Developing the Content of Institutional Repository for Academic Libraries in Nigeria

Cajetan Okechukwu Onyeneke PhD, CLN
ICT Librarian
Imo State University, Owerri

Abstract

Institutional repositories have made it possible for wide range dissemination of research outputs through the means of the internet. There are many benefits derivable from institutional repositories in scholarly communication. The benefits include collation of research output fragmented across departments and disciplines, increasing the visibility of your institution, and building intellectual leadership and credibility for authors. There have equally been apprehensions towards repository deposit, especially open access publishing which centres around lack of motivation to self-archive; concerns surrounding intellectual property, copyright and plagiarism; and negative attitudes towards open access publication and archiving as legitimate mode of scholarly communication. To overcome these challenges, institutions are now adopting mandatory deposit policies to boost repository's content and create a sustainable, accessible collection of their institutional scholarly output. This paper aims to discuss the important role institutional repositories play in archiving scholarly output of academic institutions in Nigeria. It also explores the benefits of institutional repositories to these institutions and the society at large.

Keywords: Institutional repository, Archiving software, Content, Academic libraries, Nigeria.

Introduction

Advancements in information and Communication Technology (ICT) have given rise to the emergence of open access institutional repository which is a digital archive for the collection, preservation and dissemination of institutional research outputs. Institutional repositories have made it possible for wide range dissemination of research outputs through the means of the internet. Christian (2008) posits that the growth of open access institutional repositories has been very remarkable in many developed countries. However, academic and research institutions in many developing countries like Nigeria are still battling to overcome many challenging issues in attempt to make their research outputs openly accessible by means of internet technologies like institutional repositories. Such challenging issues include lack of awareness of open access institutional repositories among researchers and academics in the country's academic and research institutions; inadequate information and communication technology infrastructures (A major problem here is the high cost of internet bandwidth in the region); inadequate funding; inadequate advocacy which is directly linked to low level of awareness of open access institutional repository in Nigeria.

As centres of intellectual and scholarly research, academic and research institutions, whether in developed or developing countries, should take interest in the creation, dissemination and preservation of knowledge. Knowledge dissemination is especially important in the third world context because the emergency of an independent intellectual life and some self-sufficiency in science is to some extent dependent on establishing the essential structure for dissemination of knowledge (Altbach, 1978).

In the past, the only conventional model adopted in preserving and disseminating knowledge from academic and research institutions remained institutional libraries and scholarly publishing. The former housed research outputs in the form of grey literatures and thus played

a greater role in form of preservation than dissemination, while the later played a much greater role in terms of dissemination through scholarly journals. However, there is shift from the past conventional model to a more networked information model giving rise to institutional repository. Institutional repositories in academic libraries in Nigeria are trending in scholarly discourse among members of the academia.

Institutional Repository (IR): Conceptual Approach
Lynch (2003) defines institutional repository as:
> A set of services that a university offers to the members of its community for the management and dissemination of digital materials created by the institutional repository is most essentially an institutional or organizational commitment to the stewardship of these digital materials including long-term preservation where appropriate, as well as organization and access or distribution.

This definition implies that an institutional repository might also include other digital assets generated by academics, such as administrative documents, course notes or conference proceedings. Deposit of materials in an institutional repository is sometimes made mandatory in some institutions.

An institutional repository is a digital archive of the intellectual product created by the faculty, research staff, and students of an institution and accessible to end users both within and outside of the institution with few (if any) barriers to access (Crow, 2002). Repositories hold both the scholarly works of an institution and, in some cases, the cultural heritage of the area hosting the institution (Ezema & Okafor, 2016). Repositories involve change in the manner research output is shared, disseminated, preserved and published. The main reason of setting up an institutional repository is to bring together, preserve and provide access to the intellectual output of an institution.

Barton and Waters (2004) describe institutional repositories as databases that provide services for capturing, storing, indexing, preservation and redistribution of a university's scholarly research in digital formats. Some of the main objectives for having an institutional repository are to provide open access to institutional research output by self-archiving in an open access repository; to create global visibility for an institution's scholarly research; and to store and preserve other institution digital assets, including unpublished or otherwise easily lost ("grey") literature such as theses, working papers or technical reports (Wikipedia, 2019).

The origin of the notion of an institutional repository is twofold:
a. Institutional repositories are linked to the notion of digital interoperability, which, in turn, is linked to the Open Archives Initiative (OAI) and its Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH). The Open Access Initiative had its roots in the notion of a "Universal Preprint Services", which has since been superseded by the Open Access Movement.
b. Institutional repositories are partly linked to the notion of a digital library – i.e., collecting, housing, classifying, cataloguing, curating, preserving, and providing access to digital content.

Benefits of Institutional Repositories in Scholarly Communication
The benefits of institutional repositories will be better appreciated when viewed from the perspective of stakeholders. This includes the activities of the researcher, the institution and the society. The benefits have been discussed thus:

Researcher benefits:
1. IR facilitates increased visibility of research output of individual researcher, the department and the entire institution.
2. IR increases impact of publications of an author at the institutional. Any research output made freely available can be disseminated more widely and thus have greater impact.
3. IR helps to manage and store digital contents connected with any research, including the underlying research data set.
4. IR Offers usage metrics so that researchers can determine hit rates on specific papers.
5. IR creates the potential for researchers to undertake citation analysis through following links to papers held in other repositories.
6. IR allows the creation of personalized publications lists for researchers.

Institutional Benefits
1. A repository can interoperate with other university systems and maximize efficiencies between them by sharing information.
2. A repository increases visibility and prestige of an institution (however, this depends on the richness of the contents).
3. The content of a repository is readily searchable locally and internationally.
4. A repository allows an institution to manage their intellectual property rights by raising awareness of copyright issues.
5. A repository that contains high quality contents could be used as a "shop window" or marketing tool to entice staff and students, and to also attract funding.
6. A repository can store other types of content that are not necessarily published, sometimes known as "grey literature".
7. A repository may be an important tool in managing an institution's Research Assessment Exercise (RAE) submission.

Societal Benefits
1. IR assists research collaboration through facilitating free exchange of scholarly information (this is made possible through the use of metadata harvesters of OAI – complaint institutional repositories).
2. IR aids in the public understanding of the research endeavours and activities.
3. IR makes it easy and possible to access any research works as long as internet presence is available.

Types of Institutional Repository Softwares
A great number of Digital Library (DL) and Digital Repository (DR) systems have been developed and published as open-source software. This large number of software systems becomes a challenge when institutions and organizations plan to build a repository infrastructure to host their collections. The good news is that there are many articles and surveys that evaluate and compare Open-Source Digital Repository and Digital Library software. Open Society Institute (2004) provided one of the first guides for selecting open-source repository software based on the features and benefits of 9 different repositories. Goh et al. (2006) equally produced an extensive checklist for evaluating Digital Library. Two papers also compare and evaluate some current open-source Digital Repositories and Digital Libraries Software (Masrek & Hakimjavadi, 2012; Tramboo et al., 2012). The main concern of these papers is the comparison of the software systems based on some quantitative quality features, in order for interested institutions and organizations to select the proper system for their digital collections. An institution or organization can be guided into selecting repository
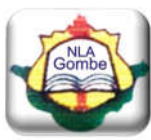
software that matches its criteria based on five collection paradigms that represent case studies of different contents and functionality.

The table below shows five repository software systems selected for a comparative study based on three restrictions. The repository systems:
1. Being publicly available using an open-source license.
2. Being complaint with the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH).
3. Having a large number of installations worldwide.

Software	Initially Developed by	License	Website
DSpace	MIT Libraries and Hewlett Packard Labs.	BSD Open Source License	https://dspace.lyrasis.org
Fedora Commons	Cornell University and the University of Virginia Library	Apache License Version 2.0	https://getfedora.org
Greenstone	University of Waikato	GNU General Public License	https://www.greenstone.org
EPrint	University of Southampton	GNU General Public License	https://www.eprints.org/us
Invenio	CERN Document Server Software Consortium	GNU General Public License	https://invenio-software.org

Table 1: The five repository software systems selected for comparison.

Repository Software Characteristics
Some characteristics and features are expected from modern repository software. The following are the characteristics selected based on models for repository and Digital Libraries Systems.

1. **Object Model:** The internal structural of the digital object. This, according to Kahn & Wilensky (2006) refers to the entity that integrates metadata, digital content and relationships with other objectives.
2. **Collections and Relations Support:** This includes collection description metadata, definition of collection hierarchies and templates that describe the format of the digital objects or the presentation of the collection.
3. **Metadata and digital Content:** This refers to the storage capabilities of the system for preserving the digital content, the metadata sets and the digital object.
4. **Indexing, search and browse:** What are the mechanisms used for indexing and searching on the metadata. It is important for the repository to support indexing and searching not only for a restricted metadata set, but also for specified metadata fields.
5. **Object Management:** There should be methods and user interfaces provided from the repository to manipulate metadata and digital content using CRUD (Create, Read, Update, and Delete) actions. There should be support for the submission of digital objects using workflows.
6. **User Interfaces:** Provide web or desktop users interfaces used as the front-end of the repository, presenting the hosted collection and the contained digital objects.
7. **Access Control:** Support for users, groups and roles, as long as authentication and authorization methods are concerned.
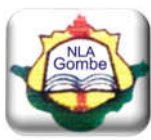
8. **Multiple Languages Support:** Multiple Languages should be supported in the user interface, the metadata fields and the digital content. The character encoding is of great importance in order for the repository to be fully multilingual.

9. **Interoperability Features:** What are the standards that the repository system supports in order to ensure interoperability with other software applications like RSS, Atom, SWORD (simple Web-service Offering Repository, Deposit) and others. The use of web services assures the proper interoperability with other applications.

10. **Level of Customization:** This is customization of the repository in collection level, the format of the digital objects and the services provided. The quality and methods provided by the application programming interfaces (APIs) of the system is equally important.

11. **Extended Service Support:** This refers to extra services that are provided from the repository framework or support of plug-in and add-ons from other systems.

12. **Preservation Support:** Support for features and processes responsible for the preservation of content, including backup, replication and migration process, preservation of metadata, versioning, checksums and archiving solutions.

13. **Installation/Community Support.** This is the support provided by a software community for the selection and usage of repository software. Also a large number of installations and an active community of users and developers usually offer warranty for the software's evolution in the future.

14. **Collection Hosting/Cloud Support:** Many organizations provide their repository software as a service for a yearly or monthly fee. The hosting is mainly offered per collection and usually in a cloud environment.

Repository and Content Development

Davis and Connolly (2007) carried out an evaluation and found that the Cornell University repository is under-populated and under-used by academic staff. The reasons they found for the non-use bordered on staff preference for existing alternatives to repository, a perception that repositories were redundant, technical difficulties, concern that their work may be plagiarized, concerns regarding quality and status of the repository, and confusion over copyright. Similar research conducted by Foster and Gibbons (2005) found that majority of researchers did not perceive the repository to be relevant to their needs, nor perceive any potential benefit from using the repository.

Apprehension towards repository deposit (and open access publishing in general) centres around three key issues: Lack of motivation to self-archive; concerns surrounding intellectual property, copyright and plagiarism, and negative attitudes towards open access publication and archiving as legitimate modes of scholarly communication. This may have informed the mandatory deposit policies adopted by some institutions to boost repository content and create a sustainable, accessible collection of their institutions research outputs. The mandatory deposit policy may apply to specific types of research output, to academic staff outputs, or to postgraduate theses. While mandates take time to be embedded in staff work processes, they have proved to be an effective way of achieving growth and guarantee the sustainability of repository collections (Sale, 2006). On the contrary, other repository developers propose that mandates damage goodwill towards the repository as an integral service and that staff buy-in is more important to sustainability than any guarantee of regular submissions of content (Palmer et al., (2008).

Ways of Organizing Content

Each institutional repository service organizes content in a way to suit its university's unique culture and academic organizations. The content of an institutional repository is institutionally defined, scholarly, cumulative and perpetual, open and interoperable. Many universities organize content according to academic research centres or departments. This is by no means the only organizing principle.

One university (the University of Kansas) organizes its content using "hybrid communities" in the following ways:

1. **Formal Community** – This consists of departments, research centers, and groups already existing. It follows established submission guidelines and workflow. Example: Neuroscience Dept.
2. **Subject Community** – Open access, all academics can submit, or by proxy. Library staff reviews content before going online. Example: Policy Research Institute.
3. **Community of Interest** – This is an ad hoc group that crosses through many departments, Scholar driven, and membership is limited to academic choice. It changes over time. Example: Social Science – cuts across departments.

How you organize content communities depends on the interests and allegiances of your academics. Most software systems call for customizations based on your content communities.

What Content should be Included in the Repository?

Content recruitment is the key to a successful institutional repository. Proper attention is paid as to what should constitute the content of a repository. Listed below are the suggested contents:

- **Contents from students** – Master's theses, Doctoral dissertations, Honours theses, and Research Colloquium papers.
- **From the Faculty** – Pre-prints or Post-prints of scholarly articles, Working papers, Book chapters of Full books, Conference presentations and Reports.
- **Institutional** – Position papers, Accreditation self-studies, Annual Inaugural lectures, and Annual reports.

Digital Collections Content Case Studies

In the following paragraphs, five different collection types are described and one or two repository software systems are proposed in each case. The software is proposed based on the special features specified by each collection as long as the flexibility that the system provides in order to implement some of the features. The five collection paradigms were selected to represent different needs and features, regarding metadata, digital content format, relationships, administration, and preservation issues. The collections are: a scientific data collection, a digitized content collection, a rare books collection, an Electronic Theses and Dissertations (ETDs) collection and a new media art collection. For each case we state a brief description, a list of the content types supported, a specification of the special features required and we propose the repository software for the collections hosting.

Scientific data collection

Case description: Scientific data extracted from research experiments, observations or surveys usually are critical and valuable data, important for researchers worldwide. For many years scientific data used to be stored in local databases or custom applications developed by the research organizations, because scientific repositories and DLs was not always high on the priority list of science and technology researchers (Wallis et al., 2010).

On the other hand, Digital Agenda for Europe (2010) states that publicly funded research should be widely disseminated through Open Access publication of scientific data and papers. Scientific repositories are needed for managing and sharing datasets, publications, reports and other types of content for public or internal use. Researchers should have the ability to submit their datasets or publications and select the access policies.

Content types: Datasets mostly in text files or spreadsheets, documents (usually Word documents and PDFs), presentations. Sometimes video and audio files from observations are available.

Special features: Submission by researchers and curation by librarians or specialized staff; uses registration support and access policies; linking between objects (datasets and publications); exporting datasets in common formats.

Proposed solution: For this case it seems that the most appropriate software systems are; Invenio and DSpace. They support workflows where a registered user may proceed with the submission and other user groups may review and edit the submitted object. Furthermore they support linking between objects of different collections (e.g. publication object with dataset object). In addition, they provide collaboration features for user groups and Invenio supports citation metrics for articles. Paradigms of such repositories are Zenodo Repository and CERN Document Server which are implemented using Invenio software and Dryad Repository developed using DSpace.

Digitized content collection
Case description: An organization is planning to digitize collections from libraries, archives and museums and host them in a single repository. The organization has human resources and the amount of time in order to customize the DR system and develop extra modules. The highest priority needs are the support of preservation issues, the use of multiple metadata standards and the different formats of digital content.

Content types: Images, videos and 3D objects of digitized items (books, paintings, objects, sculptures, etc).

Special features: Different metadata sets and digital content formats; relationships between objects; submission by librarians, archivists or curators; preservation support; detailed access policies depending on content.

Proposed solution: In that case the most suitable repository system is Fedora, since it provides a very customizable modular architecture. It supports multiple collections having different content models, various content formats that may be associated with proper service objects for their presentation and manipulation. Preservation features as technical metadata, versioning, checksums and content replication are supported. Access policies can be defined using XACML or FeSL in collection, object or data stream level. Submission of objects is feasible using the Fedora Administrator but using REST or SOAP APIs collection specific web interfaces can be developed. An example of a repository that hosts multiple digitized collections using Fedora is Pergamos Digital Library (Pyrounakis et al., 2014).

Rare books collection
Case description: A library plans to electronically publish rare books in an easy to use customizable repository system. The books are digitized as high quality images and their structure must be retained based on the book's table of contents. The full text of each book is extracted and should be searchable. Basic metadata will be stored for each book as title, author

and publication year. The library does not possess enough human resources for the installation and customization of the repository, so it needs an "out of the box" solution.

Content types: Digitized images of the book pages, extracted text and PDF files.

Special features: Submission of content and metadata by the librarians; hierarchical structure of books; full text indexing of book content; easy installation and maintenance.

Proposed solution: In that case the most appropriate repository system is Greenstone, since by default it represents books in a hierarchical manner, using table of contents. The full text of the book is searchable in paragraph, chapter or document level, using the provided search engine. Greenstone requires few human resources for its installation and maintenance, because it is designed and developed considering its distribution to organizations in developing countries.

Electronic Theses and Dissertations collection
Case description: A university needs a digital repository for ETDs as long as for publications produced by students and staff. Documents are submitted by authors and staff, using basic metadata and predefined subjects. The hierarchy of the organization should be represented in the repository. The collections will be part of a federated repository for ETDs using OAIPMH.

Content types: Mostly documents in Word or PDF format and archive files containing additional data.

Special features: Use of authority files; users authenticated using LDAP or other centralized authentication protocol simple web interfaces for the submission of documents; OAI-PMH Data Provider support.

Proposed solution: In that case, the most appropriate systems are EPrints and DSpace. They both use authority files to implement subject headings and support the organization's hierarchical structure (DSpace by default represents communities and EPrints hierarchical authority files). They provide web interfaces for the submission of metadata and digital content by registered users. The users can be authenticated by the organizations' authentication mechanism using LDAP or Sibboleth protocol. Both repositories support OAI-PMH as Data Providers, so they can contribute to the federated repository.

Conclusion
Institutional repositories are essential tools for universities. It fulfils number of roles in a university. Open access repositories would broaden the usage. Knowledge may be expensive to generate, but it should be inexpensive to use. Nigerian universities libraries need to take positive initiation in developing repositories in order to widely disseminate the intellectual output of universities. In the competitive education market, institutional repositories provide a platform on which a university can showcase its research. Formulating collection policy, standardization of bibliographic details, checking copyright policy and persuading faculty members to deposit their publications were the major challenges focused that must be addressed by institutions in order to deploy a robust repository. More difficult is the effort needed to persuade the researchers to deposit their publications. Success of institutional repository depends on the willingness of authors to deposit their work. Librarians need to continue to educate them about the importance of institutional repository. Its maintenance must be an institutional strategic goal. Sharing experience in developing institutional repository will be useful to universities in Nigeria.

References

Altbach, P. G. (1978). Scholarly publishing in the third world. *Library Trends,* 26(4): 489-504.

Barton, M. R., & Waters, M. M. (2014). *Creating an institutional repository: LEADIRS workshop.* Massachusetts: MIT Libraries.

Christian, G. 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.*

Crow, R. (2002). *The case for institutional repositories: A SPARC position paper.* Washington, D.C.: The Scholarly Publishing and Academic Resources Coalition. Retrieved from http://www.arl/sparc/IR/IRFinalrelease102.pdf.

Davis, P. M. & Connolly, M. J. L. (2007). Evaluating the reasons for non-use of Cornell University's installation of DSpace. *D-Lib Magazine,* 11(1).

Ezema, I. J. & Okafor, V. N. (2016). Open access institutional repositories in Nigeria academic libraries: Advocacy and issues in scholarly communication. *Library Collections, Acquisitions & Technical Services.* Pp. 1-45.

Foster, N. F. & Gibbons, S. (2005). Understanding faculty to improve content recruitment for institutional repositories. *D-Lib Magazine,* 11(1).

Goh, D. H., Chua, A., Khoo, D. A., Khoo, E. B., Mak, E. B., & Ng, M. W. (2006). A checklist for evaluating open source digital library software. *OnlineInformation Review,* 30(4): 360-379.

Kahn, R. & Wilensky, R. (2006). A framework for distributed digital object services. *International Journal on Digital Libraries,* 6(2): 115-123.

Lewis, S., deCastro, P., & Jones, R. (2012). SWORD: Facilitating deposit scenarios. *D-Lib Magazine,* 18(1/2).

Masrek, M. N., & Hakimjavadi, H. (2012). Evaluation of three open source software in terms of managing repositories of electronic theses and dissertations: A comparison study. *Journal of Basic and Applied Scientific Research,* 2(11): 10843-10852.

Open Society Institute (2004). *A Guide to Institutional Repository Software.* 3rd ed. New York: Open Society Institute.

Palmer, C. L., Teffeau, L. C. & Newton, M. P. (2008). Strategies for institutional repository development: A case study of three evolving initiatives. *Library Trends,* 57(2): 142-167.

Pyrounakis, G., Nikolaidou, M. & Hatzopoulos, M. (2014). Building digital collections using open source digital repository software: A comparative study. *International Journal of Digital Library System,* 4(1).

Tramboo, S., Humma, H., M Shafi, S., & Gul, S. (2012). A study on the open source digital library software: Special reference to DSpace, EPrints and Greenstone. *International Journal of Computer Applications,* 59(16): 1-9.

Wallis, J. C., Mayernik, M. S., Borgman, C. L. & Pepe, A. (2010). Digital libraries for scientific data discovery and reuse: From vision to practical reality. In *Proceeding of the 10th Annual Conference on Digital Libraries, New York, USA,* pp. 333-340.