>
<tr>
<td>2</td>
<td>VHS</td>
<td>VHS Player Machine</td>
<td>VHS Player Machine Videocassette recorder SVHS-Hi-Fi SR-S388E (status: It’s not connected with other accessories and it needs maintenance) If it gets a preventive maintenance, it used as for previewing purpose.)</td>
<td>If Beta cam Player Machine standard used to play medium and large size cassette, that Machine may also used to play the VHS resources.</td>
</tr>
<tr>
<td>3</td>
<td>Film Strip</td>
<td>Telecine</td>
<td>Not available</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>U-matic</td>
<td>U-matic</td>
<td>Not available</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>MiniDV</td>
<td>MiniDv</td>
<td>Not available</td>
<td>If DVcam Player Machine standard used to play medium and large size cassette, that Machine may also used to play the DV and Minidv resources.</td>
</tr>
<tr>
<td>6</td>
<td>DV</td>
<td>DV Player Machine</td>
<td>DV Player Digital video cassette recorder AG DV 2700 (status: -</td>
<td>If DVCam Player Machine standard used to play medium and large size cassette, that Machine may also use</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>DVCam</td>
<td>DVCam Player Machine</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Digital Video Cassette DSR-1800p (status:- it works but it needs maintenance and will be measured the capacity of how may resources will converted )</td>
<td>If DVCam Player Machine standard used to play medium and large size cassette, that Machine may also use to play the DV and Minidv resources.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>CD</td>
<td>Computer</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Can be handled on PC</td>
<td>To play the DV and Minidv resources.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

It’s not connected with other accessories and it needs maintenance) If it gets a preventive maintenance, it used as for previewing purpose.)
### Annex 2: Technology requirement and availability for audio items

<table>
<thead>
<tr>
<th>No</th>
<th>Physical media</th>
<th>Technology</th>
<th>Technology available in ECA</th>
<th>Remark</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Real tape</td>
<td>Real tape player Machine</td>
<td>Real tape player Tascam BR-20 (status:- not assembled with others accessories ) It need to be measured the capacity of how may resources will converted</td>
<td>If we have a capacity to convert those documents within outsources institution, it’s better to use for previewing purposed for the archives.</td>
</tr>
<tr>
<td>2</td>
<td>Cassette</td>
<td>Cassette player</td>
<td>1-Cassette player Tascam CD-A700 (status:- it works but it needs preventive maintenance) 2-Tascam DA-P1 digital Audio Tape Recorder (status: not assembled with others accessories) It used to listen cassette resources.</td>
<td>If we have a capacity to convert those documents within outsources institution, it’s better to use for previewing purposed for the archives.</td>
</tr>
<tr>
<td>3</td>
<td>CD</td>
<td>Computer</td>
<td>Can be handled on PC</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Minidisk</td>
<td>Minidisk Player Machine</td>
<td>Not available</td>
<td></td>
</tr>
</tbody>
</table>
## Annex 3: Technology requirement and availability for image items

<table>
<thead>
<tr>
<th>No</th>
<th>Physical media</th>
<th>Technology</th>
<th>Technology available in UNECA</th>
<th>Remark</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Image slide and Negative photo films</td>
<td>Slide image reader and converter and also Negative photo films reader and converter</td>
<td>Not available</td>
<td>Meet with an international volunteer institution</td>
</tr>
<tr>
<td>2</td>
<td>Printed Photos</td>
<td>Scanner</td>
<td>Available</td>
<td></td>
</tr>
</tbody>
</table>
Annex 4: User Acceptance Testing Questionnaire

Name: Customization of DSpace for Managing Multimedia Institutional Repository.

Description: To perform User Acceptance Testing on the developed prototype MIR application.

Purpose: To Elicit information on the performance of MIR from users’ points of view.

Questionnaire:

1. Do you agree that using MIR would definitely beneficial to you?
   1 2 3 4 5

2. Do you agree MIR prototype application user-friendly?
   1 2 3 4 5

3. Does multimedia items easily accessible on MIR?
   1 2 3 4 5

4. Does each multimedia items properly described with the metadata standard?
   1 2 3 4 5

5. How do you see the performance of the prototype? Does it load fast?
   1 2 3 4 5

6. Does prototype of MIR mobile device ready application?
   1 2 3 4 5

Completed by: ____________________________

Date: ____________________________

Thank You!