Data Management in Digital Archives

Kamilah Louw
Computer Science
University of Cape Town
Cape Town
LWXKAM003@myuct.ac.za

ABSTRACT
Digital libraries are not always successfully implemented or reliable in developing countries with low resources. SimpleDL provides users with a toolkit to create simple digital libraries while mitigating the issues experienced with digital libraries in low resource areas. This paper discusses the Simple Archive Management project with a focus on data and file management while highlighting other similar toolkits for building digital archives and the differences in comparison to SimpleDL. It also emphasizes the principles for the creation and maintenance of data management systems, and steps of digital archiving and challenges faced within these countries. This paper highlights the principles of the creation and maintenance of data management systems. SimpleDL addresses a number of issues with regards to digital archives including the implementation and usability in developing countries. The paper also highlights the necessary steps in digital archiving while discussing their importance and the role they play in the bigger picture of usability of the archive. Cloud storage and cloud computing applications and architecture are also discussed within the paper. It is concluded that, while there are many advantages to digital archiving, the biggest problem and threat is the risk of technological and resource obsolescence. Research found that resources of digital archives need to be refreshed to stay accessible and compatible with new software and hardware in order to achieve its main goal - to provide a place to store, organize, manage and preserve valuable resources for future generations.

1. INTRODUCTION

1.1 The ARCHMAN Project
The Simple Archive Management project, also known as the ARCHMAN project, involves the design and implementation of a new and improved usable user interface for administrators of the SimpleDL toolkit. SimpleDL is a toolkit used to create digital libraries [1]. It is currently being used to archive and manage collections of historical documents collected by the University of Cape Town (UCT). A digital archive is a safe place to store records electronically which have some historical importance or hold great value for a long period of time, regardless of the format of these records [2]. This is vital as it reduces the chances of these records being damaged, stolen or lost due to negligence. Digital archives also allow these documents to be accessed in numerous locations without physically handling the original documents. This is very advantageous in fields where these records are rare, fragile and treasured, such as in the field of history. It is important that this toolkit is not only intuitive for the user but also efficient and effective since historians study the past and, thus, need a good data management system in order to store old resources and files [3].

1.2 Where does this Review fit into The Project?
The focus of this review is the file and data management in digital archives. SimpleDL requires a usable user interface and since the primary use of a digital archive is to store important data in a safe place for long-term usage, a good management system will be necessary. This management system needs to be usable and intuitive for researchers and administrators to easily find, store and manage files on the system.

1.3 Subfields within Data Management of Digital Archives
This literature review will discuss subfields within data management. These subfields include the broad aspects of data management systems, issues of data management systems from end-users, the different types of management systems, a discussion about current technologies and tools used for data and file management, principles, advantages, challenges and disaster recovery of file and data management systems.

2. TOOLKITS FOR BUILDING DIGITAL ARCHIVES
There are various toolkits that have been developed to assist in the creation of repositories. DSpace and EPrints are the most popular toolkits; both are web-based and have back-end databases. There are other toolkits that are available both online or offline and some of which were created to avoid the need for software installation [1]. This paper will also discuss the Greenstone Digital Library (GSDL) which is an open source multilingual software [5]. SimpleDL was created with the intention of being a simple way to allow long-term access to digital repositories even without a stable network, with computer system failures and no active preservation [1].

2.1 Administration within File Management Systems
In a traditional file management system, once a resource is submitted, it is placed in a virtual waiting area and is inaccessible to end users until the administrator or specific users approve it. Depending on the policies of the institution, somebody might review the document to ensure that the preferred format is used, while another person verifies plagiarism. Following the approval step, another person usually assigns appropriate metadata to the document and prepares it for publishing via the file management system [6].

2.2 DSpace
DSpace is a collaborative project of the Massachusetts Institute of Technology (MIT) Libraries and Hewlett-Packard (HP) labs [7]. It

is an open source platform that serves as a repository for digital content created by organizations and institutions for research and teaching. The system contains communities, which have collections of related content that are grouped together. These communities are created for the labs, centers, schools and departments within the institution and each has their own designated area within the system. Members of the communities have the ability to upload content directly via a Web user interface.

The system also has an alternative method of loading bulk content via a batch item importer [8]. It also includes a variety of features such as visualization and simulation of stored data [7]. Figure 1 illustrates the depositing interface used by authorized users to upload material onto DSpace [8].

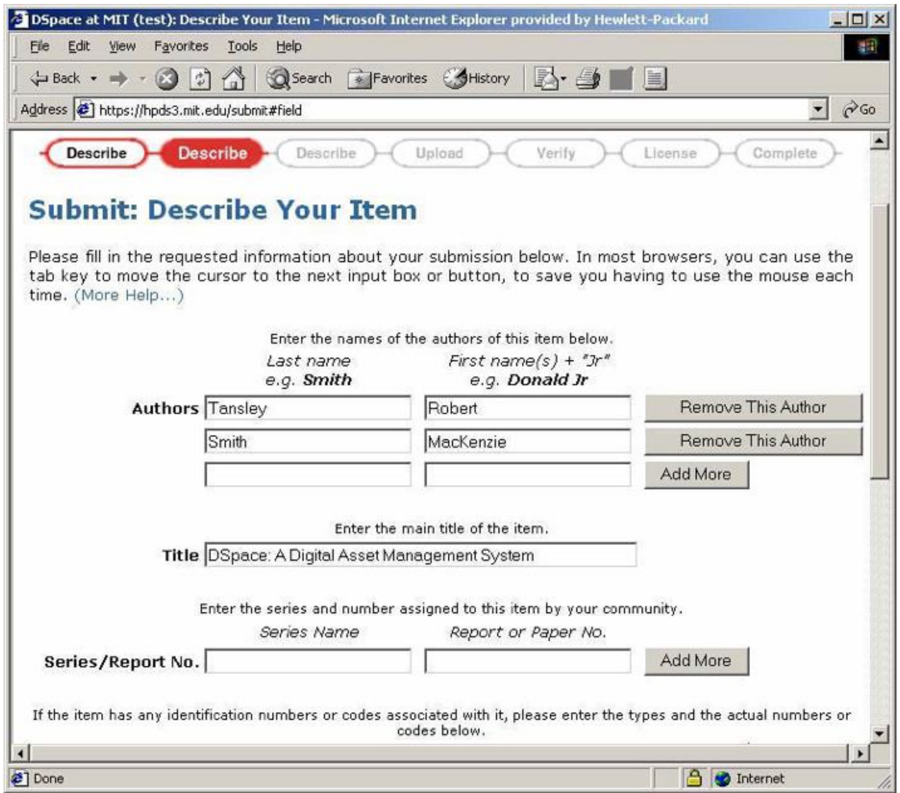

Figure 1. Depositing Interface for DSpace [7]

When building a repository in DSpace, it performs three tasks - capturing the digital content with the corresponding metadata, listing the resources systematically while assisting in searching with the use of keywords and metadata and, supporting the preservation of the digital data [5].

The two main functionalities of DSpace that will be discussed in this paper are the data model which is used for the organization of data, and the various types of metadata stored by the system.

2.2.1 Data Model

The organization of data in DSpace is meant to reflect the structure of the organization that is using the DSpace system. The resources of the archive are subdivided into bundles of bitstreams (streams of bits). Bundles are composed of bitstreams that are closely related, for example a bundle could be a collection of HTML and image bitstreams which form an HTML document while the bitstream could be a single HTML file or a single image file. A bitstream format is a distinct and consistent way to refer to a specific file format. An implicit or explicit notion of how material in that format can be interpreted is an essential component of a bitstream format. For interoperability and discovery, each archived resource has a qualified Dublin Core metadata[1] entry [8].

2.2.2 Metadata

DSpace contains three types of metadata about archived material namely, administrative, descriptive and structural metadata. Administrative metadata includes metadata for preservation, lineage, and authorization policy data. The majority of this is stored in DSpace's relational database management system (DBMS). The DBMS makes use of Oracle and PostgreSQL [9].

The resource's lineage metadata is stored in the Dublin Core metadata records. Descriptive metadata includes the metadata on the Dublin Core as well as some other descriptive data such as a name and a descriptive composition which is stored in the database management system. Structural metadata includes information about how to convey an archived resource, or bitstreams within a resource as well as the relationships between the item's individual parts [8]. Custom metadata schema can also be created using Extensible Markup Language (XML) [10].

DSpace can be easily customized to manage and preserve digital content while also making it accessible to users. DSpace can store a variety of digital records, including technical reports, images,

[1] The Dublin Core Metadata Element Set also known as "The Dublin Core", is a collection of fifteen "core" properties (attribute) for resource description [28].

theses, multimedia publications, articles, administrative records, conference papers, audio-video files, web pages, books, and so on [5].

2.2.3 Workflow

The batch item importer converts an external Submission Information Package, consisting of an XML metadata document, into an "in progress submission" item. Depending on the policy of the collection where the submission is being made, a workflow process could possibly be started. These workflows are defined within the collection and are set up by the administrator. The task that is placed into a "task pool" where one member can remove the task in order to process it. The member can either accept or reject the submission or edit the metadata [8]. DSpace also includes some elementary SHERPA/RoMEO API look-up capabilities however, it is only for authority control.

2.3 EPrints

EPrints is a free software application created by the University of Southampton in England. It allows the academic community to submit preprints, post prints, and other academic publications via a web interface [7]. EPrints employs traditional technologies and operates entirely on Open Source systems. It makes use of MySQL, Apache database and web server. Any metadata schema provided by the administrator can be used by the EPrints archive. The administrator determines which metadata fields are stored for each resource [9]. The administrative back-end provides access to configuration options [10]. EPrints makes use of MySQL, Oracle and PostgreSQL for the database management system [9]. Unlike DSpace, EPrints does not allow for bulk uploading of resources [11].

2.3.1 Data Model

Unlike DSpace, Eprints has a less structural division of sections and collections with the central concept being that all files are equivalent and do not form a hierarchical system. Since a hierarchy is needed to navigate through resources in the repository, it makes use of representations which can be used to proceed with navigation using related elements of the metadata information [12].

2.3.2 Metadata

Just as DSpace, users can upload documents with the necessary metadata for the records by filling out a web form with all the information [5]. The fields that appear in the web form are selected by the administrator and easily customizable to allow for relevant fields to be presented to the end-user depending on the resource [11]. EPrints can use a controlled vocabulary and authority lists to ensure high metadata quality. It includes native Dublin Core support as well as the ability to export to a variety of file types [10].

2.3.3 Workflows

This software connects to the SHERPA/RoMEO database, allowing authors to verify their rights to their submissions in the repository. In this manner, any unauthorized submission by the content-publisher is dealt with [5]. All back-end options, such as web interface appearance and functionality, record organization, and other server-side settings, are managed by the administrator. A record is reviewed by an editor before it is submitted online. Editors also control the metadata of the submissions. The author role facilitates document submission and management of earlier submitted documents. The administrator has control over the browsing categories that are made visible to the user. Eprints allows for documents in its archive to be indexed by google to ensure better access to resources. [11].

2.4 Greenstone

Greenstone is an open source software that was released under the General Public License (GNU) and is commonly used to create repositories and publish them online [5]. Greenstone collections can be distributed on CDROM (Compact Disk Read Only Memory) but accessing them requires software installation [1]. The New Zealand Digital Library Project at the University of Waikato, UNESCO, and the Human Info NGO collaborated to create and distribute GSDL. The goal of the software was to allow users to create their own digital libraries [5].

2.4.1 Data Model

The Greenstone model envisions a "librarian" who creates collections from existing "resources" (which include both "items" and metadata) and distributes them via the Web or on removable media [13]. Since the software is built for dabbler individuals to create their own personalized collections, these features make GSDL suitable to preserve digitized resources. In text documents, the representation of hierarchical structure is supported for chapters, sections, and paragraphs [14].

2.4.2 Metadata

Greenstone provides Dublin Core metadata standards on all of its resources but also enables users to use their own metadata scheme. This is possible through extending an existing metadata scheme or by defining a new scheme using the metadata editor [13]. Special XML tags are used to define specific sections within a text document. XLinks within the records can be used to connect it to other records or resources [14]. Users define metadata within the Librarian interface, resources with similar characteristics are grouped into collections and uploaded [15].

2.4.3 Workflows

Professional librarians are solely responsible for content management on GSDL. When files are added to a specific collection on GSDL, it must be rebuilt for full text indexing and browsing structures, whereas on DSpace collection building occurs incrementally. The 'Greenstone Librarian Interface' is a Java-based interface which allows the administrator, possibly a librarian, to manage resources and collections [14]. The "administrative" function within the software allows selected users to authorize new users to build collections, protect documents so that only registered users with a password can access them, etc. The administrator is given the ability to add users as well as delete or create groups [15]. Before files can be viewed online, they must first be imported and built. Documents are initially imported into the XML-compliant Greenstone Archive Format. The archive files are then organized into searchable indexes and a collection information database with hierarchical structures for browsing. Once completed, the collection will be available to go online and respond to information requests [12]. "Receptionists" and "collection servers" are the two primary elements in Greenstone. The user inputs information and interacts through the receptionist. Then, request routed to the appropriate collection servers after being analyzed for the user interface and the processed information is returned to the receptionist for delivery to the user [15].

3. DIGITAL ARCHIVES

3.1 Advantages of Digital Archives

Digital archives are entirely digital and do not require any physical space such as a physical archive where physical files and documents are stored. Since these archives are digital, searching, storing, organizing and managing these files can be done quite effortlessly. The digitalization of these files allows for easy, remote access via the internet and at any time. These resources can also be accessed by multiple users at a time whereas the physical hardcopy in a physical archive would only be available for single usage or depending on the number of copies available. Digitized documents also allow for longer preservation and conservation since the original copy is never physically handled and therefore never exposed to wear and tear [2]. Having digitized files also allows for referenced data to be easily accessible, greater searching capabilities, simplification of teaching and research and ability to easily cross-reference and link files [16].

3.2 Principles for Creation and Maintenance

Before creating a digital archive there are some key issues that need to be taken into consideration. When building a digital management system for a business or institution, the existing guidelines that highlight how digital records are accessed, maintained and organized should be maintained. Before implementing a digitization process, the business or institution should clearly outline the objectives, namely, the access or the combination of access and preservation [17]. Digitized records should be crossed-checked with the original file to ensure that they do not contain any uncorrectable error, and an error report should be maintained for future observation. The data carriers (machine that can record and convey machine-readable data) should be examined at regular intervals for data integrity. The old carriers, hardware, and formats can become outdated hence digital records should be copied before this occurs. At least two copies should be retained in separate locations with access at any time. Digital copies of records could be subjected to copyright legislation. The records should therefore be in accordance with intellectual property rights and other legal rights related to copying, storage and modification of digital files. The digital management system should prioritize files in accordance with their practical requirements, historical significance and risk of content loss [2].

3.3 Steps of Digital Archiving

There are various significant steps involved in the process of digital archiving. Digital archiving is not only preserving digital information for future use, but also systematically organizing and managing it so that it can be retrieved as required. The steps involved include creation, acquisition, identifying and categorizing, storage, preservation and access.

3.3.1 Creation

The creation stage involves the initial stage of converting resources into a digital form for the use of archiving and preservation. This allows for online access globally. There are two types of digital material that may be digitized, namely, analog material that includes printed books, manuscripts, etc. and machine-readable material such as digital photographs, multimedia, websites etc. There are various tools used during the creation process such as scanning, optical character recognition, multimedia technologies, metadata and many more. [2]. The techniques used when creating a digital resource have an impact on how easily the resource can be digitally archived and preserved. When issues of consistency, format, standardization, and metadata description are addressed early in the creation process, the preservation and archiving process becomes more efficient. It is best practice to create the metadata at the time the file is created [16].

3.3.2 Acquisition

This step of the process includes the virtual and physical integration of digital files into the archive. The file should be known to the administrator or the archive. There are some aspects that need to be considered in the acquisition of digital resources; these may include collection policies, selecting what materials to archive and refreshing archived materials [16]. Collection policies should be determined to include the digital files in the digital archive and assist in the tailoring of the general collection practices of the institution or organization. Selection of which materials to archive can be simplified by creating a guideline that outlines the selection of resources to be incorporated into the archive due to legal or other issues associated with the resource. Refreshing the archived resources is also an important step in acquisition. Archived materials require refreshing due to changes and updates that occur in formats and software and hardware to keep resources compatible [2].

3.3.3 Identifying and Categorizing

Organization of archived files is achieved through identification and categorizing to manage resources over time. Identification is achieved through the use of a unique key to find files and link them to other related resources. Categorizing is related to organizing what type of resources are being archived in the form of metadata [16]. Metadata (data about data) is used in archives for description, identification, summarization, structuring, administration and preservation and reuse of files. It improves the discovery and access of digital files within the system and is key to ensuring that files continue to be accessible over time. Digital identification is achieved through the use of file names, Uniform Resource Locators (URLs) or the Digital Object Identifier (DOI). It assists in the locating of digital files, however digital resources move from server to server or from one directory to another on a network, which results in the URL changing. This is why persistent identifiers (a long-lasting reference to a resource) are preferred. Despite all these possible complications, digital archiving systems continue to use URLs when specifying the location of resources [2].

3.3.4 Storage

Any storage method is reliant on a very specific combination of software and hardware for access. Archivists and developers should consider durability, feasibility, cost, capacity, obsolescence and susceptibility when selecting a storage medium for a digital archiving system. Technological obsolescence is unavoidable but choosing the correct storage medium could delay it [2]. It is critical to understand the various storage media because they require different software and hardware for access, as well as various storage conditions and preservation demands [18].

3.3.5 Preservation

The main goal of preservation is to protect digital resources for a long period of time to allow future generations to still be able to access them. It is the process through which digital files are conserved in their digital form to ensure durability, usability and integrity of the information. It includes all of the plans, strategies, allocations of resources, and actions necessary to guarantee that originally digital and reformatted information is accessible despite obstacles posed by media failure and technical advancement. The fundamental issue with digital preservation is digital obsolescence because there are no set standards, protocols, or techniques for keeping digital data safe [2].

3.3.6 Access

Changes in access mechanisms typically take place over time, therefore successful practice must take into account long-term security requirements as well as Digital Right Management (DRM). Problems in accessing and using digital resources for archiving are dealt with by Digital Right Management. DRM systems are intended to limit the copying, sharing, reformatting, and modifying of digital media while facilitating access to and use of digital resources [2]. All the steps mentioned previously are performed to ensure continuous access to resources in the archive. Security and version control have an impact on digital archiving as well. When it comes to conservation, it is critical to have metadata to manage encryption, watermarks, digital signatures, and other features that can survive changes in the format and media [16].

3.4 Challenges with Creation and Maintenance in Developing Countries

There are many challenges that need to be addressed in the creation and maintenance of digital archives especially in developing countries where resources, expertise, technological infrastructure and funding are scarce. There are various issues which will be discussed in this paper namely, dynamic software and hardware, technical expertise, funding, substandard technology infrastructure and technological discontinuance.

3.4.1 Dynamic Software and Hardware

There are several digital technologies involved in the maintenance of digital archives. These technologies allow information to be generated, manipulated, distributed and stored with ease while preserving access to the information. Digital files stored on old technologies could be lost due to the hardware or software of these technologies becoming obsolete. Digital files and resources are stored in various formats and require specific software to interpret these formats. The continuous updates to software and hardware technology can cause difficulty in the preservation of digital content and information over long periods of time [2].

3.4.2 Technical Expertise

The creation of digital archives is dependent on the technical skills available. Digital archiving requires highly skilled people but the ability to employ and develop these human resources with the necessary skills is made more difficult by the rate of which technology is changing and developing [18]. In developing countries like in Africa, the technical skilled people are not on hand to implement, initiate and maintain preservation and digitization. There is a scarcity of human resources in developing countries and those that are trained lack the expertise and necessary skills to train others [17].

3.4.3 Funding

The creation of digital archives can be expensive, with various costs involved for the creation and maintenance of the system. With rapid changing technologies, access to digital files is not always easy and requires knowledgeable staff and substantial expenditure on technological requirements [18]. The development of IT infrastructure for the digital management system also comes at a cost and without the employment of technical professionals and skilled workers, the creation and maintenance of the digital archive system is not possible.

3.4.4 Substandard Technology Infrastructure

With most countries in Africa not having a reliable supply of electricity, it creates a bottleneck in the digitization in Africa making it impossible to maintain a suitable environment for digitization practices. The telecommunications infrastructures in many African countries are either absent or poorly developed [17].

3.4.5 Technological Discontinuance

Computer software and hardware are constantly improving and being updated. This can lead to technology obsolescence, posing a threat to digital archives in developing countries. The discontinuance of these technologies is caused by the upgrading of operating systems, programming language applications and storage devices [2]. A survey was conducted across 54 institutions where respondents were asked to rank the significance of four issues as threats to digital preservation. The data collected was too small for statistically significant conclusions, but it was found that technological obsolescence was ranked the greatest threat by the 54 institutions [19]. Figure 2 displays the data collected from the 54 institutions during the survey.

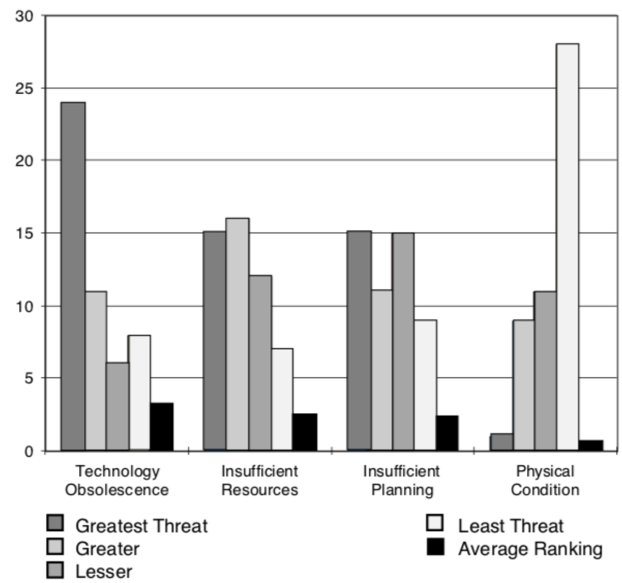

Figure 2. Threats to Digital Holdings Ranked by 54 Institutions [18]

Updating and converting software to current versions can be complicated and may cause difficulty in accessing files which were compatible with previous versions. This may lead to files being lost or inaccessible.

4. DATA AND FILE MANAGEMENT SYSTEMS

4.1 Role in Digital Archives

Digital archives contain huge amounts of files, documents and records, which could be in different formats, contain different information, be different sizes, have different levels of importance etc. Johnston et al. [2005] touches on the benefits of electronic records management systems, it was mentioned that the National Archives of Australia define an electronic records management system as an automated system which is used to control the use, creation, maintenance and distribution of electronically generated records [20]. A data management system would be used to manage, create, organize and store the digitized files in a digital archive.

4.2 Current Data and File management System of SimpleDL

Currently administrators can manage files through a Web interface. This interface allows administrators to authorize new files, comments and requests for user accounts [1]. Figure 3 shows the default administrator interface used on SimpleDL.

5.3 One Drive

In 2007 Microsoft launched their own cloud storage called "One Drive". Users can use One Drive to access their own files from any device and even offline by downloading the application. Files are also encrypted to ensure privacy and security [21].

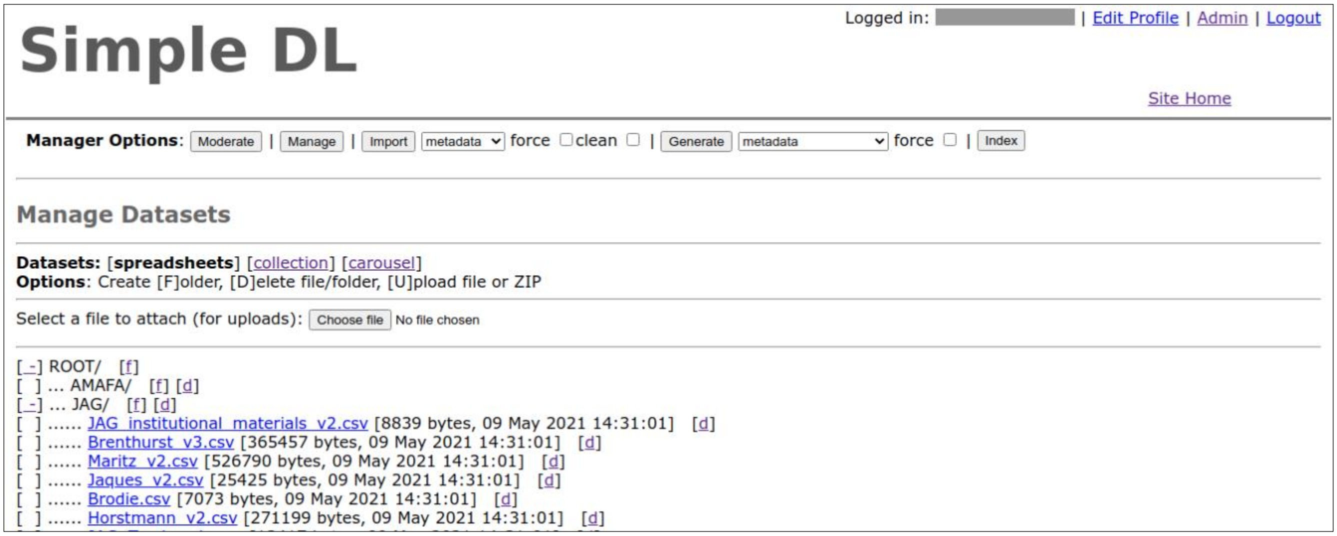

Figure 3. Web administrator interface of SimpleDL [1]

SimpleDL is a model of a digital archival system with the goal of being functional in low resource environments, without much computational power and with very little maintenance being required. It is a solution for network failures or during unsuccessful operating system upgrades but facilitates entire digital libraries to be copied as simply as individual files.

Some possible improvements include ease of installing and testing, performance and stability of various scripts, archiving of entire libraries and authoring of complex objects [1].

5. CLOUD-BASED APPLICATIONS FOR FILE AND DATA MANAGEMENT

Cloud storage applications allow users to download, upload, share and manage their files while providing them with access anywhere. It is a combination of servers, storage and networks and contains hardware and software components. [21]. Cloud computing removes the hassle of having to maintain large storage and computing devices. It allows for large amounts of data to be shared while providing customers with the ability to collaborate [22]. It has a minimal cost structure since it allows users to modify their services based on their needs and requirements [23]. Below three different cloud storage providers will be introduced as well as the general architecture of cloud storage.

5.1 Google Drive

Google Drive enables its users to store files in the cloud which sit on Google's servers, share data and synchronize files across devices. It also allows users to sign into multiple accounts and easily transition between them. It also offers shared drives to enable customers to work collaboratively [21].

5.2 Dropbox

Similar to Google Drive, Dropbox enables its users to share, sync and edit documents over the internet. Users can connect to secure servers for file sharing, file updates and ease of accessibility [21].

5.4 The Architecture

The basic services of each application are relatively similar, but each application has various cost options which allow users to unlock additional features. Storage options also vary between applications. Here, the basic architecture of cloud storage systems will be discussed. The architecture for cloud computing is divided into two, the Front End and Back End which are both connected through the internet. The front end is for the users where they can interact with interfaces and applications to access their files. The back end is responsible for the information security for users and is used by service providers [21]. The basis of a cloud storage system is that it requires a single data server connected to the internet. A client sends files to the data server via the internet, where the information is recorded. When a user wants to obtain information, they use a web-based interface to connect to the data server. The server either returns the files to the user or allows them to access and manipulate the files on the server [24]. The front end of cloud storage architectures exports an API to access storage, this is where one would find Web service front ends or file-based front ends. Behind this layer is the middleware which implements a variety of features like data reduction and replication. The back end is where the physical data storage is located. This layer can be either an internal protocol that utilizes specific features or a conventional backend to physical disks [25].

6. SUCCESS FACTORS FOR DATA AND FILE MANAGEMENT SYSTEMS

Apart from the criterion created by Maxwell [2010], there has not been a substantial amount of coverage regarding the success factors that can be used to measure the performance of a data management system. His paper is based solely on feedback from end-users and does not offer a permanent or structured criterion. There is no generally agreed set of attributes that have been outlined in order to check and measure the success and performance of a data management system. There are some papers that cover the principles that should be taken into account when creating file management systems but none of these directly address performance, they are all applicable for any generic or basic implementations [20].

Johnston et al. [2005] also briefly discuss whether data management systems meet standards since there are no set performance standards to measure against. The only paper with a title that somewhat speaks to the critical success factors for file management systems was focused mainly on local government in Poland. The aim of the paper was to propose a framework of critical success factors for data management systems in government units. Papaj et al. [2022] was able to formulate an extensive list of 23 factors that could be considered prerequisites for a successful data management system. This data was collected through interviews with employees of local and state governments who were responsible for the data management systems as well as professors of Polish universities [25].

Alshibly et al. [2016] also investigated the creation of a list of critical success factors for file management systems but their methodologies included systematic literature review, a questionnaire survey and eight in-person interviews [26].

7. DISASTER RECOVERY OF DATA MANAGEMENT SYSTEMS

Johnston et al. [2005] mention the lack of resources that highlight the recovery of data management systems after a disaster. In their paper they mention that recovery would require the database where the metadata is stored to be restored as well as restoring all settings and choices. Restoration would also include trying to recover lost electronic resources. [20].

Cervone [2006] discusses the importance of a disaster recovery plan, the development of the plan as well as type of threats and potential solutions. There are three key phases that need to be addressed in the development of a disaster recovery plan. These include the initial planning, development of the infrastructure to initiate the plan and testing of the plan. A business impact and risk analysis are the best way to start planning as it will provide a foundation for the plan, management and testing. It also assists users to focus on what can be implemented to reduce the risk before the disaster occurs, how people should respond should the disaster occur and how to recover. There are a few different types of threats that can affect an organization such as technical threats, natural threats and human threats. A risk analysis could provide a method to determine what the highest priority situations are. Cervone [2006] also discusses the use of hot-sites and cold-sites. Hot-sites provide remote recovery instantaneously by replicating the computing infrastructure and operations of the business. A cold site does not provide immediate recovery of operations and the operations on the main site are not replicated simultaneously. He also mentions that sometimes onsite recovery solutions can be used if operations are required to be at the same location as the organization [27].

Although he discusses some very important points, there is still a lack of substantial information about this topic. There are no reports to assess how businesses and organizations of recent disasters have recovered and how they were impacted both physically and financially. There are many other aspects that could be discussed.

8. CLOSING THE GAP

The SimpleDL project aims to find solutions for the challenges faced when creating digital archiving systems as well as the maintenance and costs associated with it. The proposal to improve the interface of the SimpleDL system will improve usability. SimpleDL in itself also addresses the issue of disasters and disaster recovery. Since the toolkit was designed to work without much computational power and much maintenance, it also closed the gap of the issues faced by archivists in low resource environments such as developing countries. This means that SimpleDL addresses all the issues of other digital libraries such as network failures, operating system upgrade failures and allows for entire digital libraries to be copied as easily as single items. [1]

9. CONCLUSIONS

Designing and implementing a data and file management system requires a lot of planning and a well thought out design. The design should be intuitive and allow the user to spend less time interacting with the interface when trying to access information. The storage of resources is also important as this affects the efficiency of the system and the ease of looking up resources or linking resources. There should be set guidelines put in place for the digitalization of resources to ensure longer life for resources and avoid the chances of resources becoming obsolete.

It is important that the user interface is both comfortable, easy to understand and efficient as a poor interface would deter users and could reduce their chances of having access to useful resources. Users should not be required to do any additional work such as installing software. The main focus should be placed on browsing of resources with the least number of clicks and any other user input.

10. REFERENCES

[1] H. Suleman, "UCT Computer Science Research Document Archive," 3 December 2021. [Online]. Available: https://pubs.cs.uct.ac.za/id/eprint/1512/. [Accessed 12 March 2023].

[2] D. Yadav, "Opportunities And Challenges In Creating Digital Archive And Preservation: An Overview," vol. 6, no. 2, pp. 1-11, 2016.

[3] A. Maxwell, "Digital archives and history research: feedback from an end-user," *Library Review,* vol. 59, no. 1, pp. 25-30, 2010.

[4] S. Ma, "The Function of the Establishment of Digital Archives in Colleges and Universities to the Optimization of Resources," IOP Publishing Ltd, Shenyang, 2019.

[5] L. Verma and N. Kumar, "Comparative Analysis of Open Source Digital Library Softwares: A Case Study," *DESIDOC Journal of Library & Information Technology,* vol. 38, no. 5, pp. 361-368, 2018.

[6] M. N. Masrek and H. Hakimjavadi, "Evaluation of Three Open Source Software in Terms of Managing Repositories of Electronic Theses and Dissertations: A Comparison Study," *Journal of Basic and Applied Scientific Research,* vol. 2, no. 11, pp. 10843-10852, 2012.

[7] G. Biswas and D. Paul, "An evaluative study on the open source digital library softwares for institutional repository: Special reference to Dspace and greenstone digital library," *International Journal of Library and Information Science,* vol. 2, no. 1, pp. 001-010, 2010.

[8] R. Tansley, M. Bass, D. Stuve, M. Branschofsky, D. Chudnov, G. McClellan and M. Smith, "The DSpace Institutional Digital Repository System: Current Functionality," in *2003 Joint Conference on Digital Libraries*, Houston, 2003.

[9] M. Asif and M. A. Ugc-Net, "Emerging trends and technologies for digital transformation of libraries," *IP Indian Journal of Library Science and Information Technology,* vol. 4, no. 2, pp. 41-43, 2019.

[10] M. Castagné, "Institutional repository software comparison: DSpace, EPrints, Digital Commons, Islandora and Hydra," 14 August 2013. [Online]. Available: https://open.library.ubc.ca/soa/cIRcle/collections/graduateresearch/42591/items/1.0075768. [Accessed 21 March 2023].

[11] M. Beazley, "Eprints Institutional Repository Software: A Review," *the Canadian Journal of Library and Information Practice and Research,* vol. 5, no. 2, 2010.

[12] S. Tramboo, H. S. M. Shafi and S. Gul, "A Study on the Open Source Digital Library Software's: Special Reference to DSpace, EPrints and Greenstone," *International Journal of Computer Applications,* vol. 59, no. 16, pp. 0975-8887, 2012.

[13] I. H. Witten, D. Bainbridge, R. Tansley, C.-Y. Huang and K. J. Don, "StoneD A Bridge between Greenstone and DSpace," *D-Lib Magazine,* vol. 11, no. 9, 2005.

[14] M. N. Ravikumar and T. Ramanan, "Comparison of Greenstone Digital Library and DSpace: Experiences from Digital Library Initiatives at Eastern University, Sri Lanka," *Journal of The University Librarians Association of Sri Lanka,* vol. 18, no. 2, pp. 76-90, 2015.

[15] R. Parichi and F. Nisha, "Greenstone Digital Library Management System – a functional review based on selected criteria," *Library Hi Tech News,* vol. 32, no. 10, pp. 16-21, 2015.

[16] S. J. Kumaravel, "Is Digital Archiving Simple ?," in *2nd Convention PLANNER - 2004, Manipur Uni*, Imphal, 2004.

[17] B. E. Asogwa, "Digitization of Archival Collections In Africa For Scholarly Communication: Issues, Strategies, and Challenges," University of Nebraska - Lincoln, Nebraska, 2011.

[18] G. G. Moghaddam, "Preserving digital resources: issues and concerns from a view of librarians," *Collection Building,* vol. 29, no. 2, pp. 65-69, 2010.

[19] S. M. Montgomery and M. L. Hedstrom, "Digital Preservation Needs and Requirements in RLG Member Institutions," Research Libraries Group, Mountain View, 1999.

[20] G. P. Johnston and D. V. Bowen, "The benefits of electronic records management systems," *Records Management Journal,* vol. 5, no. 3, pp. 132-134, 2005.

[21] C. Ujjainkar, "Comparison Between Google Drive, One Drive and Dropbox," *International Research Journal of Modernization in Engineering Technology and Science,* vol. 4, no. 10, p. 1, 2022.

[22] M. Aazam, E.-N. Huh and M. St-Hilaire, "Towards Media Inter-cloud Standardization – Evaluating Impact of Cloud Storage Heterogeneity," *Journal Grid Computing,* vol. 16, no. 3, pp. 425-443, 2018.

[23] M. Z. Murah, "Teaching and Learning Cloud Computing," *Procedia - Social and Behavioral Sciences,* vol. 59, pp. 157-163, 2012.

[24] K. Sharmila, T. K. Kannan, C. Shanthi and R. Devi, "Study on Cloud Storage and its Issues in Cloud Computing," *International Journal of Management, Technology And Engineering,* vol. IX, no. I, pp. 976-981, 2019.

[25] E. Ziemba, T. Papaj and D. Descours, "Critical Success Factors for Adopting Electronic Document Management Systems in Government Units," in *Proceedings of the 17th Conference on Computer Science and Intelligence Systems*, Sofia, 2022.

[26] H. H. Alshibly, Y. Bao and R. Chiong, "Investigating the Critical Success Factors for Implementing Electronic Document Management Systems in Governments: Evidence From Jordan," *Information Systems Management,* vol. 33, no. 4, pp. 287-301, 2016.

[27] H. F. Cervone, "Disaster recovery and continuity planning for digital library systems," *OCLC Systems & Services: International digital library perspectives,* vol. 22, no. 3, pp. 173-178, 2006.

[28] DCMI Usage Board, "Dublin Core™ Metadata Element Set, Version 1.1: Reference Description," 14 June 2012. [Online]. Available: https://www.dublincore.org/specifications/dublin-core/dces/. [Accessed 21 March 2023].