

https://dspace.org/

Open access institutional repositories: A requirement for academic libraries in the 21st century. A case study of four African universities

Ibinaiye, Dorcas; Esew, Michael; Atukwase, Thecla; Carte, Sean; Lamptey, Richard

2015-05-29
ICADLA4
http://hdl.handle.net/10539/18398

Downloaded from a DSpace repository. https://dspace.org/

OPEN ACCESS INSTITUTIONAL REPOSITORIES: A REQUIREMENT FOR ACADEMIC LIBRARIES IN THE 21ST CENTURY. A CASE STUDY OF FOUR AFRICAN UNIVERSITIES

Dorcas Ibinaiye, BLIS, MIM, MLS, CLN
Kashim Ibrahim Library
Ahmadu Bello University
Zaria, Kaduna State Nigeria
irewole44@gmail.com

Michael Esew, NCE, BLIS, MSc. Information Science, CLN
Kashim Ibrahim Library
Ahmadu Bello University
Zaria, Kaduna State Nigeria
yaw4mi@gmail.com

Thecla Atukwase
Ugandan Christian University
Uganda
atukwaset@gmail.com

Sean Carte
Durban University of Technology
South Africa
sean.carte@gmail.com

Richard Lamptey
Kwame Nkrumah University of Science and Technology
Ghana
phanerosis75@gmail.com

Abstract

The paper highlights the importance and expectation of a global access to information, open access institutional repositories as a requirement for academic and research libraries. The paper further investigated the challenges faced by four African universities in implementing their open access repositories. A literature review method was adopted by way of questionnaire and observation using ProQuest/Serial solutions summons which was a web scale discovery tool that indexes a wide variety of databases. Population of study comprises of four selected African universities; Ahmadu Bello University(ABU), Zaria, Kaduna State, Nigeria; Kwame Nkrumah University of science and Technology, Ghana; Durban University of Technology (DUT) South Africa; Uganda Christian University (UCU), Uganda. The data collected were analysed using Stellenbosch's best practices guide, "SUNScholar/practical guidelines for starting an institutional repository (IR)" (Gibson). Findings revealed that; KNUST have both institutional policy and open access policy that currently in draft form and no embargo on open access policy. DUTIR implemented their policy in 2009 without embargo on open access. ABU and UCU have drafted their policy but yet to be implemented. Only

KNUST space has documented a formal long term preservation policy as part of their IR policy. Based on research findings, the researchers are not willing to devote time to ensure that their research work is archived, while institutions seem unwilling to commit financially to ensure the sustainability of their repositories. It was recommended that libraries should take proactive steps to make their repositories more successful.

Keywords: Open Access, Institutional Repository, Self-Archiving, Open Access Policy, Institutional Repository Infrastructure.

Introduction

Open access repository is 'an online database on the Internet which makes the full text of items … it contains freely and immediately available without any access restrictions (Pinfield 2005: 31). Suber (2004) defines OA publications as those publications made freely available online to anyone anywhere with no changes imposed for access. ARL (2004) supported the fact that Open Access allows users to read, download, copy, distribute, print, search or link to the full text of works, permitting use of any lawful purpose, as long as access to the material is possible. Self-archiving is a term that occurs regularly whenever open access repositories are mentioned. It "refers to the right of scholars to deposit their refereed journal articles in searchable and free electronic archives." (Prosser, 2004: 26-27) Swan and Brown (2005: 2) explain that self-archiving "is not an alternative to publishing in learned journals, but an adjunct, a complementary activity where an author publishes his or her article in whatever journal s/he chooses and then simply self-archives a copy. In practice, this means depositing the file, which is usually the author's final version of the article after peer review has been completed, in an open access archive or repository."

The Budapest Open Archives Initiative (BOAI, 2002) recommends self-archiving as the first strategy to achieve open access of 'scholarly journal literature': "First, scholars need the tools and assistance to deposit their refereed journal articles in open electronic archives, a practice commonly called, self-archiving." This is where libraries come into the picture: they facilitate the self-archiving process by providing the resources that enable researchers to make their research openly accessible on the Internet. These resources include not only the librarian's expertise in mediating the submission of the research to the archive, but also in adding appropriate metadata so that other researchers can find the research, and by providing the infrastructure on which the archive runs: the repository.

It is here that the library finds itself moving into the realm of IT services, and often out of its comfort zone, and this is where problems can emerge. But there are many compelling benefits for individual researchers as well as for their institutions, and finally for society. Prosser (2004: 28) lists these as follows:

For the individual:

- they provide a central archive of the researcher's work
- by being free and open they increase the dissemination and impact of the individual's research
- they act as a full CV for the researcher

For the institution:

- they increase the institution's visibility and prestige by bringing together the full range and extent of that institution's research interests
- they act as an advertisement for the institution to funding sources, potential new researchers and students, etc.

For society:

- they provide access to the world's research
- they ensure long-term preservation of institutes' academic output
- they can accommodate increased volume of research output (no page limits, can accept large data-sets, 'null-results', etc.)

Given these benefits, it would seem worthwhile to address the challenges faced by institutions in implementing open access repositories.

Kristing (2004) argued against increase in prices of journal, which hampered the ability of libraries, universities and investigators to acquire publication necessary for research education and therefore suggested more cost effective alternatives to the traditional publishing model. Kirillidou (2006) noted that access to most of the published journals is restricted and controlled by publishers who charge libraries and consumers hefty subscriptions and articles fees to view this materials, creating the inability of academic and research institutions to subscribe to all needed journals and to provide reasonable collection.

Therefore, by identifying these challenges we will be able to suggest recommendations that will assist institutions like our own in implementing their open access repositories.

Research problem

Base on the expectation of global access to information, open access institutional repositories as a requirement for academic and research libraries, in order to provide numerous advantages to institutions and researchers. Researchers observed that challenges to their implementation still exist. It is against this background that this study was designed to investigate the challenges, causes and provide suggestions for their resolution.

Methodology

The research was conducted through literature review, interviews (by way of a questionnaire), and observation. Search tools adopted was ProQuest/Serials Solutions Summon (a Web-scale discovery tool that indexes a wide variety of databases. The following keywords used: 'open access' and 'repositories', and refined the results to include only items with full text online that were peer-reviewed and had a content type of 'Journal Article'. We considered the first 50 of the 9,905 results returned and downloaded articles, based on the relevance of their abstracts.

Generally, full text retrieval done using the DUT proxy worked successfully, but when it did not the articles were retrieved using the University of Pretoria A-Z list to find the appropriate database and then a search for title and author.

Literature review

Bhat (2010) is essentially a review of the literature available and thus provided many useful sources. Chan, et al. (2002) is a call for academics to provide open access to peer-reviewed articles by means of self-archiving and open access journals.

The Berlin Declaration builds on the BOAI with a call 'to realize the vision of a global and accessible representation of knowledge' by providing open access to scholarly material through depositing it in online repositories. Institutions can publicly commit themselves to this goal by signing the Berlin Declaration. Reich, and Rosenthal (2003: 23) provide a brief overview of the LOCKSS software, which 'enables institutions to locally collect, store, preserve, and archive web based journals thus safeguarding their community's access to that content.'

Prosser (2003) defines self-archiving, discusses why open access repositories (OAR) are necessary given the serials and permissions crises, and considers models in which they can coexist with peer-reviewed journals. Anderson, B. (2004) also discusses how self-archiving to OARs has become necessary following the serials crisis, but goes on to consider some of the objections to it: that there is no refereeing of the repository content, that this makes available untested hypotheses, and reduces demand for subscription-based journals causing subscription prices to increase and damaging

organisations that depend on journal profits. He also cautions about the public confusion that could result from the public availability of preprints in major search engines.

Swan, A. and Brown, S. (2005: 6) document an extensive survey on OARs and self-archiving; but of most immediate value is their outline of the challenges facing and benefits offered by OARs: researchers are put off self-archiving by the 'perceived time required and possible technical difficulties, yet OARs "provide a permanent record of all the research output of that institution" and consequently an "archive can also serve as a marketing tool for the institution."

Pinfield (2005: 33-34) argues that deposition should be made mandatory by grant funders as a "mandate would simply help to overcome quickly the cultural and managerial barriers that currently exist in this area." He then outlines the compelling benefits that would result: "Open access could lead to a better public understanding of science, better knowledge transfer between research institutions and industry, better dissemination of high quality content to inform clinical practice." Salo (2008), in stark contrast to the preceding articles, paints a bleak picture of the extant state of OARs, many of which suffer from lack of content, and considers some of the ways that the problems could be addressed, including digitisation of content and promoting the repository to faculty.

Salo considers the following as the problems facing OARs: reluctance of faculty to self-archive, partly due to there being no mandate forcing faculty to self-archive, but also because the software does not facilitate self-archiving and the library is not sufficiently active in promoting and assisting in the process.

Deventer and Pienaar (2008) discuss repositories in the context of bridging the information divide in South African and document the creation of the University of Pretoria repositories as well as that institution's collaboration with the CSIR in setting up the CSIRIS repository. They identify collaboration as a key to successful implementation of repositories. Romary and Armbruster (2009: 5) argue that the institutional repository is unsustainable and amounts to an 'institutional showcase' for those institutions that can afford it. They suggest that independent institutional repositories be replaced with central repositories that would be better placed to guarantee the 'long-term archiving' of material. Aguillo, et al. (2010) introduce and explain the Webometric Ranking Web of World Repositories (http://repositories.webometrics.info). They also explain the need for these quantitative statistics and their role in driving OA.

Creaser, et al. (2010) describe the findings derived from a questionnaire, which intends to ascertain the level of author awareness of OA. The questionnaire also provides a list of motivations and barriers to self-archiving, from the author's perspective. Hashim and Jan (2011) examine five OARs with a view to determining their strengths and weaknesses. In doing so, they outline several parameters

worth investigating when evaluating OARs: visual interface, resource discovery, access, system features (which include: hardware, software, file-types supported, metadata available, and workflow processes), and content management policy. Wacha and Wisner (2011) are in agreement with Salo's (2008) criticisms of institutional OARs, and argue that the problems could be addressed if libraries shift their focus from their own needs to those of the faculty.

There is a trend apparent in the literature on OARs, which begins with (possibly naïve) optimism and gradually develops into increasing scepticism that OARs are capable of meaningfully achieving the goals set out in the BOAI. The articles also consistently address the challenges they raise with suggestions as to how they can be overcome, so the outlook is ultimately not despairing, but shows an awareness of the difficulties in providing open access to academic output via institutional repositories. We are aware that we have barely scratched the surface of the literature on the topic of OARs. To go beyond what has already been written and add something new to the debate it is necessary to consider local implementations of OARs in the light of the issues identified by the current literature, and that is exactly what we shall do in the following sections.

A few of the articles that we would have liked to read were not included because their full-text was buried behind a paywall. We find this particularly ironic given that their subject matter appears to be open access repositories.

Empirical Analysis

Using Stellenbosch's best practices guide, "SUNScholar/Practical guidelines for starting an institutional repository (IR)" (Gibson), we divided our questions that emerged from our review of the literature on the subject into the following categories: general, policy-related, URL, personnel, infrastructure, software, backup and monitoring, visibility, capture of research records, promotion, submission, auditing, and statistical analyses.

We will consider each of these in turn, comparing the responses from each institution. The institutions participating in this study and their repositories are: Kwame Nkrumah University of Science and Technology's (KNUST) KNUSTspace, from Ghana; Ahmadu Bello University's (ABU) Institutional Open Access Repository (ABUIOAR) from Nigeria; Durban University of Technology's (DUT) DUT IR, from South Africa; and Uganda Christian University's (UCU) DSpace@UCU, from Uganda.

General questions

In this section we considered general questions in an attempt to gauge the level of each institution's involvement in the community. Additionally, we asked whether the institutions contribute financially

or to the source code they use, when the repositories were implemented, and what mailing lists they subscribe to.

- KNUST implemented KNUSTspace in 2009. They subscribe to a dspace-hosted mailing list as well as to Stellenbosch University's IR Talk.
- DUT's DUT IR was implemented in October 2007 in collaboration with the Eastern Seaboard Association of Libraries (esAL). They subscribe to the following mailing lists: dspace-general, dspace-tech, duraspace, and irtalk.
- ABU's ABUIOAR was implemented in 2010 without external assistance. They subscribe to Duraspace's dspace-tech mailing list.
- UCU began implementation of DSpace@UCU in 2010.

The very least community engagement that can be expected from institutions benefiting from open source code is that they participate actively in the mailing lists hosted by the organisations that provide the software. By doing so institutions not only assist other members of the community, but are able to develop their own capacity. So it is encouraging that three of the four institutions do at least subscribe to a DSpace mailing list.

As open source software depends on contributions from its community for its viability, however, it is disturbing that none of the institutions considered in this analysis contributes to the repository software's source code or financially to the organisations that provide the software. The institutions and their researchers are the primary beneficiaries of the repositories (Swan and Brown 2005: 6; Romary and Armbruster 2010: 5; and Hashim and Jan 2011: 229), so might seem reasonable to expect that they would contribute in some way to making the core software sustainable.

Policy-related questions

Policies are important because they provide clear, unambiguous guidelines under which repositories can operate and demonstrate long-term institutional commitment to ensuring sustainability.

KNUST has both an institutional policy and an open access policy that is currently in draft form. They do not have an embargo policy. Nor do they mandate archiving. They have been signatories to the Berlin Declaration since 17 October 2011.

DUT implemented their IR Policy at the outset in 2009, and this policy governs all aspects relating to the repository. Consequently they do not have separate open access or embargo policies. According to

their IR Policy, embargoes are the exception rather than the rule and archiving is mandatory for all research output as well as for electronic theses and dissertations (ETDs).

ABU and UCU have yet to draft any repository-related policies, while at the other extreme, only KNUSTspace has documented a formal long-term preservation policy as part of their IR Policy.

Institutions arguably stand to gain the most from a well-run repository since they "provide a permanent record of the research output of the institution and maximize the visibility, usage and impact of research through global access." (Hashim and Jan 2011: 229) Consequently it makes sense for institutions to implement policies that govern their repositories. So it is a little surprising that only half of the surveyed repositories were governed by policies. Perhaps less surprising is that only those two institutions were also signatories to the Berlin Declaration as signing the Declaration publicly commits the institution to making content "openly accessible" (Berlin Declaration 2003), which is not something that could be readily contemplated without first having a comprehensive policy on open access in place.

URL-related questions

The URL of a repository is an important consideration as it should be both 'marketing friendly' and 'persistent' if it is to help "provide access to the world's research" (Prosser 2004: 28).

To be marketing friendly it should be brief and memorable, while persistence is ensured by never having to change the URL.

The URL of KNUSTspace is <http://dspace.knust.edu.gh/jspui/>, ABU's OAIR uses <http://ww.kubanni.abu.edu.ng:8080/jspui>, DUT's IR is accessible at <http://ir.dut.ac.za/>, and UCU's UIS is accessible at <http://dspace.ucu.ac.ug/>.

Of these only DUT's URL conforms to the best practices recommended by in the SUNScholar guide (Gibson) as both KNUSTspace and UIS incorporate the name of the software, 'dspace'; this violates the principle of persistence as a change of software would necessitate a change in URL.

Furthermore, another challenge faced is that both KNUSTspace and ABUOAIR include the 'jspui' suffix, which complicates the URL, making harvesting of data problematic and the URL less memorable. On the other hand the brevity of DUT IR's URL could prove problematic to potential users of the service, such as students, who may not be familiar with the terminology and therefore not understand what an IR is.

Finally, ABUOAIR includes the port number on which DSpace runs. This can prevent access to the repository from institutions whose firewalls only permit access on standard HTTP ports.

Following best practices regarding repositories' URLs can only make sense. It would be best that institutions currently failing to follow these principles make it a priority to register and make use of brief, memorable and persistent URLs.

Personnel-related questions

To ensure "long-term preservation of [an] institutes" academic output' (Prosser 2004: 28) it is essential that sufficient capable personnel are employed to manage both the content and system-related aspects of the repository.

- Ghana's KNUSTspace and Nigeria's ABUOAIR are the most adequately staffed with each having both a repository director as well as a manager in charge of content in addition to a system manager and a Web 2.0 developer.
- KNUSTspace also makes use of two senior staff members for digitising content while the library takes care of systems maintenance.
- ABUOAIR involves the services of a librarian, but has a systems administrator to take care of systems maintenance.
- DUT IR lacks a repository director but does fill the positions of repository and systems manager, with the systems manager also taking care of any Web 2.0 development. Library cataloguers are involved in the final checking of metadata. Library IT is responsible for systems maintenance.
- Like DUT IR, Uganda's UIS employs both a content and a systems manager, but they differ in that their university's ICT services assume responsibility for systems maintenance.

Involvement of central IT is a very good thing as it ensures continuity of the maintenance of the repository and may ensure the provision of a budget for hardware resources. But there is a provision: central IT may not be as dedicated to the mission of providing an available, serviceable repository; much depends on the relationship between central IT and the library. Wacher and Wisner (2011: 385) contend that the focus should be on faculty, and libraries, due to their contact with faculty are better positioned to understand faculty needs. Thus repositories should be a library initiative supported by central IT rather than the other way around.

Infrastructure-Related Questions

As Deventer and Pienaar (2008:20) point out, "[i]t is much more important to make the content available than using what constitutes the state-of-the-art technology." Though, if institutions are to

reap the benefits of having a sustainable open access repository, it makes sound financial sense for management to support their repositories with the necessary resources.

Both KNUSTspace and DUT IR use virtualisation (KVM) rather than running on bare metal, like UIS. Virtualisation provides cost-effective convenience at the price of performance. Virtual servers can be backed up easily and redeployed quickly when necessary, but access to them may be compromised if too many other servers share the same hardware.

It is difficult to draw any conclusions about the repositories from the infrastructure-related questions other than that the four institutions have demonstrated commitment to hosting repositories by making some form of infrastructure available.

Software-Related Questions

Despite there being a variety of options available for repository software, DSpace is the most popular choice. According to The Registry of Open Access Repositories, there are 1,466 DSpace installations, compared to 533 EPrints, which is the second most prevalent (http://roar.eprints.org/).

All the institutions we surveyed use Duraspace's DSpace software. Salo (2008, 26) argues that "for a service intended to appeal to faculty, [Eprints] is currently the best available choice." So it would appear that, unless DSpace has improved in this respect since 2008, appealing to faculty was not a consideration when selecting repository software.

KNUSTspace runs on Linux with MS SQL Server for its database. DUT IR uses Ubuntu Linux 12.04 with Postgres. ABUOAIR also uses Linux as does Dspace@UCU, which also uses MS SQL Server for its database.

Only DUT IR makes use of persistent identifiers by way of handle.net, though KNUSTspace does have plans to purchase some form of persistent identifier.

Though DSpace@UCU does claim to, none of these repositories makes use of secure HTTP. This is troubling given the potential for log-ins to be misappropriated and data compromised, severely impacting on the trustworthiness of the repository.

Only DUT IR makes use of a proprietary add-on, by way of Atmire's Content and Usage Analysis module. It should be noted, however, that DUT did not pay for this but won it in a competition. Institutions seem very reluctant to commit financially towards their repository software; apparently they consider them cost-free because they run on open source software.

KNUSTspace and DSpace@UCU provide a mobile interface for their repository. Given the prevalence of mobile browsers, this is a serious omission for the others as the standard user interface does not provide a very satisfactory experience on small screens.

One grave omission for all but one of the institutions is the lack of a persistent identifier. Not having this limits the serious use for research of the repository, undermining the benefits provided by hosting a repository. A small financial cost allows "the name of the item to persist over changes of location and other related state information" (http://www.handle.net), so researchers can refer to articles with confidence that they will remain accessible to future readers.

In summary, it would appear that the software has been chosen with ease of installation and maintenance for systems people being the prime consideration, and (lack of) expense a close second. Further configuration such as secure HTTP, or the implementation of mobile interfaces, and additional expense for proprietary add-ons seldom bear consideration.

Backup and Monitoring Questions

As the SunScholar Guidelines state, "[p]lan for disaster recovery, this is vital for long term system sustainability" (Gibson). Romary and Armbruster (2010: 17) argue that "a more concentrated infrastructure for publication repositories" is required to ensure sustainability. If basic backup and monitoring procedures are not implemented, then they are probably correct and institutions should rather ensure sustainability by using a professionally-run service and forgo hosting what can only amount to an "institutional showcase" (2010: 5).

Only DUT IR has a comprehensive disaster recovery plan (DRP) in place; daily backups of the asset store, database, and configuration files are automatically sent to a remote server on another campus where weekly and monthly backups are retained. This could be improved on by including a secondary remote backup server at another location. It appears that even when central IT is involved, as in the case of UIS, best practices are not necessarily followed.

Monitoring creates a feedback loop whereby accessibility is continually assessed. The statistics derived from the process of monitoring can be used in the auditing process, which is discussed later. At this point the focus is on the sustainability ensured by monitoring the availability of the repository.

KNUSTspace DSpace@UCU and DUT IR make use of Google Analytics for monitoring, with DUT IR also making use of AWStats and the built-in DSpace statistics, including Solr. Digital Inspiration's Website Monitor Google Docs script is used to monitor uptime. DUT IR also uses Atmire's Content and Usage module to make the statistics available.

Google Analytics provides free statistics on repository usage and is the minimum that should be implemented. But even this is not implemented at one of our four sites, indicating a serious gap in ensuring sustainability of that repository.

Visibility-Related Questions

One of the main points made in the Budapest Open Access Initiative is that it ought to be possible "to crawl [full-text articles] for indexing." Making the repository accessible to machines, such as the Google Web crawler, so that it can be indexed will help items from the repository appear in search results, and ultimately make the repository's content available to researchers.

There are three ways in which this can be facilitated: by registering with OAI-PMH-compliant harvesters, by providing sitemaps, and by using media-filter to extract a text version of other formats such as PDF.

KNUSTspace, DSpace@UCU and DUT IR have registered with some harvesters, but only DUT IR makes use of sitemaps and media filter.

These are easily accomplished procedures that not only go some way to ensuring open access to research, but also promote the institution's research. Therefore there does not seem to be any compelling reason not take these steps, and this seems to indicate a lack of knowledge about the capabilities of the software.

Capture of Research Records Questions

A repository is worthless without content, and that content is made up of research records. Romary and Armbruster (2010: 19) lament that "the fragmented landscape of institutional repositories predominates, but their repositories are overwhelmingly empty." These research records can be provided by researchers directly archiving their content, or mediated by faculty, librarians, or can even be archived by IT.

One way of rapidly increasing content is by digitising analogue content that is already available to the library and, as Salo (2008: 31) emphasises, "Institutional repositories should be able to digitize analogue content" because a "sensible service does not limit itself to born-digital materials, nor does it expect faculty to manage digitization on their own." So it is encouraging that KNUSTspace, DSpace@UCU and AUOAIR are currently digitising materials.

All repositories have standardised on PDF. PDF/A would be the ideal format as it is suited for long-term preservation, but none of the institutions makes use of it. Another implication is that neither institution is archiving ancillary research data. This would, of course, require more disk space, but

would add value to the repository. As Salo (2008: 20) explains, "faculty could find a use for document versioning, and they frequently express the need to share raw materials such as drafts and datasets with a select few while research is in progress. If the repository does not allow faculty to do these things, they will not use the repository, not even for the final product."

Archival of copyrighted articles is sometimes permitted by the publisher, but usually only after an embargo, therefore it is preferable for a repository to permit archiving of either a pre- or post-print. Pre-prints have the disadvantage of not being peer-reviewed, but post-prints are the equivalent of the peer-reviewed article, lacking only the publisher's formatting and editing. "Postprints are not some kind of self-published, second-rate alternative to conventional journal articles: they are those articles." (Swan and Brown 2005: 3) Both KNUSTspace and DUT IR make the copyrighted article available if possible, otherwise resorting to post-prints. In the case of AUOAIR and KNUSTspace pre-prints are also acceptable. DSpace@UCU claims that it will only archive post-prints.

Since electronic theses and dissertations (ETDs) are not subject any copyright restrictions KNUSTspace, AUOAIR and DUT IR make these available.

Only DUT IR has the luxury of a copyright office to assist with intellectual property issues, but KNUSTspace, DSpace@UCU and DUT IR rely on SHERPA/ROMEO to establish copyright ownership of published articles. As Salo (2008: 7) pointed out in 2008, this aspect could be improved by automating the search using the SHERPA/ROMEO API: "because SHERPA/ROMEO has only recently begun to offer an API, repository managers have not yet automated the process of checking whether meditated deposits are legal, leaving permissions management a time-consuming manual process fraught with risk." It is surprising that this functionality is not yet available in the mainstream repository software.

But a more pressing problem among the three repositories that are archiving research records is that none of them makes use of a researcher ID field to disambiguate authors. Using an ORCID ID, for example, would provide the repository with the means to provide citation information to its researchers. This would increase the value of the repository to its most important constituents, the researchers. Equally no repository has implemented a funder ID field.

'Self-archiving' suggests that researchers themselves would archive their research, but "[i]n practice... most deposits are third-party mediated, many by librarians, some by support staff or IT personnel" (Salo 2008: 21). This is unfortunate as it means that it falls to the repository content manager to locate the material, and it also suggests that researchers may not fully support the repository's mission, even where archiving is mandated. This situation at all repositories apart from KNUSTspace, which accepts archiving from researchers, faculty and librarians.

Promotion-Related Questions

As Pinfield (2005: 34) points out, "[i]institution's would have to set up internal support procedures to facilitate deposition." Yet this most obvious promotion of the repository to its users appears to have been completely overlooked by our repositories. This can be accomplished by means of a news blog or mailing list. Users can also be guided through the submission process or assisted in finding relevant information by providing a help wiki.

Creaser et al. (2010: 157) identify the three chief barriers standing in the way of faculty self-archiving as: "concerns over copyright infringement; uncertainty over embargo periods; and unwillingness to place content where other content had not been peer-reviewed." At the very least the first two of these concerns could be addressed by providing information where it is required, and the third barrier would be overcome as a result.

Submission-Related Questions

Items may be submitted to the repository from a variety of sources, including directly by the researchers, from the faculty or research office, or mediated by the library.

KNUSTspace accepts submissions from researchers directly as well from the library, which makes sense given that they are digitising content. DSpace@UCU relies on faculty to provide submissions. Faculty and the Postgraduate School officer provide AUOAIR with material for archiving. DUT IR receives electronic material from the faculty office, but the archiving is the responsibility of the content manager.

Not having researchers directly involved in the submission process affects the repository in several ways. The researchers are not necessarily made aware of the existence of the repository and are not likely to promote the repository to their peers. Also, tracking down materials that need to be archived becomes an onerous task that falls to the library. Swan and Brown (2005: 4) point out that, to the researcher, 'the most obvious benefit of making their work open access is the enhanced citations, and therefore impact, that result.' Consequently it may seem surprising that our repositories appear to be library rather than faculty initiatives. In their Executive Summary, Swan and Brown (2005) explain that the lack of faculty involvement in self-submission stems from the "perceived time required and possible technical difficulties in carrying out this activity." They (Swan and Brown, 2005: 6) then go on to explain the reason for the institution wanting repository: "an institutional open access archive provides a permanent record of all the research output of that institution' and consequently an 'archive can also serve as a marketing tool for the institution."

Auditing Questions

Only DUT IR is audited periodically, with statistics being sent to the library manager on a monthly basis. These statistics are then reviewed at meetings of the IR Working Group and at IR Advisory Committee meetings.

Auditing is a necessary process because doing so keeps the repository visible to management, who are responsible for providing adequate finances and personnel to keep the repository operating at optimum levels.

The statistics are useful in promoting the repository internally to management and researchers. Without backing from management, which can be increased by demonstrating that the repository is a showcase of the university's research, the repository is likely to be underfunded and understaffed. If the repository is not promoted to researchers, they are unlikely to support it by archiving their research. Therefore it is important to monitor usage and regularly audit these statistics.

Statistical Analyses

Currently the most comprehensive way of comparing open access repositories is by means of Webometrics rankings. These analyse repositories in terms of their size, visibility, availability of rich files (such as PDF documents), and scholarly content. According to the Webometrics Web site, "The Ranking Web (Webometrics) provides a list of mainly research-oriented repositories arranged according a composite index derived from their web presence and the web impact (link visibility) of their contents, data obtained from the major commercial search engines." (Webometrics). Aguillo et al. (2010: 478) explain that "Webometrics is helping to describe in a quantitative way the scholarly communication processes, focusing not on traditional citation databases (Web of Knowledge, Scopus) but on the information available in the publicly accessible Web." So a repository that is properly indexed by Google should perform well in the Webometrics rankings.

This is borne out by the results. DUT IR (which uses sitemaps, extracts data from rich files using media-filter, and publicises itself to harvesters) ranks 17 in Africa and 744 in the world, while KNUSTspace is 38th in Africa and 1,260 in the world.

This may seem surprising given that KNUSTspace offers vastly more content with its 5,881 items compared to DUT IR, which only has 800. But this may be due to DUT IR marketing itself more effectively on the harvesters used by the Webometrics system as well as providing the crawlers with site-indexes and providing a higher relative percentage of articles (as opposed to other types of content): KNUSTspace's articles comprise five percent of the total compared to 21 percent in DUT IR. Conversely ETDs in KNUSTspace account for 81% of all content, compared to 79% in DUT IR.

But not all content is desirable, for, according to Salo (2008: 13), “[r]epository managers understandably focus on filling the repository at all costs, since the easiest (though undoubtedly the least useful) measure of repository success is growth in collections (even empty ones) and items (even useless ones).” She goes on to state that this 'has led to repositories becoming an outright joke among university-press publishers.

Another measure of visibility is downloads and collection visits per month. Here KNUSTspace counts 4,435 downloads and 5,432 collection visits whereas DUT IR counts 21,046 downloads and 1,166 visits. Apparently DUT IR attracts a lot of attention from search-engine crawlers.

According to Salo (2008: 22), a “useful statistics system … must count accesses per author and per collection as well as per item and per file, and must provide information about these accesses over defined periods of time. It should track referring links and reader locations as well, so that authors can watch their work spread and participate in online conversations about it. It should also filter out otiose accesses such as those from search-engine crawlers.” Even with its Solr statistics, DSpace does not meet these requirements, and the situation is not remedied by Google Analytics, which fails to provide data at the item level. In addition the excessive downloads count recorded by DUT IR's statistics suggests crawler activity, which ought to be filtered out by the statistics package.

Going Forward

When asked to consider the short and long term goals of the repositories going forward, the following were the responses of each institution:

Ghana's KNUSTspace has no short or long-term goals, but considers strong administrative support as the main challenge.

South Africa's DUT IR considered lack of content as their most pressing challenge and listed involving the subject librarians in collecting content as their short-term goal, and increasing the visibility of the institution's research as their long-term goal.

Uganda's UIS listed lack of skilled staff to manage the repository, academics who do not want to submit their articles, and lack of policy as their most pressing challenges and intend to develop a logo and domain name in the short-term.

Recommendations and Conclusion

An institutional repository is a useless excrescence unless it is part of a systematic, broad-based, well-supported data-stewardship, scholarly-communication, or digital-preservation program (Salo, 33).

Based on the research findings, we conclude that researchers are unwilling to devote the necessary time to ensuring that their research is archived, while institutions seem unwilling to commit financially to ensure the sustainability of their repositories. This is surprising given that the literature on the subject is clear that the open access repositories are of primary benefit to the institutions that house them and their researchers.

If researchers are to be involved, the software must be made more usable for them. This can be accomplished by training interventions, but these are time-consuming for all involved. More immediately available solutions are the establishment of an online help wiki, or possibly a mailing list. Another way to involve researchers is to provide improved statistics that focus on the researchers' needs. For example, citation tracking is one area that would benefit researchers directly. If a researcher ID (such as ORCID) is used citation statistics could be provided to the researcher within the repository so they could immediately assess the impact their research. Unfortunately this level of software improvement is not likely to occur without developer intervention and, unless the libraries are able to develop the necessary skills themselves, they must rely on the community to do it for them. One way to ensure that it does happen is to provide funding to the organisations that do the software development.

However this is an area in which institutions' management seem particularly reluctant to invest since the software is free of charge. Perhaps one solution is for institutions to indirectly invest in the technology by purchasing add-ons or services from companies such as Atmire, whose primary focus is development of the core software.

The institutional involvement is not all bad by any means: there is evidence of their commitment to ensuring their repositories' sustainability in their implementation of policies and appointment of personnel. There is always room for improvement, however, such as in either improving infrastructure or encouraging central IT's involvement.

Recommendations

The following recommendations arise from the findings of the research:

Libraries should take proactive steps to make their repositories more successful. One initiative that seems very worthwhile is the development of a community of practice to assist one another. Stellenbosch University has made a great start in this, following on the early steps of the University of Pretoria, and it now remains for other institutions to get involved and ensure that African research is made visible and accessible to the world.

There are two alternative ways forward for repositories: Either institutions should give up hosting their own repositories and resign themselves to making their data available through professionally hosted repositories that will take care of the technical issues for them, or they should invest in developing the capacity of their repositories. If data sovereignty is important to an institution, then the latter route is the only viable one. Investment should start by implementing each of the steps listed in Stellenbosch's best practices guide.

With the above steps in place the repository can grow into something more than a showcase of the institution's research and consider becoming a repository of raw data, ultimately achieving the goals laid out in the Budapest Open Archives Initiative. But our analysis of the challenges faced by the four African open access repositories has shown that we still have a long way to go.

References

Aguillo, I. et al. 2010. Indicators for a webometric ranking of open access repositories. *Scientometrics*, 82: 477–486.

Anderson, B. 2004. Open Access and Institutional Repositories. *Behavioral Social Sciences Librarian*, 23(1): 97–101.

Association of Research Libraries. Framing the issue: open access. [Web document]. Washington, DC: Association of Research Libraries, Office of Scholarly Communication, 2004. [cited 16 December 2004]. Retrieved from http://www.arl.org/scomm/open_access/framing.html

Bhat, M. 2010. Open Access Repositories: A Review. *Library Philosophy and* Practice, 1–8.

Budapest Open Access Initiative.[Online]. Retrieved from http://www.budapestopenaccessinitiative.org/read [Accessed June 4, 2014].

Creaser, C. et al. 2010. Authors' Awareness and Attitudes Toward Open Access Repositories. *New Review of Academic Librarianship*, 16(S1): 145–161.

Gibson, H. SUNScholar/Practical guidelines for starting an institutional repository (IR). Retrieved from http://wiki.lib.sun.ac.za/index.php/SUNScholar/Practical_guidelines_for_starting_an_institutional_repository_(IR) [Accessed June 8, 2014].

Hashim, T. and Jan, T.R. 2011. Institutional Repositories: An Evaluative Study. *Trends in Information Management*, 7(2).

Hoorn, E. 2005. Repositories, Copyright and Creative Commons for Scholarly Communication. Ariadne, (45).[Online]. Retrieved from http://www.ariadne.ac.uk/print/issue45/hoorn [Accessed June 7, 2014].

Kyrillidou M. Journal costs: current trends & future scenarios for 2020. *ARL Bimonthly Rep 2000 Jun; 210*. Retrieved from http://www.arl.org/newsltr/210/costs.html. [Accessed 19 April 2006].

Max Planck Society 2003. Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities. Berlin Declaration.[Online]. Retrieved from http://openaccess.mpg.de/286432/Berlin-Declaration [Accessed June 8, 2014].

Pinfield, S. 2005. A mandate to self archive? The role of open access institutional repositories. *Serials*, 18(1): 30–34.

Prosser, D. 2004. The Next Information Revolution – How Open Access repositories and Journals will Transform Scholarly Communications. *Liber Quarterly*, 14(1/4).

Registry of Open Access Repositories.[Online]. Retrieved from http://roar.eprints.org/ [Accessed June 16, 2014].

Reich, V. and Rosenthal, D. 2003. LOCKSS: Building permanent access for e-journals: Practical steps towards an affordable, cooperative e-preservation and e-archiving. In: Proceedings of the 7th ICCC/IFIP International Conference on Electronic Publishing. ELPUB.[Online]. Minho, Portugal: Universidade do Minho, 23–23. Retrieved from http://elpub.scix.net/cgi-bin/works/Show?_id=0306&sort=DEFAULT&search=Reich%20Rosenthal&hits=3 [Accessed June 7, 2014].

Romary, L. and Armbruster, C. 2010. Beyond institutional repositories. *International Journal of Digital Library Systems*, 1(1): 44–61.

Salo, D. 2008. Innkeeper at the Roach Motel. *Library Trends*, 57(2). Retrieved from http://digital.library.wisc.edu/1793/22088 [Accessed June 4, 2014].

Suber P. Praising progress, preserving precision. *SPARC Open Access Newsletter*. [serial online]. 2004 Sep 2(77). [cited 19 June 2005]. Retrieved from http://www.earlham.edu/~peters/fos/newsletter/09-02-04.htm.

Swan, A. and Brown, S. 2005. Open access self-archiving: An author study. Retrieved from http://cogprints.org/4385/ [Accessed June 4, 2014].

Van Deventer, M. and Pienaar, H. 2008. South African Repositories: Bridging Knowledge Divides. Ariadne, (55) [Online]. Retrieved from http://www.ariadne.ac.uk/print/issue55/vandeventer-pienaar [Accessed June 7, 2014].

Wacha, M. and Wisner, M. 2011.Measuring Value in Open Access Repositories. *The Serials Librarian: From the Printed Page to the Digital Age*, 61(3-4): 377–388.

Welcome to the Handle System. [Online]. Retrieved from http://www.handle.net [Accessed June 16, 2014].

Webometrics. Retrieved from http://repositories.webometrics.info/en/Methodology