

University of Nebraska - Lincoln

DigitalCommons@University of Nebraska - Lincoln

Library Philosophy and Practice (e-journal)

Libraries at University of Nebraska-Lincoln

2022

Evaluating Institutional Repositories' (IR) capabilities for long-term preservation with a focus on content, file format and metadata practices in selected public university libraries in Kenya

Hellen M. Ndegwa

PHD student Moi University, Eldoret, ndegwahellen8@gmail.com

Emily Bosire Dr.

Moi University, Eldoret, emilykwamboka@gmail.com

Damaris Odero Dr.

Moi University, Eldoret, oderodjn@gmail.com

Follow this and additional works at: <https://digitalcommons.unl.edu/libphilprac>



Part of the [Scholarly Communication Commons](#)

Ndegwa, Hellen M.; Bosire, Emily Dr.; and Odero, Damaris Dr., "Evaluating Institutional Repositories' (IR) capabilities for long-term preservation with a focus on content, file format and metadata practices in selected public university libraries in Kenya" (2022). *Library Philosophy and Practice (e-journal)*. 7114. <https://digitalcommons.unl.edu/libphilprac/7114>

Evaluating Institutional Repositories' (IR) capabilities for long-term preservation with a focus on content, file format and metadata practices in selected public university libraries in Kenya

Abstract

The type of content and file format influences the success of digital preservation strategies. Institutional repositories are custodians of digital resources that are to be held in perpetuity necessitating the need to consider long term preservation of these resources. The purpose of this study was to evaluate the suitability of digital content and its metadata for long term digital preservation. The study was qualitative in nature utilizing interviews as well as document analysis. websites and IR database investigations were utilized to check on content, format and metadata adequacy. The findings revealed great inadequacies in the IRs' capabilities to support long term preservation as evidenced by haphazard content and format selection, ingest procedures that did not consider long-term preservation as well as metadata that focused on access only. Recommendations included the need to involve archivist in develop selection and appraisal policies as well as development comprehensive metadata policies that ensured that preservation metadata was also captured as required. Creation of awareness among repository administrators to expose them to the importance of adopting open file formats and standard as well as benchmarking were also proposed. The paper provides insights into universities on the relationship between selection and processing of digital resources and their long-term preservation within the IRs in Kenya.

Keywords: Institutional repositories; University Libraries-Kenya, metadata, Institutional repository contents, File Formats, digital preservation

Introduction

According to Robertson & Borchert (2014) Institutional Repositories (IRs) are intended to be long-term homes for intellectual output from a college or university. At the early stages, researchers like Hockx-Yu (2006) tried to exonerate the IRs from their poor uptake of digital preservation by arguing that they had been in existence for such a short time and therefore had not yet experienced any preservation challenges with their digital resources. The Center for Research Libraries' (CRL), (2007) continued to justify this poor uptake by proposing that IRs lacked the basic characteristics that define digital preservation repositories since they instead focused on access to content. Hockx-Yu (2006) goes on to add that there were divided views between those who support digital preservation as a function of the repository and those who felt that the IRs ought to concentrate on improving access, usage and impact and leave out long-term preservation. These arguments do not negate the need for digital preservation within the IRs as it all depends on the goal of the institutional repository and the content it contains a view supported by Thibodaux (2007) who proposed that the criteria for success of an IR must be derived from its statement of purpose. By this he meant that although institutional repositories are established across the industry, their success can only be measured by what they are meant to

do and the environment in which they exist. Whether they practice digital preservation for long term or short term will therefore be established from their mission statement.

The need for digital preservation was clearly explained by Hockx-Yu (2006, p. 235) in the statement that “Researchers, students, staff and institutions will require ongoing availability and confidence in the future accessibility of the content within the repositories”. This therefore places a great responsibility to IR administrators who have to ensure that the resources are available to them for as long as needed which could be beyond decades. Knight (2005) proposed that the academic community ought to consider the need for digital preservation, to ensure that academic research in the form of e-prints and e-theses, deposited within the repositories remained accessible and offered a guarantee of integrity in the long-term. Despite this responsibility most IRs have been slow to adopt digital preservation as shown by Lavoie & Dempsey (2004) who outlined the prerequisites of digital information environments and noted that of all the prerequisites such as predictability and comprehensiveness, interoperability, transaction ability and preservability, there has been a slow uptake of preservation but on a positive note adds that a lot of research is being done on it.

ISO 14721 (2003) defines the minimum requirements that a digital archive should be able to fulfil in order to support digital preservation. Some of these minimum requirements relate to ensuring that resources accepted in the repository meet the requirements such as content that meets user needs, the right file format and enough preservation metadata (ISO 14721, 2003). In addition, researchers such as Najar and Wani (2019) have come out strongly to decry the need for the consideration of file formats in any digital preservation strategy because as they are being updated, to include new trends and features the previous ones become obsolete. The Digital Preservation Coalition (2015) cites obsolescence and proliferation as the two major reasons why organisations need to consider file formats during preservation planning. File formats become obsolete when the software that renders them is no longer supported.

A number of studies have linked content with digital preservation with Smith (2008) recommending a needs assessment to ensure that the content ingested in the IR meets the needs of the designated community. The need for selection is based on the existence of an enormous quantity of information being produced digitally, with variable quality, and the resource constraints on those taking responsibility to preserve long-term access. The Digital Preservation Coalition (2008), supported by Lunghi et al (2012) emphasized the need to select and appraise digital resources a view affirmed by Ismail & Affandy (2018) who put forward the claim that the resource constraints on those taking responsibility to preserve long-term access, makes selectivity inevitable if the objective is to preserve for ongoing access.

The role of preservation metadata in digital preservation cannot be disputed. Besser (2000) explains preservation metadata as a strategy that aims at providing sufficient technical information about the resources and supports migration and emulation as preservation strategies. By describing the technical environment of the resource, it can be migrated to newer versions of

hardware, software, storage media or file format effectively or the previous environment can be imitated in cases of emulation. This as well articulated by Woodyard (2004) who pointed out that preservation metadata should be able to: list the technical details about files and structure of the resource and how to use it; record the history of all actions performed on the resource, including any changes or decisions made about it; prove the authenticity through technical means and account for the continued custody of the resource and retain information on who has the responsibility and rights to perform preservation actions on the resource.

Cendi (2006) articulated it very well by declaring that all organisations needed to document the purposes and requirements including the purpose and requirements of preservation metadata a view that Jones (2006) agreed with and went on to propose not only the need for substantive descriptive metadata to support access but also technical metadata to aid preservation.

In this article, characteristics of resources in three major universities and their relationships to digital preservation are described.

STATEMENT OF THE PROBLEM

Institutional repositories, an innovation of libraries are fast becoming one of the best avenues utilized by institutions in making their research knowledge or output widely available and accessible to the outside world via the World Wide Web. Digital preservation aims at ensuring that digital content within the repositories remains accessible to user communities for a long period of time for future generations. Establishing the effectiveness of digital preservation strategies in use is critical to its success. Institutional repositories in Kenya have adopted several strategies such as adoption of policies, offsite storage and backups to ensure that the digital resources in their custody are availed to the future generations (Erima, Masai & Wosyanju, 2016; Moseti, 2016). Unfortunately, studies are yet to focus on the effectiveness of selection and file format decisions and practices as well as metadata creation practices in digital preservation. Adams, (2018) decries the lack of literature on the success of digital preservation practices and opines that without knowing the efficacy of digital preservation efforts it was impossible to plan for it. Tieman (2015) argued that research funders, depositors and other stakeholders need evidence that the repository is worthy of trust. According to Maemura et al (2017) there is need to assess an organization's abilities to achieve its digital preservation goals with Shajitha and Abdul (2021) proposing frequent evaluations in institutional repositories to identify and bridge any gaps identified. Donaldson (2020) proposed the use of a systematic and independent audit to determine the details of the process and identify potential weak points in order to make improvements. Frank (2018) believes it is important to understand whether the repositories entrusted with valuable digital information are trustworthy because the content that they are responsible for preserving includes valuable and sometimes unique resources. A 2018 survey of the National Digital Stewardship Alliance (NDSA) member institutions noted that the evaluation of digital preservation practices was the greatest challenge facing the community of digital stewards with few of them performing the task (Altman et al, 2019).

It is from this background that this study sought find out whether, the content selection practices, file format decisions as well as metadata creation decisions were effective enough to support long term digital preservation in selected university libraries in Kenya.

Objectives of the Study

The broad objective of this study was to evaluate content selection practices, file format decisions and practices as well as metadata practices for effective digital preservation. Specific objectives to:

- i. Establish the type of content is accepted in the institutional repositories
- ii. Assess the adequacy of content selection and acquisition policies.
- iii. Determine the type of file formats are accepted in the institutional repositories
- iv. Establish the existence of guidelines for depositors
- v. Assess the type and adequacy of metadata created

Research Questions

- i. What type of content is accepted in the institutional repositories?
- ii. DO the IRs have content selection and acquisition policies?
- iii. What file formats are accepted in the institutional repositories?
- iv. Are there guidelines for depositors?
- v. What type of metadata is created and how adequate is it?

Significance of the Study

It is hoped that the outcome of this research will provide insight into the importance of content and file format considerations as well as importance of metadata for longer-term preservation of information materials in IRs in Kenya

Literature review

A number of studies have tried to link characteristics of information materials with successful digital preservation. This paper has reviewed some of these studies in order to anchor itself with the existing body of knowledge. The review has been taken along the three major themes of the paper: content, file formats and the role of preservation metadata in successful of digital preservation.

Content in Institutional Repositories

Content in institutional repositories is selected to reflect their characteristics. Johnson (2002) describes these characteristics as: Scholarly, cumulative and perpetual, open and interoperable. The first attribute means that the contents of the repository are limited to the output of one institution (Genoni, 2004) while the third stipulates the need for their availability and usability in the long term. Digital repositories can store various types of content and file formats. Several

researchers have described the contents of a digital repository as peer reviewed journal articles, thesis and dissertations, research data, monographs and book chapters, conference proceedings, departmental and research center newsletters and bulletins, papers in support of grant applications and status reports to funding agencies (Waddington et al, 2013; Genoni, 2004). Shearer (2006) established that repositories contained a wide range of materials and went further to explain that “being scholarly” did not exclude a repository from collecting other types of materials such as university annual reports, video recordings, computer programs, data sets and photographs. Breytenbach, Lourens and Marsh (2013) had proposed that since IRs stood for permanence and accessibility of information, they were a better alternative to informally distribute the information and therefore go beyond the scholarly output to include unpublished conference papers, teaching and learning resources, unpublished research material and corporate material such as publicity material of an institution.

The inclusion of peer reviewed journal articles dates back to the establishment of the Open Society in 2002 which in its declaration stated that: “Open Access to peer-reviewed journal literature is the goal”, (Budapest Open Access Initiative, 2001) giving way to researchers to publish their research findings using either the green open access where authors self-archive their research output in digital repositories and gold open access journals where authors or their sponsors pay for their research to be published in open access journals . In most of the institutional repositories these articles include the author’s final version without the publisher’s editing style or those that have passed the embargo period. (Jones & Benson, 2016).

Research data refers to data underlying publications and/or raw data. Research data has found its way in institutional repositories as research funders especially in the United Kingdom (UK) are requiring their grant-holders to make their data Open Access, once they have themselves analyzed and published their findings from the data (Pryor and Donnelley, 2009). The UK’s Economic and Social Research Council (ESRC), for example, requires its grant holders to make their data available for reuse (unless there are convincing reasons for not doing so), and attaches financial penalties to non-compliance (Mauthner, 2013). The benefits of advancing data curation by the institutional repositories have been advanced by a number of researchers. These include allowing other researchers to use the data to verify results, to compare with their own data or to re-use (Heidorn, 2008) and in some way to generate new data and knowledge, (Krotoski, 2012).

According to Suber (2012) books are not common as content for open access since they are often written for monetary gain (royalties on sales) and authors are reluctant to deposit them for free in a repository.

In addition to the types of content described above, institutional repositories frequently contain thesis and dissertations generated by post graduate students as part of the requirements for their degree programs and other research-related outputs such as post graduate seminar presentations and speeches. As explained by Ahmed, Alreyaee & Rahman, (2014), these form an important

part of research work in an academic institution and the need to provide long-term access to them is paramount.

Researchers have proposed a multi-stakeholder involvement of especially archivists in the selection of digital repository content as they have expertise in appraisal that ensures the identification of resources that require long term preservation (Smith, 2008; Dell & Shultz, 2014). Keller, Robertson, Steinle, & Thibault (2019) in their report “Digital Preservation Task Force Update”, established that although archival expertise had been identified as critical success skills in the establishment of IRs, very few cases involved the archivist and in cases where they were involved, their expertise was visible in the scope and content of the institutional repository, adoption of preservation standards and the inclusion of structural and technical metadata for preservation purposes.

It is important to note the relationship of content to planning as summed up by Webb, Pearson & Koerbin (2013) who pointed out that when organisations clarified their preservation intentions it was a likely good starting point for preservation planning for diverse digital collections as it was adept to identify what needs to be kept and what does not warrant the use of limited preservation resources.

Content is submitted to institutional repositories in a number of ways. A report for Decker Library, Maryland Institute College of Art by Rodríguez (2019) identified self-archiving directly to the repository platform by the creators, DVDs and CDs. These are the most common methods but in some cases links to the resources are given by the creators whereupon the repository administrator downloads the content and uploads to the repository platform. The media used to submit the resources also requires poses its own preservation challenges. IRs use the CD-ROM and the DVDs as backups for their documents but lack disaster protection plans for them (Mensah & Amoah, 2019). Nadal (2007) recommended that there was need to ensure that these storage media were secure and reliable as they were known to be fragile and unstable.

File formats and digital preservation

Brown (2006) defined a file format as ‘the internal structure and encoding of a digital object, which allows it to be processed, or to be rendered in human accessible form. A digital object may be a file, or a bit stream embedded within a file’. According to Barve (2007) digital preservation aims at ensuring the accessibility of a digital object throughout time but faces challenges due to technological obsolescence that makes old file formats to become unreadable and unusable. There are different file formats for different applications such as text files, audio files, video files, image files, database files, presentation files, spreadsheet files and markup languages. Selection of file formats and preservation media is very important to digital preservation. According to Lundell (2012) file formats used for digital preservation should be independent of the hardware and software since the information encoded in them normally outlive the hardware and software and recommends the use of open file formats. JISC (2008) advocates for the selection of file formats to take into account longevity, protection, and

preservation. Rimkus, Padilla, Popp and Martin (2014) recommended the establishment of file format policies to assist the repository managers. Francke, Gamalielsson, and Lundell (2017) in their study found that majority of the repositories under their study in Sweden had some form of guidelines for which file formats were accepted, but very few considered whether or not file formats constitute open standards. Their study also established that PDF files were often encrypted to prevent manipulation and this was seen as a potential hindrance for future file migration which at one time or another has to be done for successful long-term preservation.

Some file formats have been recognized by standardization organizations (and published as standards), whereas other formats are maintained by specific companies. According to the Property Records Industry Association (PRIA) (2017) PDF was developed by Adobe Systems Incorporated but was released to the International Organization of Standardization (ISO) as an open standard. It was from this that PDF/A (archival) was developed. ISO describes PDF/A as a file format that provides a mechanism for representing digital objects so that their visual appearance is independent of the tools and systems used for creating, storing or rendering the files, while TIFF was created by the Aldus Corporation in 1986 for use in desktop publishing with the current TIFF 6.0 developed in 1992 and its right acquired by Adobe Systems Incorporated in 1994.

Role of Metadata in digital preservation

Preservation metadata refers to information about digital objects that is necessary for their long-term accessibility. Due to the dependence of digital resources on machines for renderability, and the rate at which technology becomes obsolete, it is important to preserve the resource together with some information on the hardware, software, file formats and storage media they were previously created in.

According to Arora (2006) digital formats depend on specific computing environments and a deviation from that environment results in a change of the rendering presentation of the resource. This argument gives weight to the need to keep the computing environment as close as possible to the original one to maintain the reliability and integrity of the resource. This is what give rise to the need for preservation metadata,

Woodyard (2004) recommends preservation metadata that should be able to: list the technical details about files and structure of the resource and how to use it; record the history of all actions performed on the resource, including any changes or decisions made about it; prove the authenticity through technical means and account for the continued custody of the resource and retain information on who has the responsibility and rights to perform preservation actions on the resource.

Cendi (2006) was of the view that all organisations needed to document the purposes and requirements including the purpose and requirements of preservation metadata, an idea that

Jones (2006) agreed with and went on to propose not only the need for substantive descriptive metadata to support access but also technical metadata to aid preservation.

OAIS defines a group of types of metadata under the name of Preservation Description Information (PDI) which is broken down into reference, provenance, context, fixity and rights metadata (Giaretta, 2011). The importance on metadata in digital preservation has been spelt out by a number of scholars. The OCLC/RLG (2001) pointed out that the creation and deployment of preservation metadata will be one of the key components of most digital preservation strategies. In fact, Gartner & Lavoie (2013) concluded that preservation metadata was one of the best practices of any long term stewardship of digital resource.

Methodology

This study aimed at evaluating content selection, file format decisions as well as metadata practices in selected public university IRS in Kenya. The study was qualitative in nature utilizing the case study design and incorporating the evaluative research aspect. Three Universities were purposively selected on the basis of the length of time they had been in existence; how far they had been able to develop their institutional repositories infrastructure judged by a number of factors such as: consistent top webometrics ranking by 2016; amount of content was indicated in their repository websites (over 3500 items); registration with Open DOAR and ROARMAP. The criteria for age was based on the assumption that the universities had well-established postgraduate and research programs implied a comparatively well-established research infrastructure and high quantity of research output that has or was in the process of being digitized that may necessitate the need for active digital preservation.

The library and ICT departments were also chosen purposively owing to the fact that the library department is responsible for the management of the institutional repository and that the ICT department is involved in one way or another in the management of the technical aspects of the IRs. Within these departments, purposive sampling was used to select repository administrators, the senior library management responsible for policy development and implementation. These comprised the university librarian or director and the deputy university librarians. In ICT department, the technical person attached to the library was identified.

A total of 19 (nineteen) respondents were involved in this study, Seven from UoN and six from each of the other universities under study. Interview guides were used to collect data from the respondents with document analysis used to collect primary data in order to support triangulation where by institutional repository policies, procedures and the website were analyzed to collaborate data gathered through the interviews. Observation to collaborate interview answers on metadata creation and was carried out as the respondents were creating metadata. Data was analyzed conventional content analysis where codes were derived directly from the text data collected.

Findings and Discussion

The research sought to evaluate selection of content, file format as well as metadata practices in the institutional repositories of selected universities in Kenya with a view to establishing their suitability to support successful long-term digital preservation.

Content in the institutional repositories under study

The study found that most of the content in the IRs under study were similar and only in a few cases was content unique to a particular university. Some of the common digital resources included Books, Conference /Workshop / Seminar Papers, Theses and dissertations, Policies/ Reports/ Newsletters, Public Lectures and Speeches, and journal articles, graduation resources (video and lists). Those that were unique included multimedia and undergraduate projects that were found at Kenyatta university and Jomo Kenyatta University Science and Technology while UoN had a collection called archives that had microform thesis, pictorial collection and rare collections. These finding mirrored those of Vrana, (2017) who found that digital repositories can store types of e-prints of scientific papers, research data and also e-learning materials and other forms of institutional intellectual outputs. They also collaborated with other researchers who had concluded that despite their being scholarly repositories, the Institutional Repositories were not excluded from collecting other types of materials such as university annual reports, video recordings, computer programs, data sets, photographs and even publicity materials, (Breytenbach, Lourens & Marsh, 2013; Shearer, 2006).

The study established that in all the universities under study, thesis and dissertations had the largest content distribution as follows: University of Nairobi (36143), Kenyatta university (7925) and JKUAT (1884). This was followed by research articles with University of Nairobi having (2297), Kenyatta university (3709) and JKUAT (1329) respectively. At KU, thesis and dissertations made 60% of the total content followed by research papers at 37%, with the other content took up the remaining 3%. In UoN, thesis and dissertations formed 43%, journal articles 35%, conference proceedings and seminar papers 9% and the archives collection 10%. At JKUAT, thesis and dissertations formed 52% of the collection while research papers, conference proceedings were 39%. From the above findings it can be concluded that majority of the digital resources stored in the Institutional Repositories are primary information sources that are very useful in furthering research. The large concentration of thesis, dissertations and research papers in all the repositories is an indication that the institutions under study had a commitment to preserve the digital resources to ensure that future generations had access to this research to support the research process. Previous studies have drawn attention to the problem of losing digital content in open access repositories due to poor preservation of content over time and pinpointed the necessity of preservation for the protection of open access content (Pinfield & James, 2003).

At the time of the research the largest percentage of the content (Thesis and Dissertations) had an analog equivalent mainly because it was mandatory that students presented both a soft copy and a hard copy. This may have had a great impact on the organisations view of digital preservation

because the respondents felt that since a hard copy existed somewhere then it was of not necessary to invest in expensive preservation activities as they could digitize again using the print copy if anything happened to the digital copy. This notion did not take into consideration research articles that were born digital with no analog equivalent and were in danger if digital preservation planning was not considered. These findings were attributed to a lack of awareness of the fragility of digital resources brought about by their machine dependency and technological obsolescence. As explained by Williamson, (2004) who citing Feeney (1999) a general lack of awareness on the importance of active digital preservation resulted in the misuse of public funds as a lot of money was wasted on digitization projects undertaken without due regard to the long-term preservation because the maintenance of the digital files created became too expensive for the repositories concerned or the digitization process was repeated again in future.

Procedures for content selection

The study established that procedures for content selection were broadly defined. The open access policies of the universities under study had a section on acquisition which clearly stipulated the types of resources that were to be accepted by the repositories.

Each university had a list of the type of content to be accepted in the IR as shown in the table below.

Table 1: Types of materials by content in received in IRs of selected Universities

UoN	JKUAT	KU
Theses and dissertations/research projects Scholarly Research articles (published peer reviewed and pre-prints) Open lectures Conference/workshop Proceedings; Books, monographs, chapters Image collections (paintings, pictures, drawings, illustrations, etc). Audio and audio-visual materials Technical reports and working papers Inaugural lectures, distinguished lectures, speeches Datasets Admissions lists	Journal articles Books Book chapters Conference publications Refereed designs, Creative, performance-based and visual arts outputs that have research components Masters and doctoral theses Unpublished scholarly work	Journal articles Theses and Dissertations. Learning Objects (past papers, lecture notes and presentations). Conference and Workshop Proceedings. Books and book chapters. Technical reports, commissioned reports, and other un-refereed research outputs. Newsletters of significant research groups. Other materials produced by academic/research staff and approved by Deputy Vice Chancellor, (Research, Production and

Graduation lists University policies Events programs Valedictory presentations University calendars University magazines Forms Newsletters Literary publications Journalism student media content Other materials as maybe approved by the Senate and the Vice-Chancellor		Extension), Directors of, Schools/Deans of Faculties or Heads of, Departments.
--	--	--

Despite having well defined content procedures materials not listed such as charters and graduation booklets as well as undergraduate projects were still accepted into the repositories. Appraisal was also not done nor was the content categorized in a way that allowed those that requiring long term preservation to be identified. Genoni (2004) had pointed out that this was a major challenge in the development of repositories as little consideration was given to the content accepted and went further to recommend that a vibrant acquisition policy was necessary in order to influence long-term preservation of the digital resources. He also went on recommend the frequent evaluation of the collection to enable libraries to make conservation and preservation decisions.

In one case, the study found that there was a pictorial collection consisting of photographs of university management and other dignitaries scanned and saved in PDF. This collection had been discontinued as the repository management argued that it was difficult to keep up with the changes in management. Although no new materials were being received in this collection, no decision had been made on what to do about what was already deposited in the repository except to make it inaccessible to the public. These instances revealed weaknesses due to inadequate selection policies to long term preservation. Previous studies support the use of experts to select resources that are ingested in the repository and also propose the identification of those that will be kept for long term (Smith, 2008; Lunghi et al, 2012; Dell & Shultz, 2014) with Ismail and Affandy (2018) going on to add that selection is very necessary based on the resource constraints associated with digital preservation.

The findings also revealed that procedures for withdrawal of content from the collection was well articulated in the open access polices as indicated by the following statements picked from the policies: “Items will be preserved indefinitely; Items may only be removed from the repository due to: Proven copyright violation or plagiarism; Legal requirements and proven

violations; National Security; Falsified research; and Request by author” the policy goes on to add that “withdrawn items will not be deleted but will be removed from public view” and that “Withdrawn items' identifiers/URLs will be retained indefinitely”. These procedures if well adhered to place the responsibility to the IR to put in place quality control measures to ensure that what was ingested into the repositories will not be withdrawn at any particular time.

Types of formats accepted in the institutional repositories

On what type of formats were accepted in the repository, all the respondents stated that the Portable document format (PDF) was the preferred format for text although in some instances other text formats were accepted while Jpeg was considered for images. Although there were no audio files in the repositories under study, the open access policies in the universities under study indicated that mp3 file format the was preferred format while for video the mp4 was chosen.

A look at the file formats in the collection revealed that it was true the PDF file format was the most dominant which is an indication that over 97% of the resources in the repositories were textual. The JKUAT repository had a collection of images/ photographs of graduation ceremonies, but the links were broken. JKUAT (2 files) and KU (1 file) repositories accepted video files with UoN not accepting any. These findings resonate with Robertson and Borchert (2014) who in their paper “Preserving Content from Your Institutional Repository” noted that while Portable Document Formats (PDFs) were the most common formats, other formats such as, born digital text, audio, video, images, or multimedia were used.

Despite having identified the file formats preferred by the repositories, there was no information identifying the various distinctions within the same file format families. This is against Barve (2007) recommendations that any digital repositories needed to keep detailed knowledge of the internal properties of digital formats as it was necessary to interpret properly the full information content of digital objects for digital preservation purposes.

On whether consideration was given in the choice of formats that promoted preservation for long-term access and use most respondents indicated that they had not factored in the implications of file formats as evidenced by a number of respondents who said that “I don’t think file format matters at all. After all, through time I have been able to open and use my files without any problems”. This was an indication that even those in charge of the repository do not have basic knowledge of the relationship between file format and long-term preservation which translates to poor consideration for file formats to be utilized in the repository. This collaborated with Francke, Gamalielsson, and Lundell (2017) who in their study found that although majority of the repositories in Sweden had some form of guidelines for which file formats were accepted, very few considered whether or not the file formats constituted open standards. It is also because of this challenge that pictorial collections were stored in PDF format leading to loss of formatting and leading to poor quality.

Although all the respondents said that they had clearly defined boundaries for formats, a look at their open access repositories showed that only JKUAT formally did so. The UoN repository policy described a broad criterion for format selection leaving it open for any file format that met the criteria to be used. The statement in the policy read as follows “File formats that are platform-independent, vendor-independent, non-proprietary, stable, widely supported are recommended. The Repository did not accept executable binary files if alternatives are available and UoN Digital Repository staff may convert to more appropriate formats any content that is in obscure or little-used formats for compatibility reasons”. Kenyatta University repository’s policy does not mention anything about file formats.

The above scenarios were against the recommendations of Rimkus, Padilla, Popp and Martin (2014) who had recommended the establishment of file format policies to assist the repository managers to identify file formats that would support long term preservation of digital objects.

The table below shows the formats as articulated by the JKUAT policy

Table 2: File formats Accepted in the IRs

<ul style="list-style-type: none">• Adobe® PDF (.pdf)• Audio and video file formats (. aiff, aif., aifc, .tiff, .jpeg, .gif)• Microsoft Office Excel® (.xls)• Machine-Readable Catalogue Records - MARC• Microsoft Office Excel® (.xls)• Microsoft Office Powerpoint® (.ppt)• Microsoft Office Word® (.doc, .docx)• Moving Picture Experts Group (.mpeg, .mpg)• Text file Formats (HTML, TXT (text), DAT (data: ASCII data), RTF (rich text format), and XML]

Preservation experts recommend that institutions limit themselves to fewer formats that can be preserved and made accessible over time to avoid obsolesce. At the same time, they advocate for open file formats that are stable (Arp, 2019). JKUAT has clearly defined boundaries although there was no clear definition of which Microsoft files formats were adopted. The 93-2007 file formats are no longer supported by Microsoft and their specifications have been released to the public and therefore would be suitable for preservation since they are widely used. It is important that the institutions in future take into consideration the file formats since they are the vessels that the information is encoded in.

The results point to weak file format defining policies that weaken any digital preservation efforts.

Procedures for receiving (Ingesting) materials in the IR

All the repositories under study had some guidelines on how resources were received in the repository but only textual information was covered. Audio visual and multimedia was not catered for. The resources were presented to the repository administrators who after checking for quality uploaded the items. Although all the respondents agreed that they had procedures on how the digital resources were received and processed, only the UoN was in the process of preparing a guide for self- archiving. This is a serious oversight for digital preservation as consistency and quality management should be supported by documentation in cases where new staff are being used to do the work. This reduced the organisation viability in developing trusted repositories that have digital preservation at the centre of the IRs activities.

Table 3: Guidelines for Ingest as reported Verbatim

University	Type of resource	Guidelines for ingest
University of Nairobi	Thesis	A well labeled CD with both a word and PDF copy of the work
	Journal articles	Should indicate whether it is wholly open access or it is still on embargo. The person depositing should be the owner of the intellectual content
	Others	Ownership File format Content
Kenyatta University	Thesis	Thesis checked for quality by school of postgraduate in regard to format and content Receives well labeled CD from the school of post graduate
	Journal articles	Checked by the heads of department and forwarded to library for uploading
	Others	Received and uploaded not much done to them
JKUAT	Thesis	Thesis checked for quality by school of postgraduate in regard to format and content Receives well labeled CD from the school of post

		graduate to check referencing and plagiarism
	Journal articles	Checked by the heads of department and forwarded to library for uploading
	Others	Received and uploaded not much done to them

All the three universities under study received PHD and masters’ theses in Compact Disc (CD) to upload the resource to the repository. In the universities where the final CD was picked from the school of post graduate, there were cases where the repository sometimes received blank CDs. The greatest problem was where these CDs were to be stored as they were too many and the repositories did not have specialized storage equipment for them. In the three cases under study the CDs were stored in boxes and therefore they could not be relied upon as a backup in future.

The need for both a PDF and a word document was informed by the fact that whereas the PDF is the best accepted file format, the word document was necessary in case there were correction to be made on the document.

Majority felt that Dspace was enough to guide them on types of formats since it allowed them to choose three levels of preservation format; supported, known and unsupported.

In regard to metadata all the three universities used the Dublin core metadata set. This is a general metadata schema that may not cater very well for all the types of resources in the repositories. Most of the metadata collected was descriptive to facilitate access. Elements collected to support this include: provenance (author), title, date of publication, language, type of resource (thesis, speech etc.), and the universal resource identifier (URI).

In cases where the resource is a chapter in a book or a journal article in a journal, structural metadata was provided. Structural metadata ensures that the resource is linked to other components and prevents separation over time a necessity for long term access.

The UoN repository utilizes the creative common licenses to support its open access publishing method. The repository provides the license with the metadata in order to inform user on what is expected of him/her.

As for technical metadata, all the repositories under study only indicated the file format but nothing is given on the resources technical environment. This is a serious oversight since due to media and technological obsolescence, technical metadata is what will be used by future

generations to understand the computing environment of the resource that will ensure that it is successfully rendered.

Administrative metadata collected include rights metadata that indicated whether the item was available on full text to everyone or just a section of the users. The UoN included some metadata indicating any changes that had been done to the digital resource. This was given in form of a note to show any changes that have been made to the original document such as corrections to title, etc. Event date accessioning, is also given.

Figure 1: sample metadata collected by the IRs

The screenshot displays a digital repository page for a thesis. The title is "Administrative Practices Influencing Effectiveness of Peer Counselors' Management of Secondary School Students' Discipline in Kahuro Sub-county, Kenya". The metadata table lists various fields such as contributor, date, language, publisher, rights, and title. Below the metadata, there is a section for "Files in this item" which lists two files: a PDF document and a license file. The right sidebar contains navigation options like "Search UoN Digital Repository" and "This Collection", along with a "MY ACCOUNT" section with "Login" and "Register" buttons.

Field	Value	Language
dc.contributor.author	Kishoyian, Seki P	
dc.date.accessioned	2020-03-12T07:38:08Z	
dc.date.available	2020-03-12T07:38:08Z	
dc.date.issued	2019	
dc.identifier.uri	http://erepository.uonbi.ac.ke/handle/11295/109288	
dc.language.iso	en	en_US
dc.publisher	University of Nairobi	en_US
dc.rights	Attribution-NonCommercial-NoDerivs 3.0 United States	*
dc.rights.uri	http://creativecommons.org/licenses/by-nc-nd/3.0/us/	*
dc.title	Administrative Practices Influencing Effectiveness of Peer Counselors' Management of Secondary School Students' Discipline in Kahuro Sub-county, Kenya	en_US
dc.type	Thesis	en_US

Files in this item

	<p>Name: Seki Patrick Kishoyian 1.pdf Size: 760.7Kb Format: PDF Description: full text</p>	View/Open
	<p>Name: license_rdf Size: 811bytes Format: application/rdf+xml</p>	View/Open

This item appears in the following Collection(s)

- College of Education and External Studies (CEES) [5581]

[Show simple item record](#)

Except where otherwise noted, this item's license is described as Attribution-NonCommercial-NoDerivs 3.0 United States

The repository software also collected other technical metadata like format and size of file. From these findings, it can be concluded that in terms of metadata a lot needs to be done to support digital preservation since currently the IRs are concentrating on descriptive metadata to promote access forgetting that future access is dependent on how well the digital resources are preserved.

Concerns on whether the metadata collected was enough for long-term digital preservation, the respondents felt that what they had was enough based on the fact that most of the resources received had an analog equivalent in case of failure of the digital copy. Lack of awareness on the role of metadata in digital preservation was noted with librarians putting emphasize on descriptive metadata. The study also noted that librarians did not recognize the unique characteristics of digital resources that necessitated more metadata than physical resources.

Sensitization on this issue is very important if future generations are to benefit from these IRs and also to ensure that it is not an effort that will fizzle out when technology becomes obsolete.

Conclusion

From the foregoing, it is clear that institutional repositories are the custodians of the scholarly contents of their institutions. They are tasked with ensuring that these resources will be available and understandable by their users at whichever point in time. The study noted that IRs in Kenya were focused on populating their repositories and although their open access policies pledged to ensure long term access to the resources in their custody, adoption of best practices especially in regard to content and file format selection was not a priority. Instead, effort was put to enlarge large digital collections by using content not stipulated in the open access policies in-order to gain recognition by webometrics. Also of note is that although some procedures were already in place, they were not effectively followed. Lack of awareness among IR staff on the role of content and file format selection as well as the need for preservation metadata on long term access was a major challenge that if not rectified could spell doom to the future of the digital collections in their care.

Recommendations

Developing selection policies with specialists such as archivists will ensure that there is a concrete appraisal criterion to determine the materials that should be preserved in the long-term. This will avoid haphazard collection acquisition that could make long term preservation a milestone.

In addition, the IRs should develop adequate acquisition policies based on a needs analysis that determines the need for the content and as well as formalizing the medium the content is delivered in.

Taking into consideration the financial challenges faced by the libraries, there is need to review selection policies to accommodate the resources only identified from a needs analysis and at the same time avoid collecting media formats that are expensive to maintain.

The importance of a general awareness to all stakeholders cannot be underestimated especially in the used of open file formats that support long term digital preservation.

Standards would also be very useful here as it will lead to formalizing not only the file formats, technology watch and creation of quality metadata that can support digital preservation. Benchmarking with other IRs that have made strides in the area of digital preservation will open the minds of digital custodians on the need for best practices in digital preservation

REFERENCES

- Ahmed, A., Alreyaee, S. & Rahman, A. (2014). Theses and Dissertations in Institutional Repositories: An Asian Perspective. *New Library World*, 115 (9/10),438– 451. DOI: 10.1108/NLW-04-2014-0035
- Arora, J. (2006). *Digital preservation and management: an overview*. In PLANNER-2006: Digital Preservation, Management and Access to Information in the Twenty First Century. India: Ahmedabad. retrieved from: <https://ir.inflibnet.ac.in:8443/ir/bitstream/1944/1187/1/1-20.pdf> on 08/01/2021
- Arp, C. (2019). *Archival Basics: A Practical Manual for Working with Historical Collections*. London; Rowman & Littlefield Publishers. accessed from: https://books.google.co.ke/books?id=_OiRDwAAQBAJ&pg=PA151&lpg=PA151&dq=%22fewer+file+formats%22+%22digital+preservation%22&source=bl&ots=DoXAujKWCz&sig=ACfU3U0SGvRtdAOHjfgXENkpba0m44xnGg&hl=en&sa=X&ved=2ahUKEwi77_TPy6zuAhVfQxUIHbldDH8Q6AEwAHoECAMQAg#v=onepage&q=%22fewer%20file%20formats%22%20%22digital%20preservation%22&f=false on 08/01/2021
- Barve, S. (2007). *File Formats in Digital Preservation*. In A. R. D Prasad & Devika P. Madalli (Eds.), International Conference on Semantic Web and Digital Libraries (pp. 239-248). Bangalore: DRTC.
- Besser, H. (2000). *Digital longevity*. In Maxine K. Sitts (ed). Handbook for Digital Projects: a Management Tool for Preservation and Access (pp 162-176). Andover, Mass.: Northeast Document Conservation Center. Retrieved from: <https://www.nedcc.org/assets/media/documents/dman.pdf> on 03/12/2016
- Breytenbach, A., Lourens, A., & Marsh, S. (2013). The role of Jotello F. Soga Library in the digital preservation of South African veterinary history. *Journal of the South African Veterinary Association*, 84(1), 1-7. Retrieved from: <http://www.scielo.org.za/pdf/jsava/v84n1/21.pdf> on 08/12/2020
- Brown, A. (2006). *The PRONOM PUID Scheme: A scheme of persistent unique identifiers for representation Information*. The National archives; digital Preservation Technical paper 2. Retrieved from: https://www.nationalarchives.gov.uk/aboutapps/pronom/pdf/pronom_unique_identifier_scheme.pdf on 08/01/2021

- Budapest Open Access Initiative. (2001). Budapest, Hungary: Open Society Institute. Retrieved from <https://www.budapestopenaccessinitiative.org/read> on 12/09/2018
- CENDI Digital Preservation Task Group, (2006). Formats for Digital Preservation: A Review of Alternatives and Issues. Retrieved from: http://www.cendi.gov/publications/CENDI_PresFormats_WhitePaper_03092007.pdf on 08/01/2021
- Center for Research Libraries (2007). Trustworthy Repositories Audit & Certification: Criteria and Checklist (TRAC), Version 1.0. Retrieved from: <https://www.crl.edu/archiving-preservation/digital-archives/metrics-assessing-and-certifying/trac>. on 08/01/2021
- Dell, E. Y. & Shultz, S. M. (2014) Conserving Digital Resources: Issues and Future Access. *Journal of Electronic Resources in Medical Libraries*, 11(3), 124-133, DOI: 10.1080/15424065.2014.937657
- Digital Preservation Coalition (2015). Digital Preservation Handbook Technical Solutions and Tools. retrieved from: <https://www.dpconline.org/docs/digital-preservation-handbook2/1550-dp-handbook-technical-solutions-and-tools/file>. on 14/12/2020
- Francke, H., Gamalielsson, J. & Lundell, B. (2017). Institutional repositories as infrastructures for long-term preservation. *Information Research*, 22(2), paper 757. Retrieved from <http://InformationR.net/ir/22-2/paper757.html> on 10/12/2020
- Gartner, R. & Lavoie, B., (2013). Preservation Metadata (2nd edition). *DPC Technology Watch Report* 13-3 May 2013. DOI: <http://dx.doi.org/10.7207/twr13-03>
- Genoni, P. (2004). Content in institutional repositories: a collection management issue. *Library Management*, 25(6-7), 300-306. <https://doi.org/10.1108/01435120410547968>
- Giaretta, D. (2011). *Advanced digital preservation*. London: Springer. DOI 10.1007/978-3-642-16809-3
- Heidorn, P. B. (2008). Shedding Light on the Dark Data in the Long Tail of Science. *Library Trends*, 57(2), 280–299. Retrieved from: <https://www.ideals.illinois.edu/bitstream/handle/2142/10672/heidorn.pdf?sequence=2> on 14/05/2019
- Hockx-Yu, H. (2006). Digital preservation in the context of institutional repositories. *Program: electronic library and information systems*, 40(3), 232-243. <https://doi.org/10.1108/00330330610681312>
- Identifiers for Representation Information (Digital Preservation Technical Paper 2). London: The National Archives. Retrieved from: https://www.nationalarchives.gov.uk/aboutapps/pronom/pdf/pronom_unique_identifier_scheme.pdf on 02/12/2020

- Ismail, A & Affandy, H. B. (2018). Conceptual Paper: Digital Preservation Strategies in Archival Institution. *MATEC Web of Conferences*, 150, 05052 (2018).
<https://doi.org/10.1051/mateconf/201815005052>
- ISO (2003) Space data and information transfer systems: Open archival information system Reference model (ISO 14721:2003), retrieved from http://library.nic.in/iso-record_management/14721.pdf on 11/12/2019
- Johnson, R. (2002), Institutional repositories: partnering with faculty to enhance Scholarly Communication. *D-Lib Magazine*, 8(11). Retrieved from:
<http://www.dlib.org/dlib/november02/johnson/11johnson.html> on 14/05/2019
- Jones, H., & Benson, C. (2016). *Publishing Law*. (5th ed.). Routledge
- Jones, M. & Semple, N. (2006). Mind the Gap: Digital Preservation Needs in the UK. *Ariadne*, 48. Retrieved from: <http://www.ariadne.ac.uk/issue/48/semple-jones> on 06/06/2019
- Knight, G. (2005). *SHERPA DP: establishing an OAIS-compliant preservation environment for institutional repositories*. In Digital repositories: interoperability and common services, 9th DELOS thematic workshop, Heraklion, Crete, May 11- 13, 2005 (pp. 43-48). Retrieved from: <http://delos-wp5.ukoln.ac.uk/dissemination/pdfs/knight.pdf> on 02/06/2019
- Krotoski, A.K. (2012). Data-driven research: Open data opportunities for growing knowledge, and ethical issues that arise Insights: *The UKSG Journal*, 5 (1), 28-32
- Lavoie, B. & Dempsey, L. (2004). Thirteen ways of looking at...digital preservation. *D-Lib Magazine*, 10, (7/8). Retrieved from: <http://www.dlib.org/dlib/july04/lavoie/07lavoie.html> on 06/06/2020
- Lundell, B. (2012). *Why do we need Open Standards?* In Proceedings: EURAS Annual Standardization Conference ‘Standards and Innovation’ (pp. 227–240). Aachen: G. Mainz Verlag. Retrieved from <http://urn.kb.se/resolve?urn=urn:nbn:se:his:diva-6384> on 14/12/2020
- Lunghi, M., Grindley, N., Stoklasova, S., Trehub, A. & Egger, C. (2012). *Economic Alignment*. In McGovern, N. Y’s Aligning National Approaches to Digital Preservation, (195–234). Atlanta, GA: Educopia Institute Publications. Retrieved from:
<https://aurora.auburn.edu/bitstream/handle/11200/48584/ANADP%202011%20Economic%20Alignment.pdf?sequence=1&isAllowed=y> on 30/11/2020
- Mauthner, N. (2013). Open access data sharing policies: Implications for academic roles, practices and identities. *Society for Research into Higher Education*. Retrieved from:
<http://www.srhe.ac.uk/downloads/MauthnerScopingReport.pdf> on 10/05/2020
- Mensah, A. E. M. & Amoafu, E.A. (2019). The story so far-digital preservation in institutional repositories: The case of academic libraries in Ghana. *Digital Library Perspectives*, 35(2) pp. 80-96. <https://doi.org/10.1108/DLP-12-2018-0039>

- Nadal, J. (2007) The Human Element in Digital Preservation. *Collection Management*, 32:3-4, 289-303, DOI: https://doi.org/10.1300/J105v32n03_04
- Najar, J. K. & Wani, J. A. (2019). Digital Preservation: An Overview. *Library Philosophy and Practice*. Available at: <https://digitalcommons.unl.edu/libphilprac/2989>. Accessed on 08/01/2021
- Pinfield, S. & James, H. (2003), The Digital Preservation of e-Prints. *D-Lib Magazine*, 9(9) available at: <http://www.dlib.org/dlib/september03/pinfield/09pinfield.html>
- Property Records Industry Association (PRIA). (2017). TIFF-PDF/A – A White Paper. Retrieved from: <https://docplayer.net/194067446-Tiff-pdf-a-white-paper.html> on 30/11/2020
- Pryor, G. & Donnelly, M. (2009). Skilling up to do data: whose role, whose responsibility, whose career? *The International Journal of Digital Curation*, 4(2), 158-170. <https://doi.org/10.2218/ijdc.v4i2.105>
- Rimkus, K., Padilla, T., Popp, T. & Martin, G. (2014). Digital preservation file format policies of ARL member libraries: an analysis. *D-Lib Magazine*, 20(3/4). Retrieved from <http://www.dprylib.org/dlib/march14/rimkus/03rimkus.html> on 14/05/2019
- Robertson, W. C. & Borchert C. A. (2014), Preserving Content from Your Institutional Repository. *Serials Librarian*, 66, 278–288. DOI: 10.1080/0361526X.2014.881209
- Rodríguez, C. F. (2019). Archiving Art and Design Theses: Report On Institutional Repositories: Decker Library, Maryland Institute College of Art. Retrieved from: https://assets.mica.edu/files/resources/ndsr_art_mica_fontanez_rodriguez_2019.pdf on 08/01/2021
- Shearer, K. (2006), The CARL institutional repositories project: A collaborative approach to addressing the challenges of IRs in Canada. *Library Hi Tech*, 24 (2), 165-172. <https://doi.org/10.1108/07378830610669547>
- Smith, K. (2008). Institutional repositories and e-journal archiving: What are we learning? *The Journal of Electronic Publishing*, 11(1), 1-10. DOI: <https://doi.org/10.3998/3336451.0011.107>
- Suber, P. (2012). Open access. Cambridge, Mass: MIT Press. Updates and Supplements: retrieved from: <https://mitpress.mit.edu/books/open-access> On 08/01/2021
- Thibodeau, K. (2007). If you build it, will it fly? Criteria for success in a digital repository. *Journal of Digital Information*, 8(2), 1-5. Retrieved from: <https://journals.tdl.org/jodi/index.php/jodi/article/view/197/174> on 03/05/2020
- University of London Computer Centre; UKOLN; JISC (2008). PoWR: The preservation of web resources handbook. Retrieved from: <https://jiscpowr.jiscinvolve.org/wp/files/2008/11/powrhandbookv1.pdf> on 08/01/2021

- University of Nairobi (2012). University of Nairobi institutional repository Policy available at: <http://erepository.uonbi.ac.ke/>. Accessed on 10/12/2020
- Vrana, R. (2011). Digital repositories and the future of preservation and use of scientific knowledge. *Informatol*, 44(1). 55-62. Retrieved from: https://bib.irb.hr/datoteka/510540.DIGITALNI_REPOZITORIJI_I_BUDUCNOST_OcUVANJA.pdf on 03/06/2019
- Waddington, S., Zhang, J., Knight, G., Jensen, J., Downing, R., & Ketley, C. (2013). Cloud repositories for research data – addressing the needs of researchers. *Journal of Cloud Computing: Advances, Systems and Applications*, 2(1), 13. <https://doi.org/10.1186/2192-113X-2-13>
- Williamson, A. (2004) Awareness of quality assurance procedures in digital preservation. *Library Review*, 53(4), 204-206, <https://doi.org/10.1108/00242530410531811>
- Woodyard, D. (2004), Significant property: digital preservation at the British library VINE: *The Journal of Information and Knowledge Management Systems*, 34(1), 17-20. DOI: 10.1108/03055720410530942
- Yakel, E., Rieh, S. Y., St. Jean, B., Markey, K., & Kim, J. (2008). Institutional repositories and the institutional repository. *American Archivist*, 71(2), 323-349. Retrieved from <http://hdl.handle.net/2027.42/106421> on 10/10/2019