

DIGITAL RECORDS CURATION AT THE EAST AND SOUTHERN AFRICAN UNIVERSITIES INSTITUTIONAL REPOSITORIES (IRs)

Juliet Erima
Lecturer, Moi
University School of
Information Sciences

Kenya
julieterima@gmail.com

Tshepho L. Mosweu
Senior Lecturer,
University of Botswana

Botswana
lydhoss@gmail.com

Abstract - Most of the knowledge generated in academic institutions today is in digital form. Given that institutional repositories (IRs) across universities receive, preserve and make access to digital assets. The aim of this study is to assess the status of digital curation at Institutional repositories in selected Universities in Botswana and Kenya. The study takes a quantitative approach whereby data was collected through survey questionnaires administered amongst university IR staff in Botswana and Kenya. The data collected was analysed and presented with tables and figures. The Open Archival Information System (OAIS) Functional Model was used in this study as a lens to investigate the problem. The findings of the study show that public universities in Botswana and Kenya have established IRs that ingest digital records into their custody. Most resources ingested include thesis and publications by academic staff and students. These IRs store their digital records on local servers and other storages like CDs. This study found that the majority of the IRs both in Botswana and Kenya do not have digital records preservation plans. This study recommends the use of the OAIS model to preserve, manage and make access to digital records at East and Southern African public Universities Institutional Repositories.

Keywords: Botswana, Kenya, Universities, Digital records curation, Institutional repositories

Universities and higher education institutions are in the business of generating a lot of information and knowledge resources, both in analogue and digital formats. By comparison, digital content has become increasingly ubiquitous in present day organisations. Anderson and Rainie (2012) acknowledge that “we swim in a sea of data [...] and the sea level is rising rapidly”. Institutions are increasingly finding themselves “between a rock and a hard place” when facing rapidly changing technologies and the sheer volume of digital creation (Hedstrom, 1998). Due to the exponential creation of born-digital materials, information is being lost nearly as soon as digital assets are produced. As a result of this, individuals, institutions, and society as a whole need an accurate, complete and usable record of human activities, and an appropriate legal and institutional framework in which to use that record. Without trustworthy records, people and institutions cannot make informed decisions, verify existing information, evaluate evidence, hold others accountable, construct accurate histories or develop new knowledge (Prom, 2011). An authentic record does not preserve itself, and even the best-intentioned record creators often lack the resources or expertise to act as permanent custodians for non-current records. Nor can we rely on those who provide the service of temporarily storing and transmitting

INTRODUCTION

records to permanently preserve an interpretable record of human activity (Prom, 2011).

Harvey (2010) posits that technical obsolescence or fragility, lack of resources, ignorance of good practices, and uncertainty over appropriate infrastructure – all constitute serious risks to data. In previous years, digital preservation efforts originally focused on ensuring that material survived technical obsolescence and organisational mismanagement. Preservation implied a passive state, where material would be “dumped” in an inaccessible “dark archive”, with only a few authorised users, to ensure that it retained its integrity and authenticity.

Lately, the focus has shifted to ensuring that digital material is managed throughout its lifecycle so that it remains accessible to those who need to use it. Metadata is used to both improve accessibility and discoverability; and to control authentication procedures, creating audit trails to ensure that material cannot be accessed or altered by those not authorised to do so. Digital material is actively preserved, used and reused for new purposes, creating new materials. Unfortunately, relatively few institutional repositories in African public universities have implemented systematic institutional functions to preserve digital records in their keeping. Institutional repositories need a practical method to capture, preserve and provide access to records like email, blogs, digital photographs and unpublished reports, which are at extreme risk of loss over the medium and long term (Prom, 2011).

According to Walters and Skinner (2011), the responsibility for the custody and preservation of cultural heritage lies squarely upon the shoulders of librarians and archivists. This paper assesses the status of digital curation at Institutional repositories in selected Universities in Botswana and Kenya. An Institutional repository (IR) has been defined as a library of digital objects and associated metadata from a single institution (Clobridge, 2010)

The template is used to format your paper and style the text. All margins, column widths, line spaces, and text fonts are prescribed; please do not alter them. You may note peculiarities. This measurement and others are deliberate, using specifications that anticipate your paper as one part of the entire proceedings, and not as an independent document. Please do not revise any of the current designations.

A. *Preparing Your Paper (Heading 2)*

1) *Paper Size*: This template has been tailored for output on the A4 paper size.

2) *Type Sizes and Typefaces*: Follow the font type sizes in Table I. The font type sizes are given in MS Word font size points. Open Sans is the preferred font.

3) *Paper Margins*: Set top = 0.75” (19.05mm), bottom = 1” (25.4mm), side = 15,9mm. Each column measures 8,59 cm wide, with a 0,63 cm gap in between.

4) *Paper Styles*: Left- and right-justify the columns. On the last page of your paper, adjust the lengths of the columns so that they are equal. Use automatic hyphenation and check spelling and grammar.

Research Problem

Universities and other research organizations create and amass large volumes of digital assets and information which include administrative records, theses and dissertations, research publications, multimedia collections, digital surrogates of cultural material, learning objects, course materials, among others (Schmidt, Ghering and Nicholson 2011). Tindermans (2009) addressed the subject of digital preservation in the community pointing out that the huge volume of digital content, diverse variety of digital objects formats coupled with rapid technological changes that gave rise to an influx of new versions was a red flag that could not be ignored. Institutional repositories in many public universities in Africa such as Botswana and Kenya lack comprehensive, campus-wide digital preservation programmes or guidelines. Intentional digital preservation strategies are necessary in order to respond to the increase in digital content - especially in technology-dependent formats - and to provide prolonged access to digital records and archives. The goal of this research study is therefore to determine what is occurring in institutional repositories of selected universities in Botswana and Kenya with regard to digital records curation and to eventually propose a strategy that can be adopted by these institutions to support the long-term preservation and access of digital records and archives.

Research Objectives

The objectives of this paper are as follows:

1. To establish how IRs in selected universities acquire digital records
2. To evaluate the methods used to store digital records in the IRs
3. To investigate how digital records are managed in the IRs
4. To establish the preservation strategies for digital records in the IRs
5. To find out the procedures for access and use of digital records in the IRs
6. Propose recommendations to enhance digital curation practices in the IRs.

The Concept of Digital Curation

The term digital curation was first used in 2001 to refer to digital preservation, data curation, and the management of assets over their lifecycle (Yakel, 2007). Today, the term digital curation is increasingly being used for the actions needed to add value to and maintain these digital assets over time, for current and future generations of users (Beagrie, 2008). According to Yakel (2007), "Digital Curation is the active involvement of information professionals in the management, including the preservation, of digital data for future use". The Digital Curation Centre (2020) defines digital curation as "maintaining and adding value to digital research data for current and future use" and adds that "it encompasses the active management of data throughout the research lifecycle". According to Yakel (2007), "Digital Curation is the active involvement of information professionals in the management, including the preservation, of digital data for future use." Given the diversity of its stakeholders and of the environments in which it is conducted, digital curation potentially involves anyone who interacts with digital information during its lifecycle.

For purposes of this study, digital curation shall be defined as the active involvement in the management, including the preservation, of digital resources for future use. This intentionally broad definition is slightly adapted from Yakel. It omits the restriction to who is involved and uses the term "digital resources". Note that the focus on future use can be a very close or a very distant future. Ball (2010) defines digital curation by stating that digital curation in IRs must be seen and understood together with terms of preservation and archiving.

In almost all areas of society, but in particular in science, research, and scholarship, the ability to

effectively create, share and use digital resources has risen to form a crucial ability. The ability to manage these assets for current and future use is equally critical for a sustainable society. Institutional repositories play a crucial role in the preservation and making access to digital data and records through IRs. A study by Kakai, Musoke, and Okello-Obura (2018) found that libraries at Universities in East Africa were taking the lead in initiating and implementing IRs.

Models And Standards of Digital Archives Curation

In 2008, Higgins proposed a lifecycle in seven phases, namely the Digital Curation Centre (DCC) curation lifecycle model (2008), based on Pennock's (2007) lifecycle approach to digital curation. This lifecycle is composed of the following phases: create or receive; appraise and select; ingest; preservation action; store; access, use and re-use; and finally, transform, which links back to the first phase. According to Higgins:

"This lifecycle approach ensures that all the required stages are identified and planned, and necessary actions implemented, in the correct sequence. This can ensure the maintenance of authenticity, reliability, integrity and usability of digital material." (Higgins, 2008).

National Archives of Australia (2006) opines that Intellectual and physical management systems that are employed to store, manage, retrieve and deliver digital objects should, ideally, be based on open standards to ensure sustainability of the systems over time. Open standards exist for format types, for operating systems, disk drives and so on. If proprietary systems are used, digital objects could be lost or rendered uninterpretable over time. The Archives Domain is advocating that digital archiving solutions be based on open standards such as the Open Archival Information System (OAIS) Reference Model ('Blue Book' digital preservation framework – ISO 14721: 2003).

Theoretical Framework

There are different models that may be used in the management of records such as the Records Life-Cycle Model, the Records Continuum Model and the OAIS Model. However, this study uses the Open Archival Information System (OAIS) Functional Model

as a lens to investigate the status of the preservation and access of digital records by the public universities in Botswana and Kenya. The OAIS model categorizes the core set of tools with which an OAIS-type archive meets its primary mission of long-term preservation of information and access by the users (Digital Preservation Coalition 2015). Figure 1 depicts the OAIS model.

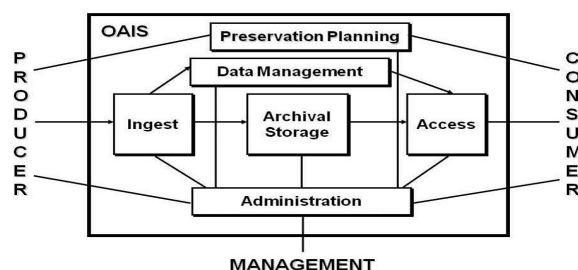


Figure 1 The OAIS Functional model. Source: Digital Preservation Coalition 2014)

The adoption of OAIS was purposely for its wide applicability for long-term preservation to any context, but principally in a digital environment, hence its relevance to the present study. The model is also suited for application in organizational and institutional set-ups such as public universities.

Methodology

This study used the quantitative approach whereby data was collected through questionnaires administered amongst university institutional repositories staff in Botswana and Kenya. The researchers desired only one response from each institution, preferably the staff in charge of the IRs. Online survey questionnaire was sent to four (4) public universities in Botswana and Kenya, giving a total of eight (8) questionnaires. The survey did not include private universities, colleges, or vocational training institutions. In total 8 responses were received. The data collected was analysed and presented in tables and figures.

Results And Discussions

The following section presents the results as per themes drawn from the research questions of this paper which are: acquisition of digital records in the IRs, management of digital records in the IRs, preservation strategies for digital records in the IRs, access and use of digital records in the IRs as well as recommendations to enhance digital curation practices at the IRs.

Ingest Of Digital Records

The ingest function as per the OAIS functional model relates to the receipt of information from sources, its packaging, acceptance of a Submission of Information Package (SIP), verification and the transfer of the created Archival Information Package (AIP) to the archival storage.

Types of digital records - Some of the digital records received by IRs include e-prints (both pre- and post-prints), grey literature (especially e-theses), working papers, technical reports, books and book chapters, conference papers, posters and administrative records (Jones, n.d). Respondents were asked to identify the types of digital records they received at their Institutional repositories. As shown in Figure 1, the majority of the respondents (42.9%) indicated that they receive thesis at their repositories followed by scientific research papers, artefacts, research publications, peer reviewed and published prints all at 14.3% each. Public universities generate large volumes of digital content emanating from three broad activities namely teaching, research and extension and outreach.

What types of digital records do you acquire in your repository?
7 responses

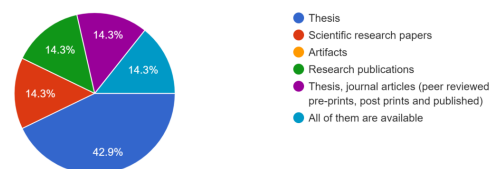


Figure 2 Types of Digital Records

Source of digital information - The study also sought to find out the source of digital records by IRs. Figure 3 shows that academic staff was the most cited source of digital records at 37%, other sources included students at 25%, administrative staff at 25% while the rest (publishers, postgraduate students and academic staff, students and administrative staff) stood at 12.5 % each. A study by Kakai (2018) revealed that lack of open access policies operating within institutions and lack of awareness of open access IRs among researchers and academicians were some of the factors that contribute towards limited acquisitions.

What are the sources of your digital records?
8 responses

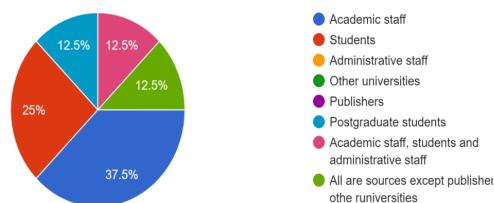


Figure 3 Sources of Digital Records

Best practice in digital archiving demands that archival repositories should formulate and implement collection development policies addressing the materials that the archives retain and what that which is not collected (Noonan and Chute 2014). The study findings indicated that majority of the institutions (75%) had collection development policies while two of the institutions were in the process of developing such documents.

Information Attached to the Digital Content

When asked about the essential information that should be attached to the digital content before acceptance into the IR, respondents answered as follows:

- R1: Plagiarism similarity check report and author consent form.
- R2: Delivery list from postgraduate or IR submission Form.
- R3: It's provenance.
- R4: For hardcopy publication you have to scan it to digitize and soft.
- R5: Item description (author, title, publisher, citation etc.).
- R6: Thirteen elements from Dublin Core metadata Standard.
- R7: ORCHID ID is critical

Digital records verification - OAIS functional model also requires that information be verified during the ingest function. Exlibris Knowledge Centre (2022) is of the view that the responsibility for the quality and accuracy of Institutional Repository content belongs to the source of data. The study respondents were asked to indicate how the information they received was verified. The R1 indicated that they have an office designated to repository administrator who is responsible for verification; R2 indicated that they have a Correction of Thesis form; R3 revealed that they do not verify information yet; R4 revealed that they use a Sherpa

Romeo; R5 said that they compare the information with the physical document. One respondent did not answer this question; R6 indicated that metadata is verified by the Repository manager before the content can be published while R7 said metadata is verified through the registry of researchers.

Packaging Of Digital Information

The study sought to find out how the digital information received was packaged. Two of the respondents did not answer this question; however, the rest of the respondents gave the following answers:

- R1 When the digital content is received at the office of repository administrator, its first run through "Turnitin" the anti-plagiarism software to verify the level of plagiarism whether it is within the University's accepted standard. Secondly, it's processed by classifying to determine the repository community and subject which is treated in the document. Finally, the record is entered to the IR.
- R2 Once uploaded, the work can be searched via author, subject, title etc.
- R3 We have not yet received digital information, only print.
- R4 For hardcopy publications you have to digitize by scanning.
- R5 Information is arranged into groups called communities which are subject-specific. In the case of [University X] the communities have similar names as university Faculties. So basically, the information is packaged according to faculties.
- R6 The repository is made up of different communities within the University.

Storage Of Digital Records by IRs

Archival Storage function is about the storage, maintenance, and retrieval of archival information packages (AIPs). When asked how they store digital records, the majority of the respondents (85.7%) indicated that they store their records on servers while only 14.3% indicated CDs and hard copies as depicted in Figure 4. None of the respondents indicated that they store their records on either clouds or servers. One respondent did not attempt this question.

How do you store your digital records?
7 responses

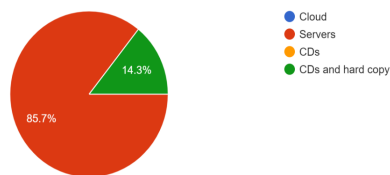


Figure 4: Storage of digital records

Management Of Data

Archival Storage function is about the storage, maintenance, and retrieval of archival information packages (AIPs). It accepts AIPs submitted from the ingest function, assigns them to long term storage, migrates AIPs as needed, checks for errors, and provides requested AIPs to the Access function. Some University IRs in this study stated that they use the DSpace software; however, Kakai et al (2018) argue that software is not easy to install and maintain.

Query requests - The respondents were asked to state the procedure for executing query requests and generating results. The responses were as follows:

- R1 The users have been assigned to a specific email that receive and answer users' questions.
- R2 Via manual or online request through email and the same for results.
- R3 Not yet applicable [IR not yet established]
- R4 Searching using Author, title and subject.
- R5 Contact the IR Manager.
- R6 The D-Space platform sends emails to administrators and if there is a query then the admins will address it.
- R7 DSpace's JSPUI. The JSPUI defines several filters, listeners and servlets to process a request.

Reports generation - The respondents indicated that they generated the following reports: deposits reports and entries, usage statistics (downloads, views), most popular items and authors, content statistics, search statistics, storage statistics, statistics by country, items added in a given certain of time, total items in a repository

Preservation Planning

Preservation planning function supports all activities meant for long term preservation and accessibility of digital records.

Preservation tasks - Respondents were asked to state whether they had preservation or migration plans in place and the majority of them (71.4%) indicated that they do have preservation plans, while 28.6% indicated that they do not have preservation plans as depicted in Figure 5.

Preservation Tasks Do you have detailed preservation/migration plans?
7 responses

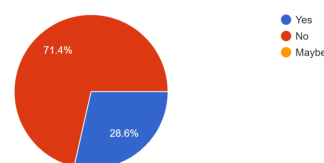


Figure 5 Preservation Tasks

Evaluation and risk analysis of content - When asked how often they do evaluation and risk analysis of content, 80% of the respondents indicated that they do not do any risk analysis of content or they don't know about it, while only 20% indicated that they do it daily and quarterly.

Access And Use of Digital Records

The Access function relates to the user interface that allows users to retrieve information from the archive on request. Kakai, Musoke and Okello-Obura (2018) argue that in the digital environment, library users are interested in easily accessing full-text information resources, and these should be readily available from IRs. Respondents in this study were asked to comment on how user- friendly their interface was based on a scale of 1-5 where 1 was fairly friendly and 5 very friendly. The majority of them (50%) chose 5, 33.3 chose 2 while 16.7 chose 1 as depicted by Figure 6.

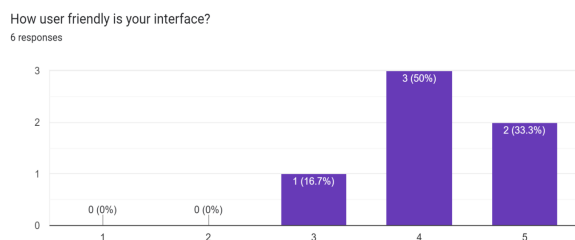


Figure 6 User friendliness of the interface

To explain further responses in on the user friendliness of the IRs interfaces as presented in Figure 5, the respondents had this to say:

R1 The terminologies used are common English and can easily be understood by anybody who understands English.

R2 It easy to retrieve a document by author or title.

R3 Not applicable [Yet to establish an IR].

R4 It is easy to navigate.

R5 It provides greater visibility and accessibility at all times.

R6 It is easy to navigate since there are no pictures that can distract the user, less customization.

R7 All features are clear.

Recommendations To Enhance Digital Curation Practices

Based on the findings from the data collected, this paper presents the following recommendations:

Expansion of digital information content that is received by IRs to other materials such as digital archival materials and photographs

Universities should diversify their sources of digital records to ensure that more information is preserved and to comply with the legal deposit legislation. Other sources which are not target for the IRs surveyed in this study include other universities and publishing houses.

Cloud storage is recommended with its capacity to improve access to sharing of information and its preservation.

Implementation of the OAIS model for digital records preservation by IRs as it promotes long term preservation of digital records and may allow of interoperability with other IRs.

Conclusion

This study has established that some universities in Botswana and Kenya do have IRs that ingest digital records into their custody though the resources are mostly limited to thesis and publications by academic staff and students. The study also found that the storage of digital records by IRs in Botswana and Kenya is mostly on local servers and CDs. Furthermore, the majority of the IRs surveyed indicated that they do not have preservation plans for their digital records. As digital records are increasingly being generated, IRs would play a crucial role in the preservation of digital records in the Southern and Eastern countries such as Botswana and Kenya. The current survey may serve as the basis for bigger research to include more IRs in Southern and Eastern Africa.

1. REFERENCES

- [1] Anderson, J. and Rainie, L. (2012). The future of the Internet. Washington, DC: Pew Research Center.
- [2] Ball, A. 2010. Review of the state of the art of the digital curation of research data. Bath, UK: University of Bath.
- [3] Beagrie, N. (2008). Digital curation for science, digital libraries, and individuals. *International Journal of Digital Curation*, 1(1), 3-16.
- [4] Clobridge, A. 2010. Introduction in Ruth Rikowski. *Building a Digital Repository Program with Limited Resources*. Chandos Publishing (Oxford) Limited. UK.
- [5] Digital Curation Centre (DCC). 2020. What is digital curation? Available at: <http://www.dcc.ac.uk/about-us> [Accessed 28 February 2020].
- [6] Digital Curation Centre. 2008. DCC curation lifecycle model. Available at: <http://www.dcc.ac.uk/docs/publications/DCCLifecycle.pdf> [Accessed 7 February 2019].
- [7] Digital Preservation Coalition. 2015. *Digital preservation handbook*. 2nd ed. Glasgow: University of Glasgow.
- [8] Exlibris Knowledge Centre. *Institutional Repositories in CDI*. Available online at https://knowledge.exlibrisgroup.com/Primo/Content_Corner/Central_Discovery_Index/Knowledge_Articles/Institutional_Repositories_in_CDI (Accessed 22 January 2023).
- [9] Harvey, R. (2010). *Digital Curation: A How to Do It Manual*. New York: Neal Schuman.
- [10] Hedstrom, M. (1998), "Digital preservation: a time bomb for digital libraries". *Computers and the Humanities* Vol. 31: 189-202.
- [11] Higgins, S. 2008. The Digital Curation Centre Lifecycle Model. *The International Journal of Digital Curation* 3(1): 134-140. Available at: <http://www.ijdc.net/article/download/69/48/0> [Accessed 24 March 2019].
- [12] Kakai, M., Musoke, M.G.N and Okello-Obura, C. 2018. Open access institutional repositories in universities in East Africa. *Information and Learning Science*. Vol. 119 No. 11, 2018 pp. 667-681.
- [13] National Archives of Australia. 2006. *Designing Information and Recordkeeping Systems (DIRKS): a strategic approach to managing business information*. Canberra: National Archives of Australia.

- [15] Noonan, T. and Chute, D. 2014. Data Curation and the University Archives. *The American Archivist*, Vol. 77, No. 1 (SPRING/SUMMER 2014), pp. 201-240.
- [16] Pennock, M. 2007. Digital curation: a life-cycle approach to managing and preserving usable digital information. *Journal of Library and Archives* 1: 12-24.
- [17] Prom, C. 2011. Making Digital Curation a Systematic Institutional Function the *International Journal of Digital Curation*. Vol 6(1): 139-152.
- [18] Schmidt, L., Ghering, C. and Nicholas, S. 2011. Digital curation planning at Michigan State University. *Association of Library and Technical Services* 55(22): 104-118.
- [19] The Open Archival Information System (OAIS) Reference Model: Introductory Guide. 2014. The Digital Preservation Coalition. Great Britain (2nd Edition).
- [20] Walters, T. and Skinner, K. 2011. *Digital curation for preservation*. Washington, DC: Association of Research Libraries.
- [21] Yake, E. (2007). Digital curation. *OCLS Systems and Services* Vol. 23 No. 4, pp. 338-339.