Managing a Sustainable Institutional Repository: The Covenant University Experience
by
Nkiko Christopher (Ph.D), Bolu Christian (Ph.D) and Michael-Onuoha Happiness

Abstract
The paper examined the ubiquitous nature of digital institutional repositories and their concomitant capabilities. It specifically show-cased the robust process and technical details involved in the development and sustenance of Covenant University Institutional Repository which has consistently ranked number one repository in Nigeria based on webometric rankings. Relying on Strajda Process Management model, Covenant University adopted two open source software- Dspace and E-print based on adjudged robustness of the metadata, relative easy to setup as well as amenable to customization. The paper recommended among others, the sensitization of researchers on the dangers of ceding copyright to Journal publishers as the practice divests them of inherent rights and may preclude deposition of their works on the Institutional repository thereby undermining growth of content and visibility of research outputs. It further recommended backup systems that forestall intrusion and mitigate effect of potential disasters on sustainability of Institutional Repository.

Introduction
The digitalization of documents and publications as well as uploading of same in an electronic searchable platform is the hallmark of world class universities. Great institutions devote a great deal of time, manpower and financial resources designing Institutional Repositories to enhance learning, teaching and research experiences. The Digital Repository system is a departure from the hitherto paper-base archives. Most Institutional Repositories have dual functionalities of online and offline capabilities. The offline compartment provides very user-friendly interfaces with clickable access to digital holdings, thus bringing advantages of a web experience devoid of bandwidth requirements. On the corollary, the online capability provides access to all institutional documents and publications through an Internet browser.

Institutional Repository provides multiple access to documents. The same publication can be retrieved and used simultaneously by multiple patrons consisting of students, researchers and administrators. Another major advantage of electronic Institutional Repository over manual archives is that digital contents require no physical space for storage. Other benefits include: secure storage of documents, easy retrieval, excellent search capabilities, controlled environment for updates to content, complex security rules to control access, increased global visibility of content, one source and one set of rules for managing documents, reduced time and effort spent on document management and ability to maintain document history to meet legal requirements (Bolu, 2011).

The paper essentially seeks to document and show-case the robust processes and technical details involved in the establishment and sustenance of Covenant University Institutional Repository. The specific objectives are to:

1. examine the factors involved in the development of sustainable platform and processes for Covenant University Institutional Repository

2. determine selection of documents and publications for inclusion as contents into the Institutional Repository
3. Examine process Management questions raised and answered in the design and implementation of the Institutional Repository
4. Identify Infrastructural and Technological Requirements for effective Institutional Repository.
5. Highlight considerations for digital preservation, sustainability, security and interoperability issues.
6. Discuss the issues of intellectual property policies, personnel requirement/training, Back-up and disaster management plans as well as funding.

Literature Review
Livingston and Naltasie (2009) conceptualized the creation of Institutional Repository as a huge support to academic activities in higher institutions. However they pointed out that the challenges of sustainability, preservation, security and interoperability are capable of undermining the entire project if not clearly articulated and resolved. There is therefore the question of who is going to maintain these Institutional Repositories in the long run and to what degree are various units of the University responsible to the sustainability of a centralized repository? How will the information stored in Institutional Repositories be preserved and disseminated to the right people? Who should be authorized to assign access control. The observations provide non-negotiable fundamentals that must be dispassionately settled before embarking on Institutional Repository.

In an attempt to locate the unit in the University that should be chiefly responsible for Institutional Repository, Jones et al (2006) posited that the ideal home within an institution for the institutional repository is the Library. This is premised on the fact that librarians already possess requisite skills needed to set up and run such a venture. The skills

include bibliographic control, metadata creation (cataloguing and classification) indexing, archiving and information retrieval mechanisms.

The issues of metadata has been described as critical in the viability of digital resources. This underscores best practices, consistency, the use of standards in research repositories, quality and ease of access. Bruce and Hillmann (2004) identified and recommended seven metadata criteria as constituting the following: completeness, accuracy, provenance, conformance to expectations, logical consistency and coherence, timeliness and accessibility. The choice of metadata to be adopted for institutional repository has to be guided by these time tested criteria.

Digital Institutional Repository is a desideratum in most African Universities. It is gratifying to note that there exists a robust digital Institutional Repository derived from continuous efforts at digitizing all relevant Covenant University publications and documents for archiving. The processes include: scanning, bookmarking, rasterizing, categorizing and classification, PDF Conversion and uploading. The two scanning devices currently used are (1) HP Scanjet 5590 Doc Twain, (2) HP Scanjet N9120 Doc train. The entire gaunt of activities revolve around digitisation of existing publications through scanning and rasterising the records to become searchable, and then warehousing same using an enterprise software solution.

Background of CU
Covenant University is a Nigerian private Christian institution founded in 2002 whose institutional repository has consistently ranked number one in Nigeria and twelve in Africa (Cybermetric Lab, 2014).

Metadata selection
There are several repository software used in the industry. In determining the appropriate metadata for adoption, Covenant University considered four open sources and one proprietary types as detailed:
- DSpace- a digital repository developed as a joint project of the Massachusetts Institute of Technology (MIT) Libraries and the Hewlett-Packard Company, USA. **DSpace** is an open source software application that allows for the management of digital contents, and provides platform for creation of institutional repository. It supports a wide variety of data, including books, theses, 3D digital scans of objects, photographs, film, video, research data sets and other forms of content. The data is arranged as community collections of items, which bundle bit-streams together. (Kumar and Reddy, 2013).

- Eprints- the GNU Eprints self-archiving software, that has been developed at the Electronics and Computer Science Department of the University Of Southampton, UK. An **eprint** is a digital version of a research document (usually a journal article, but could also be a thesis, conference paper, book chapter, or a book) that is accessible online, whether from a local Institutional, or a central (subject- or discipline-based) Digital Repository(Open Archives Forum, 2003).

- Fedora – Fedora (or Flexible Extensible Digital Objective Repository Architecture) is a modular architecture built on the principle that interoperability and extensibility is best achieved by the integration of data, interfaces, and mechanisms (i.e, executable programs) as clearly defined modules. Fedora is a digital asset management (DAM) architecture, upon which many types of digital library, institutional repositories, digital archives, and digital libraries system might be built. (AIMS, 2012).

- Greenstone is a suite of software for building and distributing digital library collections. It provides a new way of organizing information and publishing it on the Internet or on CD-ROM. Greenstone is produced by the New Zealand Digital Library Project at the University of Waikato, and developed and distributed in cooperation with UNESCO and the Human Info NGO. It is open-source, multilingual software, issued under the terms of the GNU General Public License. (UNESCO, 2014).

- **Harvest Road Hive** is a **digital repository** designed to be the foundation of an enterprise knowledge management strategy. **Harvest Road Hive** is the **content management engine** at the core of the organisation's infrastructure, integrated with authoring and delivery solutions. Hive federates and interoperates with other repositories using open standards for data exchange. The product line includes integration with leading learning Management system, such as **Moodle, Blackboard (Learn and Vista)** and **Sakai.** (Pockley, 2008).
- SAP Netweaver – SAP Document management System developed by SAP AG of Germany. It is proprietary digital asset management software included in the SAP Netweaver technology.

Having recourse to installation requirements, functionality, performance, cost, security, usability, workflow, scalability, application programming

interface and interoperability considerations, DSpace and Eprints were selected. The two appear much more robust and easy to set up while DSpace is amenable to customization.

Process Management

In the design and implementation of academic and administrative content for effective utilization and collaboration, Stajda (2009) suggested that certain pertinent and fundamental questions must be raised and honestly answered as precursor to creating sustainable repositories. The questions are:

- How do documents fit into the overall business process? What is the Business process Flow? Are documents created or required at certain steps in the business process? Which business objects are documents associated with? What are the documents and what is their significance?
- How do we want to search for documents? What are the attributes of the document? Standard attributes-Description, Owner, Responsible Lab/Office Additional Attributes- Application, Release; Full Text Search
- Define Lifecycle of Documents. What are the steps in the lifecycle of the document? In Work, Pending Approval, Approved, Released – No more change-Released version remain as history
- What is the change control process? Are updates controlled through a change control process? Changes of document through Engineering Change Management – Capture reason for change, element of workflow, and digital signature for release; provides a complete history of when and why a document was updated.
- Is there a formal approval process? Before a document is officially released, does it go through a formal approval process/ Facilitated through a workflow process; Might require digital signature; Formal approval results to a released version of document with record of the approver; Further change to document must be by a new version.
- What are the security requirements? What roles in the national business are allowed to change each document? Consider status for changes – In Work seen by select group and Release seen by all.
- What type of application files will be stored? What output file of a specific application is stored? Microsoft Word, Adobe Acrobat, Autodesk AutoCAD; Application configured to behave in a certain manner when associated file is launched for display or change.
- How are versions and revisions used in the business? What do the term Version and Revision mean to the Business; Version is a separate instance of a document that has its own status such as in Work or Released. It is a snapshot in time; Revision level is assigned to a document version and is associated with a release state. Represents a major change; for each document you can store multiple versions with each version, you can assign a revision identifier.

- Do you need to support searching and maintenance in multiple languages? Maintain some attributes in multiple languages? Attribute Description could be maintained in English and French; Capability to maintain entry, display and search attributes in multiple languages.

- What is the volume and size of documents to be stored? How is document to be stored? Infrastructure requirements to be considered; Content Server to be seized appropriately, say 100 TB; Size of each file help in Network Sizing, say 4Mbps; Document consumers may be in a number of different geographic locations – Require Cache Server.

- Location of Creation versus Consumer; What are the different locations of creators and consumer? Creator is someone who generates and stores document in the system; Consumer is someone who searches and displays documents; for large number of Creators install Content Server at that location; for large number of Consumers install a Cache Server; this helps reduce the impact on the performance of WAN.

- Are there document retention requirements? How long should document be stored or made available based on business and legal requirement? How to handle document when retention period expires, say archived or deleted; what does the national & international law require?

- Do documents need to be converted to a neutral format for long term retention? What neutral format should be used – PDF or TIF/ Conversion can be carried out automatically by DMS Conversion Server when the status of released is reached.

Infrastructure and Technology

Basic technological infrastructures remain largely the fulcrum on which Institutional Repository revolves. The requirements vary according to the size and nature of the repository. However, table 1 represents an irreducible minimum hardware requirements and cost implication in naira as exemplified by the Covenant University experience.

Table 1: Hardware Capital Requirement for CU Institutional Repository

S/No	Description	Quantity	Unit Cost	Amount
1	Repository Servers – Hp BL 380 G6	1	600,000	**600,000**
2	24-port CISCO Switch	1	120,000	120,000
3	26U Cabinet	1	250,000	250,000
4	Internet Radio	1	80,000	80,000
5	Content Server – HP PC 4GB RAM	1	150,000	150,000
6	A3 Scanner with feeder – Network	1	450,000	450,000
7	Barcode Scanner	1	40,000	40,000
8	Cache Server – HP PC 4GB RAM	1	150,000	150,000
9	Network Printer	1	65,000	65,000
10	5KVA Inverter & 5KVA UPS	1	250,000	250,000
11	Backup Server – HP BL380 G6	1	600,000	600,000
12	KVM Switch – 8 Port	1	45,000	45,000
			TOTAL	**2,800,000**

Digital Preservation and Interoperability Issues

Beyond the desirability of availability and access to a robust institutional repository, a major concern among repository stakeholders is "ensuring both the long-term maintenance and continued accessibility of the contents" (RLG-OCLC, 2002). The issue of preservation therefore revolves around the technology, metadata structure, deposit interface file formats, security, reformatting disaster, backups recovery, sustainability and funding requirements. In most cases, external specialised service providers undertake the preservation of contents and the platform. However, Hitchcock et.al (2014) maintained that international standards and best practices as exemplified by Open Archival Information System (OAIS) model and Life Cycle Information for E-literature model must be followed. The hallmark of preservation is the need to minimise the risks of digital obsolescence with a view to guaranteeing perpetual access to digital contents devoid of loss, distortion and encumbrances.

Covenant University has an Open Access Policy which requires from its faculty, mandatory deposit of publications and other related documents to the institutional repository. This practice acts as an administrative impetus to stimulate awareness, increase and enrich content accumulation for the university's repository. It further accelerates its webometric ranking profile as well as heightens visibility of the institution's robust talents and research output. Similarly only publications deposited on the repository are recognised and assessed for promotion purposes. The strong management support coupled with entrenched policies help to ensure continuous flow of content.

Critical to the issue of preservation is the back-up system that safeguards the documents against loss, damage, alteration, hacking and disaster. In Covenant University repository, transactions or operations are backed up on cloud storage server located in disparate external locations. This makes for seamless irretrievability in the event of disaster. Firewalls, secured socket layers, authentication, encryption are deployed to prevent the server and entire network from attacks and intruders (Nkiko, 2014). Funding for the repository is usually provided for in the library budget whereas incremental or major researches geared towards better and innovative institutional repository are funded from the Covenant University research grant.

Copyright is a crucial factor in determining the sustainability of intellectual contents into the repository, noting that the contents represented individual intellectual property protected by law.

The above assertion therefore, demands the need for copyright considerations in building Covenant University repository.

Copyright Considerations and Institutional Repository

Copyright presupposes the need to encourage the authors and publishers for continuous scholarship resulting in greater publications in different formats. Copyright is the exclusive right accorded by law to the creator of a literary work, composer or artist with regards to the use, reproduction, and exploitation of his works (Nkiko, 2012). The permission and the consent of the right holders is required to host their articles in the Institutional Repository or else it would amount to copyright infringement.

Some Faculty members of Covenant University were reluctant to release their papers noting that copyright was ceded to the commercial journal publishers when the article was submitted for publication. It is important to point out that one of the exceptions of copyright provides that a literary work produced by the author under contract of employment may vest the right on the employer but will be vested in the author in the absence of any express agreement to the contrary (Gideon, 2008). In the case of Covenant University, since there was no such agreement, only abstracts of articles with transferred or assigned rights were uploaded to the institutional repository. The lecturers need to be educated on the dangers of signing away their rights under the guise of publishing agreements rooted on the pressure of publish or perish syndrome. Suber (2007) has described as ludicrous the justifications by the publishers of enabling them to effectively protect the author's work from plagiarism. The practice of transferred rights, automatically and legally divests the author of any rights to the publication. This is not a fair deal. The publisher does not require more than a non-exclusive licence as opposed to assignment to publish and facilitate the distribution of the work.

Another issue of concern to Covenant University faculty members was the compulsory upload of theses and dissertations on the Institutional Repository for open access. Some faculty members contended that a moratorium period of two years should be allowed before the upload to enable such persons sufficient time to publish from their research works. Others were of the view that rather than full-text, only abstracts should be uploaded. It should be recognized that documents submitted to the University Library are ipso facto documents within the public space and are open to public consumption. However, such publications are duly protected from any form of copyright infringement and plagiarism. Authors must be reassured that nobody can take undue advantage of their works on open access. Again the mere fact that some intellectual contents appear on institutional repository does not make them published items and cannot vitiate the propensity to be published. Theses and dissertations on the digital institutional repository are classified as unpublished items as long as they are not contained in any journal or book. Turnitin or other anti-plagiarism software would definitely track the publication in the repository. This confers authenticity and originality on the work since it enjoys widest online visibility and could be challenged if it were a plagiarised material. It is therefore scholars with unsubstantiated or spurious data that should be afraid of depositing their work on the institutional repository.

Conclusion and Recommendations

Digital institutional repositories provide strong basis for the crystallization of open access to intellectual outputs and enrichment of scholarship in the universities. It is the hallmark of world-class universities as it confers institutional prestige and global visibility. In determining the appropriate metadata to be adopted, attention must be paid to installation requirements, functionality, performance, cost, security, usability, work flow, scalability and interoperability issues. Covenant University selected DSpace and Eprints which are open source software based on the fact that they appear much more robust and easy to set up as well as amenable to customization.

It is recommended that Universities should mount intensive sensitization and enlightenment programmes to educate their faculty members on the dynamics of copyright in relation to their intellectual output. The practice of researchers ceding their rights to the Journal publishers has counter-productive effect on sustainability of repository contents.

Development of institutional repository is a capital intensive project. Universities need to create appropriate budgetary heads for it under the library budget provisions and ensure continuous funding of the project. Requisite training and replacement of relevant equipment must be guaranteed. It is also imperative that back-up systems are strategic to forestall intrusion and mitigate effects of possible disasters.

References

AIMS (2012). Fedora Commons. Retrieved from http://www.aims.fao.org/vest-registry/tools/fedora-commons-o. Accessed on March 7, 2015.

Bolu, C. (2011), Implementing Departmental and Central Administration Document Management systems in Covenant University. Being a Technical Report submitted to Covenant university Management.

Bruce, T., and Hillmann, D. (2004). The Continuum of Metadata Quality: Defining, Expressing, Exploiting. In Dianne. I. Hillmann and Elaine L. Westbrooks (Eds.), *Metadata in practice,* Chicago: ALA. (pp. 238-256).

Cybermetric Lab (2014), Ranking Web of Repositories. Retrieved from http://www.repositories.webometrics.info/en. Ranking-africa/sub-saharan_Africa. Accessed on December 17, 2014.

Gideon, E (2008), Issues and Challenges to the Development of Open access Institutional Repositories in Academic and Research Institutions in Nigeria. A research paper prepared for the International Development Research Centre (IDRC), Ottawa, Canada. pp.40-44.

Hitchcock S, Brody T, Jessie M and Carr L (2014), Preservation for Institutional Repositories: Practical an Invisible. Preserve project, University of Southampton.

Jones, R. Andrew, T. & MacColl, J. (2006). *The Institutional Repository*. Oxford: Chandos Press.

Kumar, K. and Raghunadha, R.(2013). Analytical study on knowledge about open source software in technological institutional LIS professionals. *International Journal of Library and Information Science.vol. 5(11), pp.439-446.*

Livingston, H. And Nastasie, D. (2009), The Role of Academic Libraries in the Sustainability, Preservation and Access control of Digital Repositories. In *Proceedings of* 60 DCC State of the Art Report *EDUCAUSE Australasia '0*9. Melbourne Exhibition and Convention Centre, Melbourne, Australia. Retrieved from http://www.caudit.edu.au/educauseaustralasia09/authors_papers/Livingston.pdf. Accessed on December 18, 2014.

Nkiko, C. (2012), Copyright Exemption and Application of Information Technology as Factors influencing the Transcription of Reading materials into Alternative Format for the Visually Impaired in Nigeria, University of Ibadan, Nigeria unpublished Ph.D thesis pp.

Nkiko, C. (2014), Managing ICT-driven Libraries: Critical Leadership Issues. Paper presented at the leadership and Innovation: Cutting Edge practices for the 21st Century Librarian Conference, Covenant University, Nigeria, July 7-11 pp.8

Open Archives Forum (2003). Comparison of Dspace and Eprints. Retrieved from http://www.oaforum.org/resources/tvtoolscomp.php. Accessed on March 7, 2015.

Pockley, S. (2008). Harvest Road Hive: Analysis report. Retrieved from http://www.duckdigital.net/Research/HRD_report_final.doc. Accessed on March 7, 2015.

RLG-OCLC (2002), Trusted Digital repositories: attributes and Responsibilities. RLG-OCLC Report, May. Retrieved from http://www.rlg.org/longterm/repositories.pdf. Accessed on December 5, 2014.

Suber, P. (2007), Balancing author and publisher Rights. SPARC Open Access Newsletter, Issue 110, June. Retrieved from http://www.eariham.edu/-peters/fos/newsletter/06-02-07.htm#balancing

Stajda, E. (2009), Effective Document Management with SAP DMS, Galileo press.

UNESCO (2014) Greenstone digital library software 2.80: a software suite for building and distributing digital library collections. *Communication and Information.* Retrieved from www.unesco.org/new/en/communication-and-information/resources/publications-and-communication-materials/publications/full-list/greenstone-digital-library-soft. Accessed on March 7 2015.