

Literature Review - Three Archives

NOXOLO MTHIMKULU, University of Cape Town

Digital cultural heritage archives provide a mechanism for cultural heritage preservation. The audience of a digital cultural heritage spans from community members to academics therefore context is a factor when constructing a digital archive as it influences the level of acceptance of the technology. A well-designed digital archive focusses on the interface, the services offered to the user and the adoption of an efficient and reliable repository. Services span the standard search and browse and cross-over to the interactive services such as exhibitions. The repository governs the management of the digital objects and repository tools such as Omeka, Fedora, EPrints and DSpace provide mechanisms for management while adopting the Dublin Core metadata standard. The application of these principles can be seen in the Europeana and Bleek and Lloyd collections and are to be further applied and investigated in the implementation of the Three Archives cultural heritage archive.

Additional Key Words and Phrases: digital libraries, cultural heritage preservation, digital archives

1. INTRODUCTION

Cultural heritage preservation has migrated from physical archives to digital heritage archives. Digital archives manage and preserve multimedia information [Ludäscher et al. 2001]; providing accessibility of digitised artefacts in the present and the future [Gwangjing-gu 2014].

The topic being investigated, Three Archives, is aimed at the introduction of three cultural heritage archives pertaining to Cape Town. These digital cultural heritage archives are to represent the *Sequins, Self and Struggle* archive: a collection containing multimedia objects from the "Miss Gay Western Cape" and "Spring Queen" beauty pageants [<http://sequins-self-and-struggle.com/>]; the *Harfield Village* collection: an aggregation of artefacts about the forced removals of the Claremont residents; and *Movie Snaps*: a collection of photographs taken in and around central Cape Town before and during the apartheid [<http://www.cca.uct.ac.za/projects/movie-snaps-cape-town-remembers-differently/>].

This paper discusses the users and usability of digital archives, the services offered by digital archives, the architectures available for adoption when constructing a digital archive, and repository tools available for use as the foundation layer of a digital archive. Insight gained throughout the investigation about the components and concepts of digital archives will assist in concluding a way forward for the the development of the Three (Digital) Archives.

2. USERS AND USABILITY

Digital cultural heritage archives are designed for use by cultural heritage institutions; cultural heritage professionals, academics and scholars, tourism users and the general public [Chowdhury 2010]. Consideration of archive users and their specific requirements is a necessity in the construction of a digital cultural heritage archive. This section will outline what users require from a system and will apply the notion of technology acceptance to digital archives; in order to conclude what to focus on to achieve a well-accepted system by the users.

During the design of a digital archive, it is important to consider the users that

interact with the system; the system itself as well as the interaction between the users and the system [Huang et al. 2012]. Users expect the collections represented to be easily and efficiently accessible [Chowdhury 2010]. Users respond negatively when archives are perceived too difficult to operate as a result of the incorrect target audience [Hong et al. 2011]. It was noted that since the archives span across many different users, it is infeasible to assume that users have prior knowledge of the archives content. Users dismiss the search option and prefer a browse option [Ruecker et al. 2011; Blandford and Buchanan 2003] which resulted in the concept of the 'rich-prospect-browser' which satisfies the users' requirement of an overview of the archive contents [Ruecker et al. 2011].

Digital archive designers focus on two measures when evaluating whether a system will be accepted by its users. The measure of Perceived Usefulness (PU): the user seeing the system as something that would enhance their work efficiency; and Perceived Ease of Use (PEU): how effortless the system is to use and learn [Hong et al. 2011]. These measures are described in the Technology Acceptance Model (TAM) [Hong et al. 2011]. TAM provides an analysis of variables affecting a user's decision to use a particular technology. The application of TAM emphasises the importance of user-centered design and concludes that a system that is perceived useful and easier to use will be better accepted [Ruecker et al. 2011]. A well-designed user interface will encourage crowd interaction [Chowdhury 2010] which may be in the form of crowd-sourcing. The focus is shifting largely from the functional to the user experiences and interactiveness of the system [Chowdhury 2010] where previous approaches focussed primarily on the efficiency and usability functions but neglected the presentation [Huang et al. 2012]. Digital libraries are becoming focussed on people as opposed to the collection and the services they provide. The migration is from systems that only provide access to systems encouraging collaboration and user input [Candela 2007].

With users in mind, a successfully designed system should reduce the users cognitive load [Jones et al. 1995], and should reduce their fear of computers while increasing their ability to learn [Huang et al. 2012].

3. SERVICES

Following from the importance of the users of a digital archive, digital libraries are required to have services allowing for the searching (personalised and context-based), browsing, selecting, grouping and often personalisation of the the presentation of the digital objects contained [Paneva et al. 2005]. Over and above these standard services, the Web 2.0 services which can be offered, include geo-location of collections; user contributions to build the collections; and user tagging to help in categorising elements of the collections [Scheinfeldt 2008]. Services available in digital archives indicate a convergence of a library and a museum but with archival content. This can be noted by the increase in exhibition and show casting services being included [Gwangjing-gu 2014].

Digital archives allowing for crowd-sourcing in the form of users contributing their private collection are emerging; accompanied by chat services as well as services allowing users to create their own story [Barak et al. 2009]. The MOSAICA [Barak et al. 2009] system had tools and services providing functionality to investigate, which is a search engine allowing searching; exploration, which involves browsing collections; and the tagging of resources as per *Fig 1*.

The above indicates the wide range of services becoming available via archives

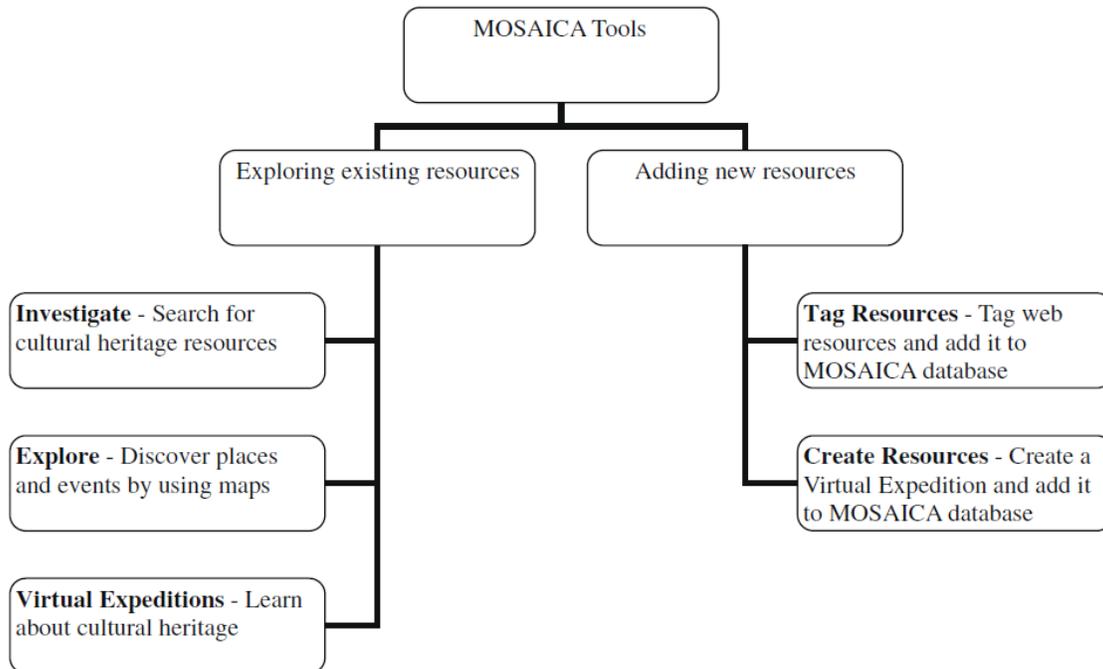


Fig. 1. MOSAICA tools and functionalities

rendering the archives more interactive and personalised. The services provided are supported by the architectural elements of the digital archives.

4. ARCHITECTURE

In order to accomplish the services stipulated, the designer needs to consider architecture requirements for the archive. A well-designed archive which will be interoperable and extensible needs to follow a flexible design with a minimalistic approach; there needs to be software and hardware independence as well as adherence to general archival standards and principles [Phiri and Suleman 2015].

Digital archives adopt a three-layered architecture with a user interface layer; the client services layer and the repository layer [Arms et al. 1997; Phiri and Suleman 2015]. The repository is a data structure used for the storage and management of digital objects [Arms et al. 1997; Kahn and Wilensky 2006].

The user interface and services layer have been extensively discussed in the above sections, this section will continue with a discussion on the repositories and their digital object management mechanisms. *Fig 2* depicts the components of a pilot system implemented by [Arms et al. 1997]. Here we take note of the handle system and the repository. The elements contained in the repository layer of a digital archive are the digital objects, which are used to manage the digital material and which are accessed and distributed across the network; an identifier for the digital object : handle; and the storage location of the digital object being the repository. [Arms et al. 1997; Kahn and Wilensky 2006].

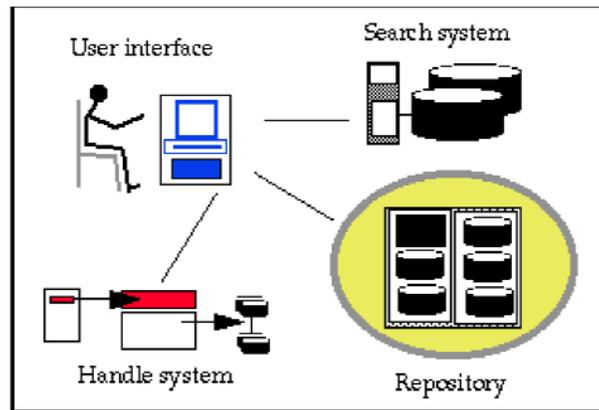


Fig. 2. Major components of a pilot system

The digital object is key here as it is what is being stored and represented in the archive. Digital objects have a structure of an explicit data type and metadata [Arms et al. 1997]. Interactions between the repositories and the digital objects they store are accomplished via a repository access protocol which governs the management of the digital objects [Kahn and Wilensky 2006].

Metadata is of crucial in digital libraries(archives) it allows for the organisation of the objects by good indexing, storage of the objects, and access and preservation [Hoe-Lian Goh et al. 2006]. The use of Dublin Core metadata standard increases the interoperability of the archives guaranteeing adherence to the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH)[<http://www.openarchives.org/>]. The OAI-PMH is a tool allowing a particular search engine to query an archive and obtain results of all the elements in the archive along with their Dublin Core metadata element [Gutteridge 2002].

Other architecture considerations are the federated architectures [Phiri and Suleman 2015] where these archives function through the integration of distributed archives and collections providing independent services [Phiri and Suleman 2015].

The high-level three-tier architecture is implemented throughout digital archives and the differences often arise in the repository layer and in the management of the digital objects. There exist tools allowing users to leverage off their repositories standards and management of the digital objects, these tools are discussed in the below section.

5. TOOLS

A digital archive can be entirely newly constructed or the developer can leverage off of repository and archiving tools available. These tools provide the infrastructure for the storage and management of the data objects and may provide a user interface that can be used in the actual archive.

The open-source digital archiving repository tools to be discussed are DSpace, an open-source software tool geared at providing management tools for digital objects [Kurtz 2013]; EPrints, an online archive of data objects usually consisting of research papers and theses [Gutteridge 2002]; Flexible and Extensible Digital Object and Repository Architecture (Fedora)[Payette and Lagoze 1998], an extensible digital content repository service providing services for the storage, management and distri-

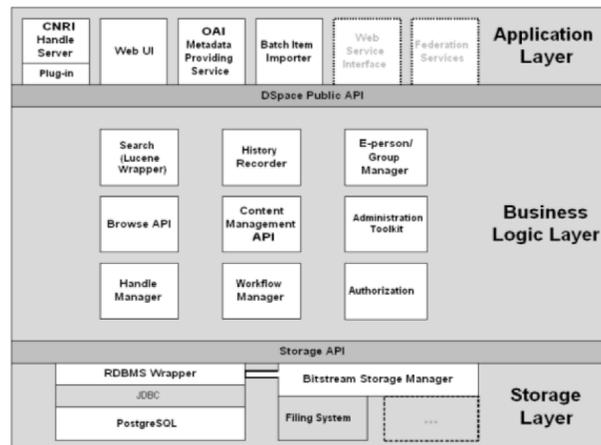


Fig. 3. DSpace technical Architecture[Smith et al. 2003]

bution of digital objects [Lagoze et al. 2006]; and Omeka, a tool for web publishing functioning as an archival service with mechanisms for the organisation and exhibition of collections[Morton 2011].

The digital archiving tools all boast an audience of members of the community wanting to self-archive their information and institutions and organisations wishing to do similar. DSpace targets organisations and institutions and provides a platform that can be used as-is or one that can be built and modified [Smith et al. 2003]. EPrints houses content of a scholarly and academic manner [Gutteridge 2002] indicating that their target audience are people wishing to self-publish their work or organisations involved in research. Fedora provides capabilities for users ranging from the inexperienced user to research institutions [Staples et al. 2003]. Omeka is aimed at cultural heritage professionals, academics and scholars [Scheinfeldt 2008]. It can therefore be observed that although some platforms offer and require more technical expertise, all are applicable to a technical/organisational background as well as for the needs of an archivist who would not have the expertise to extend the tools.

The differences in the tools emerge in their management of the digital objects. The DSpace has a technical architecture comprising the application layer, which is the interface into the system; the business layer which involves the work flow and the content management; and the storage layer which is where the information is being stored being via the PostgreSQL [Smith et al. 2003] as per *Fig 3*.

The system architecture comprising of digital object identifiers or handlers and they adhere to the OAI-PMH in order to ensure interoperability and the harvesting of their metadata [Smith et al. 2003].

The EPrints tool has a plugin architecture (as can be seen in DIAGRAM) consisting of a back-end which is the data model, followed by a plugin framework onto which plugins can be developed and added [Adewumi and Omoregbe 2011].

The Fedora repository architecture focusses on the object model which are templates for data objects and links to tools and services for managing these data objects. The behaviour objects as can be seen below store metadata describing operations of the

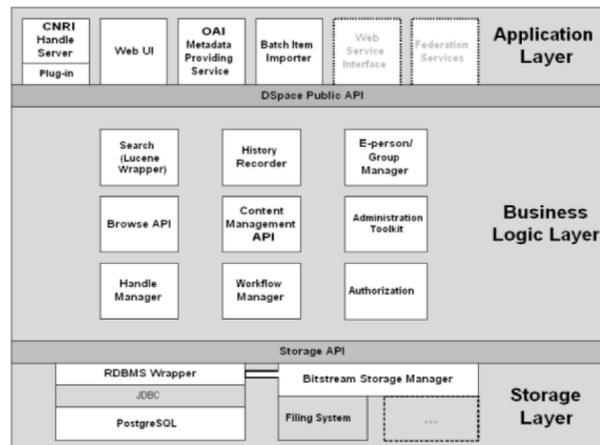


Fig. 4. Architectural Framework of EPrints [<http://www.eprints.org/software/training/programming/api-techniques.pdf>]

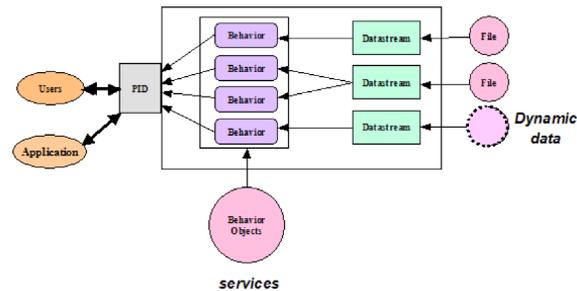


Fig. 5. The Fedora Object Model]

tools and services which the data objects leverage off. The datastreams in the image represent the digital resources being stored and their metadata. A URL describes the contents of the datastreams and this is what is used to identify and manage the objects in the Fedora repository [Staples et al. 2003].

Omeka adopts the Open Archives Initiative indicating that the data architecture was designed to abide by the standards of the Dublin Core metadata structure[Scheinfeldt 2008].

From the above it is clear that the different repository management tools differ in the services provided and in the management of the digital objects. The tools adopt the rationale behind the construction of any Institutional Repository which is flexibility, storage of various formats, accessibility, interoperability, and standard adherence [Adewumi and Omoregbe 2011].

6. FAMOUS ARCHIVES

6.1. Nelson Mandela Archive

The Nelson Mandela Archive [<http://archive.nelsonmandela.org/home>] is a multi-media online archive including books, photographs, videos about Nelson Mandela's

interactions with family, comrades and friends; as well as digitised versions of his diaries and books.

The archive was compiled by the Google Cultural Institute and pays tribute to this man by displaying his life for the world to see on this online archive.

The archive is well presented boasting services of browsing, searching, and exhibition service where we see certain elements of the archive dependent on the search and the browsing functionality. This archive is well presented and is easily learned by the user.

6.2. The Digital Bleek and Lloyd

The Bleek and Lloyd Collection [Skotnes et al. 2007] is a collection of notebooks and drawings that contain the work about the Xam and !Kun speakers of Southern Africa as interpreted by Lucy Lloyd and Wilhelm Bleek [Phiri and Suleman 2015]. The Digital Bleek and Lloyd collection is a digitised version of this comprising scans of the material collated in the original Bleek and Lloyd collection [<http://lloydbleekcollection.cs.uct.ac.za/>].

The architecture adopted in the construction of the digital archive does not make use of a database, instead, hyperlinks are used. This is the Hyperlink architecture where each object being is accessed via a hyperlink [Suleman 2007]

This strategy was successfully implemented using XML technology and worked given the static nature of the collection.

6.3. Europeana

Europeana is a European Digital Library which provides access to multimedia material located in digital libraries, museums and archives across Europe [Chambers and Schallier 2010].

The Metadata scheme adopted by Europeana is based on OAI-PMH [Koulouris and Garoufallou 2009] therefore ensuring interoperability. Europeana is an example of a federated architecture [Phiri and Suleman 2015] mentioned above. This architecture provides seamless access to services and data objects distributed amongst a variety of archives and involves data being periodically loaded into the database to ensure a consistent view of the data worldwide.

Although this archive presents a large amount of information, user interaction was criticised with justification that the search functionality was inadequate. The search was criticised as the mechanism did not search within the document but relied on bibliographic details of the objects. Another criticism was the lack of quick browsing given the users' desire for an overview of the objects contained [Maxwell 2010].

7. CONCLUSIONS

Digital cultural heritage preservation is growing and is moving from access based archives to interactive archives. The material discussed above highlighted the importance of the consideration of the user when designing a digital archive. Additionally, a consideration of services with the emergence of interactive services is necessary. The digital archival shift moving focus from the standard browse and search services to more interactive services. The analysis of the architectures available for adoption when constructing a digital archive and tools that assist in this emphasised the need

for an efficient data object management mechanism which pointed to the possibility of the adoption of and adaptation of existing archiving tools.

This information will assist in the construction of the Three Archives to be designed where a brief consideration of tools to best fit the archive architecture will be practically explored as well as the services that are implementable and in demand. It has become clear that there is no set structure in the development of an archive but rather tools to leverage off and build upon. This design process will be adopted in the design of the Three Archives digital cultural heritage website where multimedia information ranging from photographs to videos to recordings and to maps will be presented.

REFERENCES

- AO Adewumi and NA Omoregbe. 2011. Institutional repositories: features, architecture, design and implementation technologies. *Journal of Computing* 2, 8 (2011).
- William Y Arms, Christophe Blanchi, and Edward A Overly. 1997. An architecture for information in digital libraries. *D-Lib Magazine* 3, 2 (1997).
- Miri Barak, Orit Herscoviz, Zvia Kaberman, and Yehudit J Dori. 2009. MOSAICA: A web-2.0 based system for the preservation and presentation of cultural heritage. *Computers & Education* 53, 3 (2009), 841–852.
- Ann Blandford and George Buchanan. 2003. Usability of digital libraries: a source of creative tensions with technical developments. *IEEE Technical Committee on Digital Libraries Bulletin* 1, 1 (2003).
- Leonardo Candela. 2007. The DELOS digital library reference model: Foundations for digital libraries. (2007).
- Sally Chambers and Wouter Schallier. 2010. Bringing research libraries into Europeana: establishing a library-domain aggregator. *Liber Quarterly* 20, 1 (2010), 105–118.
- Gobinda Chowdhury. 2010. From digital libraries to digital preservation research: the importance of users and context. *Journal of documentation* 66, 2 (2010), 207–223.
- Christopher Gutteridge. 2002. GNU EPrints 2 overview. (2002).
- Seoul Gwangjing-gu. 2014. A Study of Larchiveum Data Model for the Design of Digital Heritage Museum1. (2014).
- Dion Hoe-Lian Goh, Alton Chua, Davina Anqi Khoo, Emily Boon-Hui Khoo, Eric Bok-Tong Mak, and Maple Wen-Min Ng. 2006. A checklist for evaluating open source digital library software. *Online Information Review* 30, 4 (2006), 360–379.
- Jon-Chao Hong, Ming-Yueh Hwang, Hsuan-Fang Hsu, Wan-Tzu Wong, and Mei-Yung Chen. 2011. Applying the technology acceptance model in a study of the factors affecting usage of the Taiwan digital archives system. *Computers & Education* 57, 3 (2011), 2086–2094.
- Angus FM Huang, Stephen JH Yang, and Shu-Sheng Liaw. 2012. A study of user's acceptance on situational mashups in situational language teaching. *British Journal of Educational Technology* 43, 1 (2012), 52–61.
- Marshall G Jones and others. 1995. Using Metacognitive Theories to Design User Interfaces for Computer-Based Learning. *Educational Technology* 35, 4 (1995), 12–22.
- Robert Kahn and Robert Wilensky. 2006. A framework for distributed digital object services. *International Journal on Digital Libraries* 6, 2 (2006), 115–123.
- ALEXANDROS Koulouris and EMMANOUEL Garoufallou. 2009. Managing cultural heritage content in the context of Europeana. In *Proceedings of the International Conference on Tourism Development and Management (ICTDM)*. 610–613.
- Mary Kurtz. 2013. Dublin Core, DSpace, and a brief analysis of three university repositories. *Information Technology and Libraries* 29, 1 (2013), 40–46.
- Carl Lagoze, Sandy Payette, Edwin Shin, and Chris Wilper. 2006. Fedora: an architecture for complex objects and their relationships. *International Journal on Digital Libraries* 6, 2 (2006), 124–138.
- Bertram Ludäscher, Richard Marciano, and Reagan Moore. 2001. Preservation of digital data with self-validating, self-instantiating knowledge-based archives. *ACM Sigmod Record* 30, 3 (2001), 54–63.
- Alexander Maxwell. 2010. Digital archives and history research: feedback from an end-user. *Library Review* 59, 1 (2010), 24–39.
- Amanda Morton. 2011. Digital Tools: Zotero and Omeka. *Journal of American History* 98, 3 (2011), 952–953.

- Desislava Paneva, Lilia Pavlova-Draganova, and Lubomil Draganov. 2005. Digital Libraries for Presentation and Preservation of East-Christian Heritage. In *Proceedings of the Second HUBUSKA Open Workshop Generic Issues of Knowledge Technologies, Budapest, Hungary*. 75–83.
- Sandra Payette and Carl Lagoze. 1998. Flexible and extensible digital object and repository architecture (FEDORA). In *Research and Advanced Technology for Digital Libraries*. Springer, 41–59.
- Lighton Phiri and Hussein Suleman. 2015. Managing cultural heritage: information systems architecture. (2015).
- Stan Ruecker, Milena Radzikowska, and Stéfan Sinclair. 2011. *Visual interface design for digital cultural heritage: A guide to rich-prospect browsing*. Ashgate Publishing, Ltd.
- Tom Scheinfeldt. 2008. Omeka: open source web publishing for research, collections and exhibitions. *Open Source Business Resource* December 2008 (2008).
- Pippa Skotnes, Wilhelm Heinrich Immanuel Bleek, and Lucy Lloyd. 2007. *Claim to the country: the archive of Lucy Lloyd and Wilhelm Bleek*. Jacana Media.
- MacKenzie Smith, Mary Barton, Mick Bass, Margret Branschofsky, Greg McClellan, Dave Stuve, Robert Tansley, and Julie Harford Walker. 2003. DSpace: An open source dynamic digital repository. (2003).
- Thornton Staples, Ross Wayland, and Sandra Payette. 2003. The Fedora Project. *D-Lib Magazine* 9, 4 (2003), 1082–9873.
- Hussein Suleman. 2007. *Digital libraries without databases: The bleek and lloyd collection*. Springer.