

Beyond Digitization: Access and Preservation

By

GBAJE Ezra Shiloba and BOT N.G

Abstract

This paper is a desk study that examines issues around and beyond digitization within the context of Nigeria. The paper focuses on access and preservation as some of the main objectives of digitization. It identifies Institutional Repositories (IR) as panacea to achieving these aims. The enabling technologies and challenges for the deployment of IR are also discussed. The need for the libraries to form a consortium to share cost of infrastructures and ICT skilled personnel for the successful deployment of IR is recommended.

Introduction

More and more information are created in digital format ranging from, public records, engineering design, legal documents, records of financial transactions, examination records, and scholarly works amongst others. The 21st century has seen the explosion of digital materials in Nigeria, resulting from the use of computers and Internet in many organizations especially universities, research institutions, the business sector, and many government institutions. While new information is created in digital form daily, many organizations are making efforts to digitize their non-digital born materials to digital format. Digitization involves the process of making non-digitally created materials available in digital format. Eadie (2005), describes digitization as the process of making an electronic version of a 'real world' object or event, enabling the object to be stored, displayed and manipulated on a computer, and disseminated over networks and/or the World Wide Web. The physical or analogue object is 'captured' by some device such as a scanner, digital camera or recorder, which converts the analogue features of the object to numerical values, enabling them to be 'read' electronically. In other words, the digital object can only be a version of the real thing being digitized.

According to the Nigerian Universities Commission website, Nigerian Board of Technical Education and Nigerian Commission for Colleges of Education, there are 92 Universities, 100 polytechnics, 63 colleges of education respectively. There are also many research institutions in Nigeria. Scholars and researchers in these institutions generate large volume of data and knowledge as part of the criteria for their promotion. Their findings are normally published in scholarly journals, many of whom are circulated in only print format. Scholarly publications of scholars and researchers in Nigeria are of great importance to solving local and regional problems as well as contributing to the body of knowledge, hence the need for wider dissemination and preservation for the future generation to build on.

Similarly, majority of the universities in Nigeria run different postgraduate programs and as a requirement the postgraduate students are expected to conduct a study and submit their findings in form of theses and dissertations. These constitute a very large portion of digital assets that need to be disseminated and preserve for future generations.

Statement of Problem

Current investment in the creation of digital asset and digitization is threatened by the rapid change in computer software and hardware technologies thereby jeopardizing the objectives of digitization. Unless significant effort is put into digital preservation to secure long-term access to these digital assets we stand to lose our entire digital asset. Studies by Lawrence (2000), Granger (2002), Macdonald and Lord (2003), Rosenthal, et.al (2005) and Gladney (2007) have revealed that digital objects will cease to be accessible without active management and intervention. These make a clear understanding of the preservation and access issues surrounding digital assets imperative to help relevant stakeholders take timely and informed decisions for the sustainability of digital asset.

The researchers has also observed that many digitization project in Nigeria has failed to adopt appropriate software that facilitate remote access, collection of appropriate metadata for each digital asset, feature that will index, sort and make retrieval of digital asset easy.

Objective of the study

The objectives of this study are:

- i. To highlight the vulnerability of digital resources and strategies for digital preservation.
- ii. To highlight the role of Institutional Repositories in digital preservation and access

Threats to Digital assets

Digital assets either digital born or digitized are stored and preserved on storage media. The most

popular digital storage mediums available today include, hard drive, optical storage media (CD-ROMs, DVD, Magnetic tapes), and zip drive. These storage medium are vulnerable to deterioration in *addition to* the problem of obsolescence in retrieval and playback technologies. Beagrie (2004), and Rosenthal (2005), over the years have identified the following as major threats to digital information:

- i. Storage medium deteriorate overtime;
- ii. the hardware that reads particular carriers wears out and cannot be replaced when it has become obsolete;
- iii. file formats become obsolete in the course of software evolution, as backward compatibility is lost over a succession of versions;
- iv. older versions of software, even when these are available, may not work on new hardware or operating systems.

Valuable digital assets of institutions are at risk of being inaccessible in the short-term as well as long-term as a result of these threats. These inaccessible digital assets may contain highly crucial data and information.

The success of preserving digital materials for long-term also requires standards for file formats. Common examples of this include Tagged Image File Format (TIFF) or JPEG2000 for images and WAV or MPEG for audio file. Dahl, Banerjee and Spalti, (2006) assert that non-proprietary standard file format maximize the use of digital content and greatly simplify migration to new file formats and new generations of computer technology

Digital Access and Preservation

The primary purpose of digitization is to improve access to digital resources. Remote access is essentially very important to disseminating digital asset widely without any geographical limitation. Apart from providing remote access digitization produce digital surrogate of rare or fragile material which reduces damages associated with the frequent physical handling of the original material. The volume of digital asset created requires that appropriate digitization software that will allow information provider enter all the appropriate metadata for easy accessibility is required.

According to Hussein and Priestley(2002) there are at least 400–500 scholarly journals published throughout Africa the largest proportion emanating from South Africa and Nigeria. However because South Africa has long embrace digital publishing the visibility of their scholarly work and contribution to

the body of knowledge is perceived to be relatively higher than that of Nigeria. By digitizing and allowing digital access to their scholarly works, Nigerian scholars will improve the visibility of their work locally and globally.

It has also been argued that digitization is sometimes aimed at digital preservation which is to ensure that digital information remains readable and useable in the future. However, digital preservation goes beyond saving digital asset on storage medium, deliberate actions must be taken to **maintain** digital asset with a view to ensuring continued **accessibility over long-term**. Long-term in this context is referred to mean long enough to be concerned with the impact of changing technologies - and should include timescales of decades and even centuries. The Joint Information Systems Committee (JISC) digital preservation briefing paper defined digital preservation as “the series of actions and interventions required to ensure continued and reliable access to authentic digital objects for as long as they are deemed to be of value”. In general, digital preservation involves a number of organised tasks associated with a variety of technical approaches or strategies for ensuring that digital resources are not only stored appropriately, but also adequately maintained and thus consistently useable over time.

There are essentially three possible strategies for digital preservation which include technology preservation, technology emulation and data migration. The first strategy is the preservation of the original technology used to create the digital asset in order to preserve the functionality and "look and feel" of the product. This is a "museum style" approach which is probably only suitable as a short-term solution. Hardware and software from the object itself are maintained so that access can be guaranteed. However, over a number of years degradation of the hardware makes this approach problematic as a long-term strategy.

Technology Emulation is based on the need to preserve the technological environment and therefore original functionality. Unlike technology preservation, an emulation strategy seeks to preserve that environment not through the preservation of original hardware/software but by using current technology to mimic the original environment. This might involve emulation of the original software or (more likely) emulation of the original hardware (in this case the original software and operating system are stored along with the digital object itself). Either way, the strategy relies on a detailed description of

the original environment on which to base the emulation in future.

Data migration focuses on maintaining digital material in current formats. The attraction of this strategy is that material is maintained in an accessible format. Data migration means the material is maintained in the archive in a currently useable format. As digital collections increase in size, the process of migration may become extremely time consuming and costly, and it could become an almost continuous process. Granger (2000) points out that migration is currently the only practical strategy open to large-scale digital archives. Choice of strategy must reflect fitness for purpose. Certain technical factors will impact on this choice: the basic data types employed in each category; the application programs used to create them; the structures applied to them; the systems used to manage or distribute them prior to deposit.

Role of Institutional Repository in Digital preservation and Remote Access

An Institutional repository (IR) is a set of services that an institution offers to the members of its community for the management and dissemination of digital materials created by the institution and its community members. Westrienen and Lynch (2005) assert that IR is most essentially an organizational commitment to the stewardship of these digital materials, including long-term preservation and facilitating access to digital asset. IR can be used to collect, preserve and provide free, unrestricted access to all types of institutional digital asset. Most IRs are designed to accept and facilitate access to all forms of digital asset including text, images, video and audio files and ensure their availability to future generations in its original digital form.

Institutional repository systems are Open Archives Initiative for Metadata Harvesting (OAI-PMH) compliant. OAI-PMH is a lightweight protocol enabling access to web accessible material from repositories interoperable for metadata sharing, publishing and archiving. It also helps facilitate access via multiple search engines and other discovery tools. This simplicity lowers the barrier to repository operation for many institutions, as it only requires a file system to hold the content and the ability to create and share metadata with external systems (Johnson, 2002).

Institutional repositories particularly DSpace provides a set of tools for helping institutions keep track of their data, organize it in meaningful ways and migrate that data to new formats as old ones

become obsolete. They are now being created to manage, preserve, and maintain the digital assets, intellectual output and histories of institutions. This is because it provides a good platform for ingest of different types of digital asset with their metadata, ability to index documents, features for easy searching, retrieving and long-term preservation among others.

Institutional repositories may be limited to a field of discipline, an institution, or a consortium of several institutions. Governments and government agencies may use digital repositories to store and preserved their digital assets such as white papers, policy documents and other relevant documents. Wilczek and Glick (2006) assert that preservation system requires natural and juridical people, institutions, applications, infrastructure and procedures. The organization of the IR together with the commitment of the sponsoring institution will provide greater assurance than just uploading onto a web server with unknown support and commitment to preservation. Libraries are already taking leadership roles in planning and building digital repositories, fulfilling their roles as experts in collecting, describing, preserving, and providing stewardship for documents and digital information. Wheatley (2004) asserts that digital repository like DSpace have been design to achieve the following in long term:

- i. Data can be maintained in the repository without being damaged, lost or maliciously altered.
- ii. Data can be found, extracted from the archive and served to a user
- iii. Data can be interpreted and understood by the user in the face of hardware and software obsolescence.

Achieving digital preservation process requires that a digital repository to have these functions and infrastructure:

- i. A process of ingest that creates or extracts the metadata necessary to ensure preservation.
- ii. A framework, within which the required Representation Information can be stored, managed and utilized (a Representation System).
- iii. A process of "technology watch" which monitors technology dependencies and the recorded Representation Information, and takes action to ensure continued preservation where technology obsolescence occurs.
- iv. A process of rendering (displaying or making sense of) retrieved digital objects.
- v. A process and related framework for recording change metadata

Enabling technology for deploying IR

Key to the deployment and implementation of Institutional Repositories (IRs) are the basic National Information Infrastructure (NII) which includes; telecommunication, electricity, web technology skilled, ability to set up and maintain a web server and information technology skilled man power. Open source software for building and deploying IR programs such as E-print, DSpace, FEDEORA, ETD-db run on a web server platform. Even though open source solution requires no licensing fee and is available on the Internet free, the major drawback is that, it is not free from requiring human expertise and web technology skill sets to install, configure and customize. Gbaje (2007), Ashcroft and Watts C. (2004), posit that Nigeria has an acute shortage of web technology and Information and Communication Technology (ICT) skilled librarians to deploy basic ICT projects in the library.

The current traditional dial-up and VSAT Internet access available and used in Nigeria is not sufficient and is expensive to deploy and maintain a web server. A high – speed broadband internet access is required for the successful deployment and sustenance of traffic for any ETD programs. High-speed broadband requires fibre optic cables and projects such as the Nepad ICT Broadband Infrastructure network, backed by the African Union, and the World Bank Regional Communications Infrastructure Network (RCIP) are a number of projects underway to connect African countries with the rest of the world through undersea cables. This will go a long way to support the deployment and sustenance of web server programs like IR.

Conclusion

Digitization is only a step toward providing remote access and preservation of digital asset. The amount of digital assets created in various institutions is large and many institutions are not aware of the threats to these digital assets. Librarians as the vanguards of access to information and preservation must continue to play their roles in providing access and persevering digital asset. They must promote the adoption of appropriate technology that will facilitate access and preserve digital asset for long-term. Institutional repositories (IR) which runs on a web server platform and has all the features for long-term preservation of digital asset can be considered as a panacea to digital access and preservation. With the availability of the enabling technology open source IR software can be configured and deployed with minimal cost.

Recommendations

Nigerian researchers, librarians, digital curators and archivists need to sensitize all the relevant stakeholders involved in creating valuable digital asset of the impending threats to digital asset and help mitigate against it. National strategies and policies need be developed to encourage the adoption of digital institutional repository like Dspace that facilitate access and digital preservation of the ever growing digital assets generated in our various institutions.

The scarcity of web technology skilled librarians make its necessary for institutions that want to adopt and deploy IR to collaborate with the ICT unit of their parent institution. While institutions with similarly mission can come together and set up an IR thereby sharing the human and technological infrastructures.

References

- Ashcroft L. and Watts C. (2004). *ICT Skills for Information professionals in developing countries: perspective from a study of the electronic information environment in Nigeria* . Retrieved on 11/30/2006 from: <http://www.ifla.org/V/ifa/IFLA-Journal-1-2005.pdf>
- Beagrie N. 2004. *The Continuing Access and Digital Preservation Strategy for the UK Joint Information Systems Committee (JISC)*. *D-Lib Magazine* Vol. 10(7/8) Retrieved on 21/04/09. From <http://www.dlib.org/dlib/july04/beagrie/07beagrie.html>
- Dahl M., Banerjee K, and Spalti M (2006). *Digital libraries: integrating content and systems*. Great Britain: Chandos Publishing Limited.
- Eadie M. (2005), *The Digitization Process: An introduction to some key themes*. <http://ahds.ac.uk/creating/information-papers/digitisation-process/index.htm>
- Gbaje E.S.(2007). *Implementing a National Virtual Library for Higher Institutions in Nigeria*. *LIBRES Library and Information Science Research Electronic Journal* Vol. 17 (2). Retrieved on 22/10/2007, from <http://libres.curtin.edu.au/libres17n2/index.htm>
- Gladney M.H. (2007). *Preserving Digital Information*. Springer Berlin Heidelberg New York.
- Granger S. (2000). *Emulation as a Digital Preservation Strategy*. Retrieved 10/11/2008, from

- <http://www.dlib.org/dlib/october00/granger/10granger.html>
- Hussein J. and Priestley C. (2002) 'Current Status and Challenges Ahead for Scholarly Journal Publishing in Africa'. Seminar on Strengthening Scholarly Publishing in Africa, 1-4 July 2002, Zanzibar. Dag Hammarskjöld Information Systems Committee (2006). Digital Preservation briefing paper. Johnson K.R (2002). Institutional Repositories: Partnering with Faculty to Enhance Scholarly Communication. Retrieved on 2/05/2009, from <http://www.dlib.org/dlib/november02/johnson/11johnson.html>
- Joint Retrieved on 2/05/2009, from http://www.jisc.ac.uk/publications/documents/pub_digipreservationbp.aspx
- Lawrence W.G et.al (2000). Risk Management of Digital Information: A file format investigation.
- Macdonald A. and Lord P.(2003). Data curation for e-Science in the UK: an audit to establish requirements for future curation and provision. Retrieved on 12/03/2009, from http://www.jisc.ac.uk/uploaded_documents/e-ScienceReportFinal.pdf
- Rosenthal, et.al (2005). Requirements for Digital Preservation Systems. Retrieved on 12/03/2009, from <http://www.dlib.org/dlib/november05/rosenthal/11rosenthal.html>
- Westrienen and Lynch (2005). Academic Institutional Repositories Deployment Status in 13 Nations as of Mid 2005. Retrieved on 06/05/2009, from <http://www.dlib.org/dlib/september05/westrienen/09westrienen.html>
- Wilczek, E. and Click K. (2006). Fedora and the Preservation of University Records Project: Reports and Findings, Tufts University and Yale University. Final Narrative Report to National Historical Publications and Records Commission, September 27.2006. Retrieved on 15/04/2009, from
- Wheatley P. (2004). Institutional Repositories in the Context of Digital Preservation. Retrieved on 22/04/07 from <http://www.dpconline.org/docs/DPCTWf4word.pdf>