A Lightweight Digital Library Framework

A Dissertation
Submitted to the Faculty of Science
of University of Cape Town
In Partial Fulfillment of
the Requirements for the Degree of
Master of Science
in
Computer Science

by
Lighton Phiri

Supervised by
Professor Hussein Suleman

January, 2011

Copyright © 2012 Lighton Phiri.
All rights reserved.

Abstract

Abstract iyambila pamene apa.

Acknowledgements

Kaya kapena niza fillinga liti iyi space –mwina pakapita myezi akhumi ndi awili?

Table of Contents

List of Figures

List of Tables

Chapter 1 Introduction

1.1 Problem Statement

Vuti ilikuti...

1.2 Hypotheses

Hypotheses kuyachita list apa...

1.3 Research Questions

Ma research questions yonse apa...

1.4 Motivation

Motivation ifunika apa...

1.5 Approach

Approach nayeve pamene apa...

1.6 Outline of Dissertation

Outline iyambila apa...

Chapter 2 Background

Research in the field of Digital Libraries has been going on for over two decades. The mid 1990s in particular saw the emergence of a number of government funded projects [14], conferences [1], technical committees [10, 24] and workshops [8, 22] specifically set up to foster formal research in the field of Digital Libraries. The rapid technological advances and more specifically Web technologies have resulted in a number of different Digital Library System frameworks, conceptual models, architectural designs and Digital Library software tools. The variation in the designs can largely be attributed to the different design goals and corresponding specific problems that the solutions were aimed to address.

This chapter is organised as follows. Section 2.1 presents an overview of Digital Libraries, including definitions, and introduces the key concepts behind Digital Library Systems. A comprehensive discussion of some popular free and open source software (FOSS) tool used to create digital collections is also presented in this section. Finally, a number of projects and on-going research that helped form the working hypotheses for this research are also described; these are simple, custom tailored, digital repository tools specifically designed to be used in environments with resource limitations --a characteristic common in many developing countries. Section Error: Reference source not found presents related work and is thus a discussion of various software frameworks and reference models that have been applied to the implementation of Digital Library Systems.

2.1 Digital Libraries

2.1.1 Definitions

The field of Digital Libraries is a multidisciplinary field that comprises disciplines such as data management, digital curation, document management, information management, information retrieval and library sciences. Fox et al. [10] outline the varying impressions of Digital Libraries from persons in different disciplines and adopt a pragmatic approach of embracing the different definitions. They further acknowledge the metaphor of the traditional library as both empowering and recognise the importance of knowledge systems that have evolved as a result. Arms [2] provides an informal definition by viewing a Digital Library as a well organised, managed collection of information --with associated services-- stored in digital formats that is accessible over a network.

In an attempt to overcome the complex nature of Digital Libraries, Gonçalves et al. [12] define a Digital Library, using formal methods, by constructively defining a minimal set of components that make up a Digital Library. The set-oriented and functional mathematical formal basis of their approach facilitates the precise definition of each component as functional compositions.

The European Union co-funded DELOS Networked Excellence on Digital Libraries working group proposed a reference model and drafted The Digital Library Manifesto with the aim of setting the foundations and identifying concepts within the universe of Digital Libraries [7]. The DELOS Digital Library reference model envisages a Digital Library universe as a complex framework and tool having no logical, conceptual, physical, temporal or personal borders or barriers on information. A Digital Library is perceived as an evolving organisation that comes into existence through a series of development steps that bring together all the necessary constituents each corresponding to three different levels of conceptualisation of the universe of Digital Libraries [6]. The DELOS Digital Library reference model is discussed in depth in Section 2.3.1.2 .

2.1.2 Application Domains

The use of Digital Libraries has become widespread mainly due to the significant technological advances that have been taking place since the 1990s. The advent of the Internet has particularly influenced this widespread use. There are various application domains in which Digital Libraries are used and researchers are continuously coming up with innovative ways of increasing the footprint of Digital Library usage.

Academic institutions are increasingly setting up institution repositories to facilitate easy access to research output. Digital Libraries play a vital role by ensuring that intellectual output is collected, managed, preserved and later accessed efficiently and effectively. Figure 2.1 is an illustration of an institution repository system --an online[1] research document archive for the Computer Science Department of the University of Cape Town.

Figure 2.1 UCT Computer Science Department Research Document Archive

[1] http://pubs.cs.uct.ac.za

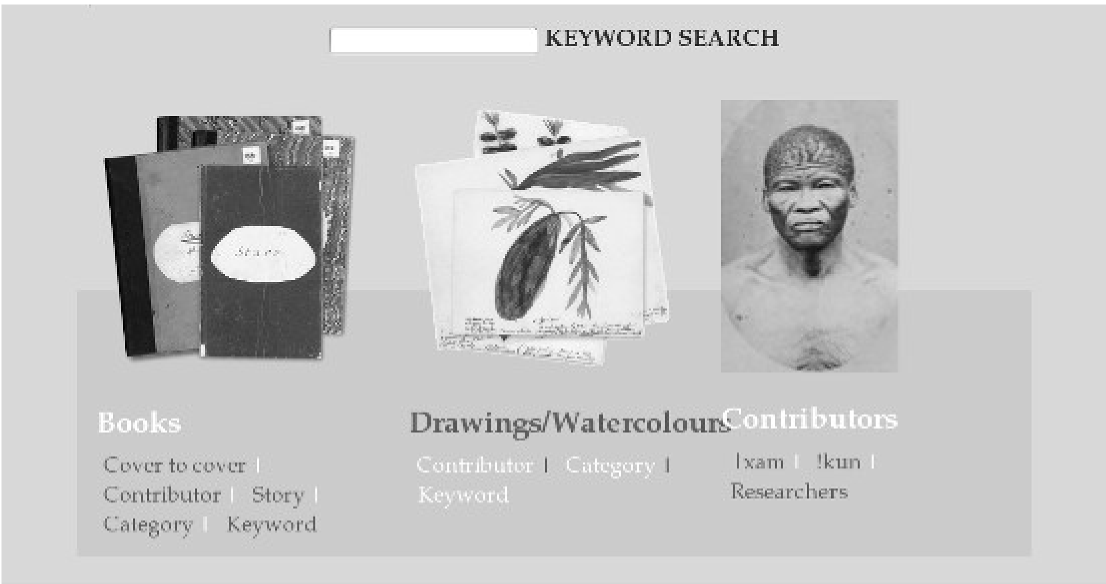

Figure 2.2 The Digital Bleek and Lloyd Collection

Cultural heritage organisations are increasingly digitising historical artefacts in a quest to display them online to a much wider audience. In light of this, Digital Libraries Systems are being developed to enable easy access to this information. Figure 2.2 is an illustration of the Digital Bleek and Lloyd Collection[2], which is a digital collection of historical artefacts that document the culture and language of the |Xam and !Kun groups of Bushman people of Southern Africa.

There has also been an increasing number of large scale archival projects that have been initiated to preserve human knowledge and provide free access to vital information [15].

2.2 Fundamental Concepts

2.2.1 Identifiers

An identifier is a name given to an entity for future reference. Arms [3] classifies identifiers as vital building blocks for Digital Libraries and emphasises their role in ensuring that individual digital objects are easily identified and changes related to the objects linked to the appropriate objects. He also notes that they are also essential for information retrieval and for providing links between objects.

The importance of identifiers is made evident by the widespread occurrence of standardised naming schemes such as Digital Object Identifiers (DOI)[3], handles[4] and Persistent Uniform Resource Locators (PURL)[5].

URIs are considered a suitable naming scheme for digital objects primarily because they can potentially be resolved through standard Web protocols. Another interesting point to note is

---

2 http://lloydbleekcollection.cs.uct.ac.za
3 http://www.doi.org
4 http://www.handle.net
5 http://purl.oclc.org

that facilitates interoperability, a feature that is significant in Digital Libraries whose overall goal is the widespread dissemination of information.

2.2.2 *Interoperability*

Interoperability is a system attribute that enables it to communicate and exchange information with other heterogeneous systems in a seamless manner. Interoperability makes it possible for services, components and systems developed independently to potentially rely on each other to accomplish certain tasks with the overall goal of having individual components evolve independently, but be able to call on each other, thus exchanging information, efficiently and conveniently [28]. Digital Libraries interoperability has particularly made it possible for federated services [13] to be developed mainly due to the widespread use of the OAI-PMH protocol for metadata harvesting.

There are various protocols that have been developed to facilitate interoperability between heterogeneous Digital Library Systems. Prominent interoperability protocols include Z39.50 [27] --a client server protocol used for remote searching, OAI-PMH [21], which has been extensively used for metadata harvesting, and RSS [32], a Web based feed format commonly used for updating Web resources.

It is also interesting to note that XML has emerged as the underlying language used to support a number of these interoperability protocols largely due to its simplicity and platform independence.

2.2.3 *Metadata*

Metadata is basically representational information that includes all pertinent descriptive annotations necessary to understand a digital object. Arms [4] describes different categories of information as being organised as sets of digital objects --a fundamental unit of the digital library architecture-- that are composed of digital material and key-metadata. He defines the key-metadata as information needed to manage the digital object in a networked environment.

The role performed by Metadata is both implicit and explicit and its functions can be more broadly divided into distinct categories. A typical digital object normally has administrative metadata for managing the digital object, descriptive metadata to facilitate the discovery of information, structural metadata for describing relationships within the digital object and preservation metadata which store provenance related information.

Metadata is made up of elements that are grouped into a standard set, to achieve a specific purpose, resulting into a metadata schema. There are a number of metadata schemas that have been developed as standards across various disciplines and they include among others Dublin Core [9], LOM [17], METS [25] and MODS [26].

Metadata can either be embedded within the digital object --as is the case with Portable Document Format (PDF) and Hypertext Transfer Markup Language (HTML) documents-- or stored separately with links to the resources being described. Metadata in Digital Libraries is generally stored in databases for easy of management and access.

2.2.4 *Standardization*

The fast pace at which technology is moving has spawned different breeds of application software tools. This means that the choice of which technology to use in any given instance differs, thus complicating the process of integrating application software with other heterogeneous software tools. Standards become particularly useful in such situations because they form the basis for developing interoperable tools and services. A standard is a specification --a formal statement of a data format of protocol-- that is maintained and endorsed by a recognised standards body [37].

Adopting and adhering to standards has many other added benefits, and Strand et al. [35] observe that applications that are built on standards are more readily scalable, interoperable and portable, constituting to software quality attributes that are important for the design, implementation and maintenance of Digital Libraries. Standards also play a vital role in facilitating long term preservation of digital objects by ensuring that documents still become easily accessible in the future. This is done by ensuring that the standard itself does not change and by making the standard backward compatible.

Notable use of standards in Digital Libraries include the use of Extensible Markup Language (XML) as the underlying format for metadata and OAI-PMH as an interoperability protocol. Digital content is also stored in well known standards, as is the case with documents that are normally stored in PDF/A format.

The use of standards in Digital Libraries, however, has its own shortcomings; in certain instances, the use of standards can be a very expensive venture as it may involve a lot of cross-domain effort [23].

2.3 Related Work

Several Digital Library frameworks [12, 19], reference models [7] and repository software tools [5, 16, 20, 30, 34] have addressed specific problems in Digital Library System architectural design and implementation. A discussion of two prominent reference models --the 5S framework and the DELOS reference model-- and the influence of design decisions on resulting architectures of existing repository software tools now follows.

2.3.1 Reference Models

A reference model is an abstract framework that provides basic concepts used to understand the relationships among items in an environment. The Organization for the Advancement of Structured Information Standards (OASIS) [24] states that a reference model consists of a minimal set of unifying concepts, axioms and relationships within a particular problem domain, and is independent of specific standards, technologies, implementations or other concrete details.

A number of Digital Library reference models have been proposed; a discussion of two prominent ones now follows.

2.3.1.1 5S Framework: Streams, Structures, Spaces, Scenarios and Societies

The 5S framework is a unified formal theory for Digital Libraries. It is an attempt to define and easily understand the complex nature of Digital Libraries in a rigorous manner. The framework is based on formal definitions, and abstraction of five fundamental concepts --Streams, Structures, Spaces, Scenarios, and Societies [12]. The five concepts together with their corresponding definitions and examples are summarised in Table 2-1.

In the context of the overall aims of a Digital Library, Gonçalves et al. [12] outlined an association between the 5S to some aims of a Digital Library System, with Streams being aligned with the overall communication and consumption of information by end users;

Structures supporting the organisation of information; Spaces dealing with the presentation and access to information in usable and effective ways; Scenarios providing the necessary support for defining and designing services and Societies defining how a Digital Library satisfies the overall information needs of end users.

However, Candela et al. [6] state that the 5S framework is very general-purpose and thus less immediate. The 5S framework is also arguably aimed at formalising the Digital Library aspects.

S-Concept	Concept Definition	Examples
Streams	Streams represent a sequence of elements of an arbitrary type	Text, video, audio, software
Structures	Structures specify the organisation of different parts of a whole	Collection, document, metadata
Spaces	Spaces are sets of objects with associated operations, that obey certain constraints	User interface, index
Scenarios	Scenarios define details for the behaviour of services	Service, event, action
Societies	Societies represent sets of entities and the relationship between them	Community, actors, relationships, attributes, operations

Table 2-1 Summary of Key Aspects of the 5S Framework

2.3.1.2 DELOS Digital Library Reference Model

The DELOS Network of Excellence on Digital Libraries[6] was a European Union co-funded project aimed at integrating and coordinating research activities in Digital Libraries. published a manifesto that establishes principles that facilitate the capture of the full spectrum of concepts that play a role in Digital Libraries [7]. The result of this project was a reference model --the DELOS Digital Library reference model-- comprising of a set of concepts and relationships that collectively attempt to capture various entities of the Digital Library universe.

A fundamental part of the reference model is the Digital Library Manifesto, that presents a Digital Library as a three-tier framework consisting of a Digital Library (DL), representing and organisation; a Digital Library System (DLS), for implementing DL services; and a Digital Library Management System (DLMS), comprising of tools for administering the DLS. Figure 2.3 shows the interaction between the three sub-systems.

The reference model further identifies six core concepts that provide a firm foundation for Digital Libraries. These six concepts --Content, User, Functionality, Quality, Policy, and Architecture-- are enshrined within the Digital Library and the Digital Library System. All concepts, with the exceptions of the Architecture concept, appear in the definition of the Digital Library. The Architecture is, however, handled by the Digital Library System definition [6].

6 http://www.delos.info

The Architecture component, addressed by the Digital Library System, is particularly important in the context of this research as it represents the mapping of the functionality and content on to the hardware and software components. Candela et al. [6] attribute the inherent complexity of Digital Libraries and the interoperability challenges across Digital Libraries as the two primary reasons for having Architecture as a core component.

Another important aspect of the reference model, directly related to this research, are the reference frameworks needed to clarify the Digital Library universe at different levels of abstraction. The three reference development frameworks are: Reference Model, Reference Architecture, and Concrete Architecture. In the context of architectural design, the Reference Architecture is vital as it provides a starting point for the development of an architectural design pattern, thus paving way for an abstract solution.

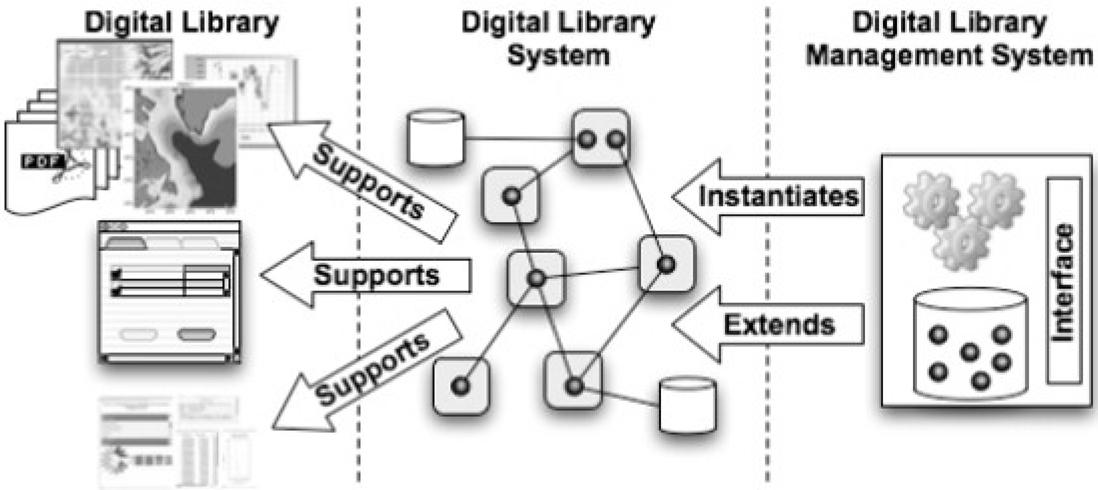

Figure 2.3 DL, DLS and DLMS: A Three-tier Framework

2.3.1.3 Summary

Table 2-2 , an adaptation from the DELOS reference model[31], is a mapping showing the relationship between the DELOS reference model and the 5S framework.

The motivation behind building both the reference models was largely influenced by the need to understand the complexity inherent in Digital Libraries. The idea of designing a Digital Library architecture based on direct user needs is not taken into account in existing reference models, although the DELOS Reference Architecture does have a provision for the development of specific architectural design patterns. The DELOS Reference Architecture is in actual fact considered to be mandatory for the development of good quality Digital Library Systems, and for the integration and reuse of the system components.

		Streams	Structures	Spaces	Scenarios	Societies
Content	Data and information handled and made available by the Digital Library	X	X			
User	Actors (human and/or machine) entitled to interact with the Digital Library					X
Functionality	Services provided to different users of the Digital Library			X	X	
Quality	Parameters used to characterise and evaluate content and behaviour of the Digital Library					
Policy	Sets, rules and regulations governing the interaction between actors and the Digital Library					
Architecture	Entities that represent a mapping of functionality and content offered by the Digital Library					

Table 2-2 Mapping Between the DELOS Reference Model and the 5S Framework

2.3.2 Repository Software

The different architectural designs of existing Digital Library software tools can be more broadly classified into two categories, namely: Centralised Component Architectures and Distributed Component Architectures.

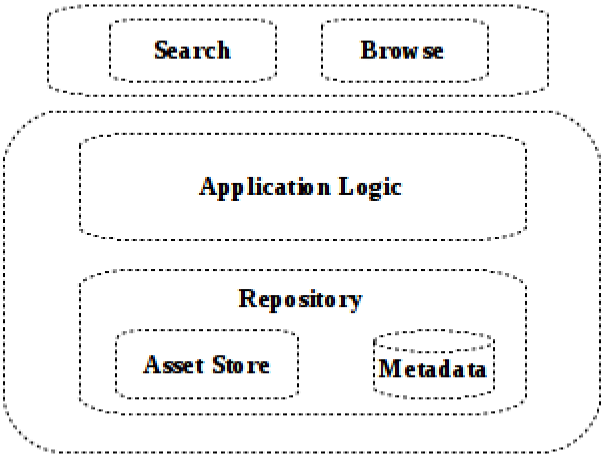

Figure 2.4 Digital Library Centralised Component Architecture

2.3.2.1 Centralised Component Architectures

In centralised component architectures, the components are located on one machine or in a cluster of coupled machines in a single location. All repository management and administrative related activities are performed from one location. The high level architecture is layered and is based on the traditional three-tier architecture, comprising of a client tier, a business logic tier, and a storage tier. A notable feature common to the storage tier of systems that conform to this architecture is the use of database management systems (DBMS) as the underlying technology for storage of metadata. Figure 2.4 is a high level view of a centralised component architecture. Notable software tools that confirm to this architectural design include CDS Invenio [40], DSpace [39], EPrints [43], and ETD-db [33].

CDS Invenio runs on GNU/Unix platforms, with a MySQL database backend server for storing metadata, and an Apache/Python Web application server. The architecture of CDS Invenio is composed of modules, with specifically defined functionalities, that interact with each other, the database and the interface layers [30].

DSpace resulted from a collaboration between Massachusetts Institute of Technology (MIT) and Hewlett Packard Labs and the architectural design was largely influenced by the desire to the system to function as a repository for digital research and educational material. DSpace is explicitly organised into a three-tier architecture, consisting of an application layer; a business logic layer and a storage layer. Digital run within a Servlet container with digital content stored in a designated area on the filesystem. However, the corresponding representational information is stored in a relational database management system (RDMBS) [39].

EPrints is a generic archive software tool that is designed to facilitate the creation of highly configurable Web-based archives. EPrints runs within an Apache HTTP server, stores digital content on the filesystem, and the metadata information is stored in a MySQL database [16].

ETD-db was developed at Virginia Tech and is among the most widely used E-thesis packages, particularly due to the fact that it enjoys widespread support from the Networked Digital Library of Thesis and Dissertation (NDLTD) [18]. ETD-db is a digital repository software for searching, browsing and collections. It is designed specifically for management of Electronic Thesis and Dissertations (ETDs);

A limitation of centralised component architectures is arguably their inability to scale, thus making it difficult to store very large digital collections.

2.3.2.2 Distributed Component Architectures

Distributed component architectures have their components and contents spread across multiple machines that could potentially be in different geographic regions. The integration between local and distributed components is done via Web service Application Programming Interfaces (APIs). The distribution of services helps with scalability by distributing the load and storage among several repositories. Figure 2.5 shows a typical architecture for a distributed Digital Library System.

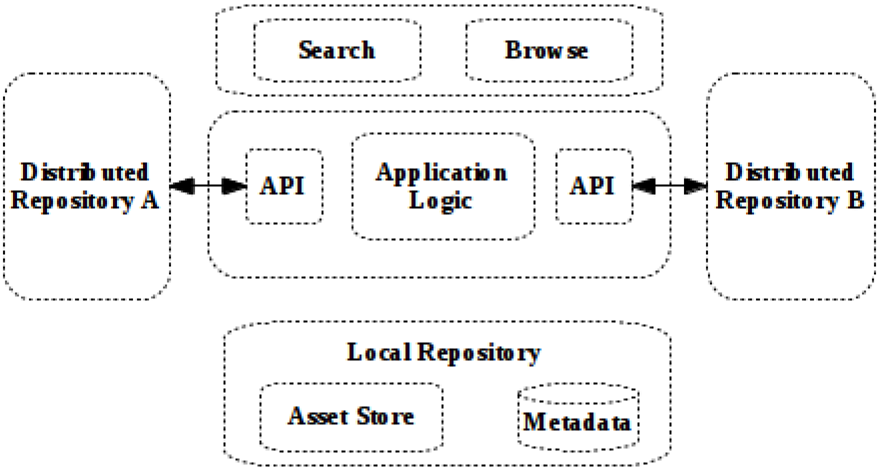

Figure 2.5 Digital Library Distributed Component Architecture

Fedora Commons [29] and Greenstone [5] are typical examples of Digital Library tools with a distributed architectural design.

Fedora is a scalable digital content repository software specifically designed for managing and delivering complex digital objects. Fedora's architecture was designed to handle any type of digital content and its key strength is its inherent support for long term preservation. Fedora's distributed model also makes it possible for complex digital objects to make reference to content stored on remote storage systems. However, Fedora is complex and lacks a dedicated user interface –a characteristic that has led to further development of Web interface tools such Fez[7] and Islandora[8].

Greenstone was designed for building and distributing Digital Library collections [42]. The motivation behind the development of Greenstone was to enable universities, libraries and other public institutions to build their own Digital Libraries. The ability to redistribute collections on a self-installing CD-ROM has particularly made it a popular tool in regions with very slow Internet connections. Greenstone's architecture is decentralised, making the system scalable, flexible and extensible. The flexibility enable Greenstone to support

7 http://fez.library.uq.edu.au
8 http://islandora.ca

distributed collections capable of being served from different machine, but at the same time maintaining a consistent presentation view to the end user. The digital objects in a Greenstone implemented Digital Library are classified into documents and resources. Documents are expressed in XML and are encoded, using specific plugins, into a METS framework representation. The METS encoded representation is then stored into an SQL database for rapid retrieval from the live Digital Library.

2.3.2.3 Summary

The repository software tools surveyed are Web based, support OAI-PMH and all, but Greenstone use a database management system as the underlying storage component for metadata. It is also interesting to note that only Greenstone supports an alternative way to distribute digital collections.

Table 2-3 is a summary of the architectural inclined features of repository software tools discussed in this section.

		CDS Invenio	DSpace	EPrints	ETD-db	Fedora Commons	Greenstone
Storage	Complex objects support					X	
	Dublin Core support for metadata		X	X		X	X
	Metadata can be stored in database	X	X	X	X	X	X
	Metadata can be stored on filesystem						X
	Supports distributed repositories		X	X		X	X
	Support for relationships among digital objects					X	
Technologies and Services	Extensible via plugins		X	X		X	X
	OAI-PMH compliant	X	X	X	X	X	X
	Platform independent		X	X		X	X
	Supports Web Services		X			X	X
	URI support (e.g. DOIs, PURL or handles)		X				
Features	Alternative mode of distribution (e.g. CD-ROM)						X
	Easy to setup, configure and use (out-of-the-box installation)			X			X
	Handles different file formats	X	X	X		X	X
	Hierarchical collection structure		X	X		X	
	Horizontal market software	X	X	X		X	X
	Web interface (GUI)	X	X	X	X	X	X
	Workflow support		X	X	X		

Table 2-3 Feature matrix for some popular Digital Library System tools

2.3.3 Simple Collection Architectures

Various custom solutions, in the form of lightweight systems that utilise simple architectures, have been devised to address problems that result from inadequate funding allocated to digitisation projects and limited broadband Internet connectivity. A discussion of two approaches to simple architectural designs now follows, outlining features that are most prominent and how simplicity is supported.

2.3.3.1 Hybrid Online-Offline Collections

The conventional way of accessing digital collections has been via online digital repositories, however, in certain instances, offline systems might help foster replication, simplicity and cost-effectiveness. Suleman et al. [38] describe an attempt to exploit advantages of both online and offline techniques in the form of a combined online-offline digital repository. A prototype lightweight, distributable, generic repository management and access system was implemented to illustrate that a hybrid online-offline solution can seamlessly provide current updates via and online component integrated into a larger offline collection. The online-offline integration is particularly interesting as the online component silently fails if there is no Internet connectivity, effectively providing a seamless user experience.

It can be argued that most digital collections are characterised by a larger initial digitisation effort, followed by a trickle of updates over time [38]. The potential for such a hybrid solution is thus limitless, especially in environments with limited broadband Internet connectivity.

However, such a hybrid solution might not be desirable for digital collections that periodically change over time; a typical example would be an institutional repository system. It would perhaps be particularly applicable to specific categories of collections, such as cultural heritage collections.

2.3.3.2 File-based Storage Systems

Digital Libraries System architectures have generally been designed to make use of database management systems (DBMS) as the underlying storage layer for metadata. Databases are indeed useful in that they enable efficient ways to insert and update records, have a standardised query language and generally come pre-installed on most machine platforms.

However, it can be argued that most operations on Digital Library Systems generally involve reading data, with minimal insert and update operations. A filesystem is an efficient place for storing data, and particularly becomes useful for metadata since search operations are performed on metadata in Digital Library Systems. The process of backing up and restoring data or moving the collection is also simplified as it generally only involves copying the collection to another server. Additionally, because there is generally one server --the Web Server-- as opposed to two separate services --a DBMS and a Web server--, installation of a digital collection is easier, and the overall server requirements are generally lower.

Suleman [36] demonstrates how XML, XSLT and XHTML can be used to generate a usable static, portable XML-centric Digital Library; he also illustrates how scalability and efficiency concerns are addressed during generation and access of the static Digital Library System.

2.3.3.3 Summary

There is increasing a growing number of architectural designs, for special repository software tools, that are being developed and aimed at solving specific problems in the developing world. However, it is interesting to note that the problems can be more broadly be classified into similar categories; including lack of adequate funding, lack of skilled human resource and lack of reliable broadband Internet connectivity.

There is thus a great need for the development of a reference model specifically aimed at guiding software engineers building tools for such environments.

Bibliography

[1] Adam, N.R., Bhargava, B.K. and Yesha, Y. eds. 1995. *Digital Libraries Current Issues*. Springer-Verlag.

[2] Arms, W.Y. 2000. *Digital Libraries*. The MIT Press.

[3] Arms, W.Y. 1995. Key Concepts in the Architecture of the Digital Library. *D-Lib Magazine*.

[4] Arms, W.Y., Blanchi, C. and Overly, E.A. 1997. An Architecture for Information in Digital Libraries. *D-Lib Magazine*.

[5] Bainbridge, D., Buchanan, G., Mcpherson, J., Jones, S., Mahoui, A. and Witten, I.H. Greenstone : A platform for distributed digital library applications.

[6] Candela, L., Castelli, D., Ferro, N., Ioannidis, Y., Koutrika, G., Meghini, C., Pagano, P., Ross, S., Soergel, D., Agosti, M., Dobreva, M., Katifori, V. and Schuldt, H. 2007. *The DELOS Digital Library Reference Model (Version 0.98) - Foundations for Digital Libraries*.

[7] Candela, L., Castelli, D., Pagano, P., Thano, C., Ioannidis, Y., Koutrika, G., Ross, S., Schek, H.J. and Schuldt, H. 2007. Setting the foundations of digital libraries: The delos manifesto. *D-Lib Magazine*.

[8] Dempsey, L. and Weibel, S.L. 1996. The Warwick Metadata Workshop: a framework for the deployment of resource description. *D-Lib Magazine*. Online Computer Library Center.

[9] Dublin Core Metadata Element Set, Version 1.1: 1999. *http://www.dublincore.org/documents/dces/*. Accessed: 2012-01-04.

[10] Fox, E.A., Akscyn, R.M., Furuta, R.K. and Leggett, J.J. 1995. Digital libraries. *Communications of the ACM*.

[11] Fox, E.A., Eaton, J.L., McMillan, G., Kipp, N.A., Weiss, L., Arce, E. and Guyer, S. 1996. National Digital Library of Theses and Dissertations. *D-Lib Magazine*.

[12] Gonçalves, M.A., Fox, E.A., Watson, L.T. and Kipp, N.A. 2004. Streams, structures, spaces, scenarios, societies (5s). *ACM Transactions on Information Systems*. 22, 2 (Apr. 2004), 270-312.

[13] Gonçalves, M.A., France, R.K., Fox, E.A. and Tech, V. 2001. MARIAN : Flexible Interoperability for Federated Digital Libraries. *Proceedings of Research and Advanced Technology for Digital Libraries, 5th European Conference (ECDL 2001)* (2001).

[14] Griffin, S.M. 1998. NSF/DARPA/NASA Digital Libraries Initiative. *D-Lib Magazine*.

[15] Gutenberg: the history and philosophy of Project Gutenberg: 1992. *http://www.gutenberg.org/wiki/Gutenberg:The_History_and_Philosophy_of_Project_Gutenberg_by_Michael_Hart*. Accessed: 2012-01-06.

[16] Gutteridge, C. 2002. GNU EPrints 2 Overview. *11th Panhellenic Academic Libraries Conference* (2002).

[17] IEEE 2002. Draft Standard for Learning Object Metadata, IEEE 1484.12.1-2002, 2002.

[18] Jones, R. 2004. DSpace vs. ETD-db: Choosing software to manage electronic theses and dissertations. *Ariadne*.

[19] Kahn, R. and Wilensky, R. 2006. A framework for distributed digital object services. *International Journal on Digital Libraries*. 6, 2 (Mar. 2006), 115-123.

[20] Lagoze, C., Payette, S., Shin, E. and Wilper, C. 2005. Fedora: an architecture for complex objects and their relationships. *International Journal on Digital Libraries*. 6, 2 (Dec. 2005), 124-138.

[21] Lagoze, C., Van de Sompel, H., Nelson, M. and Warner, S. 2002. Open Archives Initiative-Protocol for Metadata Harvesting-v. 2.0.

[22] Lagoze, C., Summary, E. and Daniel, R. 1996. The Warwick Framework A Container Architecture for Aggregating Sets of Metadata. *Organization*. (1996), 1-26.

[23] Lorist, H. and van der Meer, K. 2001. Standards for digital libraries and archives: digital longevity. *1st International Workshop on New Developments in Digital Libraries* (2001), 6–7.

[24] MacKenzie, M., Laskey, K., Mccabe, F., Brown, P.F., Metz, R. and Hamilton, B.A. 2006. *Reference Model for Service Oriented Architecture 1.0*.

[25] Metadata Encoding and Transmission Standard (METS) Official Web Site: *http://www.loc.gov/standards/mets/*. Accessed: 2012-01-04.

[26] Metadata Object Description Schema: MODS (Library of Congress): *http://www.loc.gov/standards/mods/*. Accessed: 2012-01-04.

[27] National Information Standards Organization 1994. Information Retrieval (Z39.50): Application Service Definition and Protocol Specification. *Information Retrieval*.

[28] Paepcke, A., Chang, C.-chuan K., Winograd, T. and García-Molina, H. 1998. Interoperability for digital libraries worldwide. *Communications of the ACM*.

[29] Payette, S. and Lagoze, C. 1998. Flexible and Extensible Digital Object and Repository Architecture (FEDORA). *Management*. 4 (1998), 41-59.

[30] Pepe, A., Baron, T., Gracco, M., Le Meur, J.Y., Robinson, N., Simko, T., Vesely, M. and Meur, J.-y L. 2005. CERN Document Server Software: the integrated digital library. *ELPUB 2005 conference, Heverlee (Belgium)* (2005), 8–10.

[31] Programme, S.F., Programme, I.C.T., Heritage, C., Learning, T.E. and Number, P. 2010. DL . org : Coordination Action on Digital Library Interoperability , Best Practices and Modelling Foundations. *Library*. September 2009 (2010).

[32] RSS Advisory Board 2007. RSS 2.0 Specification. (2007).

[33] Resources for Developers of ETD databases: 2002. *http://scholar.lib.vt.edu/ETD-db/developer/*. Accessed: 2012-01-07.

[34] Smith, M., Barton, M., Branschofsky, M., McClellan, G., Walker, J.H., Bass, M., Stuve, D. and Tansley, R. 2003. DSpace - An Open Source Dynamic Digital Repository. *D-Lib Magazine*.

[35] Strand, E.J., Mehta, R.P. and Jairam, R. 1994. Applications thrive on open systems standards. *StandardView*.

[36] Suleman, H. 2007. Digital Libraries Without Databases: The Bleek and Lloyd Collection. *Proceedings of Research and Advanced Technology for Digital Libraries, 11th European Conference (ECDL 2007)* (Budapest, Hungary, 2007), 392-403.

[37] Suleman, H. 2010. Interoperability in Digital Libraries. *E-Publishing and Digital Libraries: Legal and Organizational Issues*. Information Science Publishing. 31-47.

[38] Suleman, H., Bowes, M., Hirst, M. and Subrun, S. 2010. Hybrid online-offline digital collections. *Proceedings of the 2010 Annual Research Conference of the South African Institute of Computer Scientists and Information Technologists on - SAICSIT '10* (New York, New York, USA, 2010), 421-425.

[39] Tansley, R., Bass, M., Stuve, D., Branschofsky, M., Chudnov, D., McClellan, G. and Smith, M. 2003. The DSpace institutional digital repository system: current functionality. *2003 Joint Conference on Digital Libraries, 2003. Proceedings.* (2003), 87-97.

[40] Vesely, M., Baron, T., Meur, J.-Y. and Simko, T. 2004. CERN document server: Document management system for grey literature in a networked environment. *Publishing Research Quarterly*. 20, 1 (Mar. 2004), 77-83.

[41] Weibel, S., Kunze, J., Lagoze, C. and Wolf, M. 1998. Dublin core metadata for resource discovery. *Internet Engineering Task Force RFC*.

[42] Witten, I.H., Boddie, S.J., Bainbridge, D. and McNab, R.J. 2000. Greenstone: a comprehensive open-source digital library software system. *Proceedings of the fifth ACM conference on Digital libraries - DL '00*. (2000), 113-121.

[43] Eprints.org Software for Creating Institutional and Individual Open Archives.