

https://dspace.org/

Institutional Repositories: Towards Harnessing Knowledge for African Development

Moahi, Kgomotso

2010-12-17
http://hdl.handle.net/10539/8950

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

Institutional Repositories: towards harnessing knowledge for African development

Kgomotso H. Moahi
University of Botswana
Moahikh@mopipi.ub.bw

Presented at the First International Conference on African Digital Libraries and Archives (ICADLA-1), Addis Ababa, Ethiopia, 1st-3rd July 2009

Introduction

Information and knowledge are the drivers of socio-economic development. According to a World Bank report, weakness in the application of knowledge is a major factor behind the economic stagnation in Africa. Compared with other countries, Africa has not had much success in acquiring and using knowledge for development. The application of knowledge is directly linked to the availability and access to information and communication technologies (ICTs) and in the African setting there are documented challenges in availability and access to ICT. However, the greater challenge in applying knowledge for development lies in the fact that although knowledge is generated in universities and research centres, it is either disseminated in expensive international journals, or gathers dust in the offices (and computers) of the generators, as well as those that have funded or commissioned the research. The tragedy is that after a number of years, such studies are replicated without the knowledge that they have been carried out before.

Much of the knowledge that is produced is in digital form as a result of the ubiquity of ICTs in many universities and research centres. However, the challenge is that the information and knowledge is not captured, organized for easy access and use by others. The application and use of information and knowledge can only become a reality where that information is collected, processed, and made visible for dissemination and use. This can only occur if developments in ICT are leveraged to develop digital libraries that can make African-grown knowledge visible. The trend worldwide has been to establish information repositories in order to make knowledge visible and accessible.

Generally, African universities are ranked lowest in terms of research output. According to the World University Rankings, the highest ranked university – the University of Cape Town – is ranked at 359 in the world out of 6,000 universities. The University of Botswana is ranked at 5375 in the world and 41 in Africa. A scholarly research presence online is one of the criteria used in ranking universities (Kgautlhe, 2009); clearly this indicates that the online presence of African universities is very low to say the least. This is something that institutional repositories can address and in this paper I make the case that the development of institutional repositories can ensure the connection of Africans to their own knowledge and information. The paper will

therefore consider policy issues as well as approaches and challenges to the development of institutional repositories.

What are Institutional Repositories (IRs)?

Bailey et al (2006) define an IR as means of collecting and providing access to diverse, locally produced digital materials. Donovan and Watson (2008) describe the IR as a means of collecting the intellectual digital outputs of an organization. An IR is a resource or system that facilitates the capture, storage, preservation and dissemination of the intellectual output of an institute in electronic form. Such output varies from institution to institution; some will capture theses and dissertations, whilst others will capture published papers, unpublished pre-prints, working papers, conference presentations, data sets, teaching materials and other similar material (Rosenblum, 2008). Whatever output is captured into an IR will be described using standard metadata formats and protocols, the tags that describe the output and enable recognition and retrieval of the output on the World Wide Web (Rosenblum, 2008). The goal of implementing an IR is mainly to have the intellectual output of an institution in a central source; some IRs will extend content beyond published materials to include others that may not necessarily be published, such as conference presentations, working papers, technical reports sand similar material; in short, grey literature that would otherwise be lost to organizations. IRs also provide access to others who may have an interest in the output, and they promote the visibility of an organization on the Internet.

Most IRs are based in colleges and universities, although they do to a lesser extent exist in government agencies. In most universities IRs are spearheaded and staffed by university libraries. When IRs were first mooted, the major consideration was to address the rising costs of serials and provide another avenue for scholarly research output. Some saw IRs as a means of showcasing the research output of institutions, whilst others saw them as a means of providing management and preservation of research output.

In order to implement an IR, facilitating software must be obtained. There are mainly three types of software: those based on open access; proprietary software; and software that is custom-built. The uptake of IRs in many institutions has been largely the result of the availability of open source software that can be used to build the IR (Van Westerienen and Lynch, 2005, quoted in Kingsley, 2008). The most widely used open access IR systems are DSpace and EPrints.

In order for an IR to be successful as far as these goals are concerned, there must be availability of content, which depends upon authors and researchers depositing output into the IR. Usually this is the hardest part of implementing IRs because researchers as stakeholders have to be brought on board to appreciate the benefits of having their output on an IR; they also need information, knowledge and skill for using the IR. Some writers have lamented the "failure" of IRs because of the dearth of content on existing IRs.

The case for Institutional Repositories

As a result of developments in digital scholarship, more and more scholars are creating content in digital form. That content may however end up in some out-of-reach expensive journal, or in the author's computer or even in a subject database, where it is really out of reach of most academics who would benefit from its use. In addition the content that is kept by authors is not guaranteed long-term preservation and curation. It is partly to address this challenge that repositories were first mooted.

According to Davis and Connolly (2007), there are a number of perspectives on the role and importance of IRs. Some view IRs as providing an alternative to the expensive journal subscription costs that make access to published materials extremely difficult; it is argued that, even though research may be funded by the institution or government, the fact that outputs are published in international journals, out of the reach for those institutions, provides a very strong case for implementation of an IR. A number of authors have urged researchers to publish in open access journals in order to force publishers to lower their subscription costs.

Others view IRs as complementary avenues for scholarly publishing and not necessarily competition for publishers. Lynch, cited in Davis and Connolly, argues that an IR is not a journal but rather an avenue for the collection and dissemination of grey literature. Markey et al. (2007), cited in Rosenblum (2008), state that there are different ideas about the primary role of an IR; some see the preservation of scholarship as the primary purpose; others see an IR as an assessment instrument to evaluate the output of an institution; and yet others see it as a means of increasing access and the impact of research. In reporting on the University of Boston IR, Taylor (2009) stated that the purpose of the IR was to be a cumulative and perpetual store for the intellectual output (of Boston); to be a support for research assessment; to be used for the creation of standardized CVs; for fostering wider dissemination and the impact of research; to provide increased marketing opportunities for the University by creating a tangible record of research; and to provide the possibility of identifying pockets of research within the University.

According to Rieger (2007) IRs facilitate a number of activities that include digital asset management; preservation of digital assets; ensuring the visibility of institutions; and facilitating discovery of content. IRs can also provide access to outputs of public research initiatives. There are many benefits that can accrue from the creation and implementation of IRs. One is that the institute or university will be able to maintain a centralized collection of its output. This obviously depends on the type of IR, and whether it is for research outputs (theses, dissertations, working reports, published papers and similar material) or for entire institutional output, beyond what is academic and research-based. This means that output of value, and what would otherwise be lost to the institution, is collated and made available for current and later use by members of the institutions and indeed others who may have an interest. Generally universities and research institutions produce considerable output material that may not be published and therefore accessible, and having an IR means that the grey literature that is produced in an institution can be compiled, preserved and disseminated as needed. The fact

that an institution maintains an IR, and uses metadata that enables discoverability on the WWW, means that the institution will become highly visible in terms of its output on the Internet. This visibility will then ensure that awareness of the institution is heightened and its activities known and, it is to be hoped, recognized.

The main driver for IRs in the African context has been similar to that of other countries in the west: rising costs of serials, limited library budgets, visibility and similar factors. What has been even more critical however has been the fact that research is funded by African and other donors and organizations at great cost, but the results are not shared. In this context, therefore, IRs can provide a means of ensuring that the output coming from Africa is registered and accessible on the Internet.

Currently it is said that Africa accounts for less than 2% of the research output of the world. Although there is clearly not enough, there is a significant amount of research activity in African universities but the stumbling block comes in publishing the work in scholarly journals. African academics strive to publish in internationally renowned peer-reviewed journals in order to ensure academic promotion. Not many of these academics do make it into such journals, and when they do, the journals are out of reach of most university libraries, rendering access difficult. Another factor is that with the scarcity of research resources (funding) it is necessary to avoid duplication of research. Because many academics keep their work in their computers, and do not necessarily have a forum for sharing the knowledge, this leads to considerable duplication. For African universities, therefore, an IR is seen to be a vehicle for enhancing visibility, access and impact because research outputs will be visible on the Internet and WWW.

Although IRs are a relatively new phenomenon, two registers of IR, ROAR (http://roar.eprints.org) and OPENDOAR (http://www.opendoar.org) report over 1000 IRs in the world; of those, 20 are in Africa, and most of these are in South African universities.

Approaches to implementation of IR

It is crucial to recognize and appreciate the fact that IRs are mainly about the users and the content rather than simply a matter of technology. It is therefore imperative to understand the demand side of institutional repositories, lest an expensive mistake is made to implement an IR that simply has no depositors or users.

Writing about the choice of IR model to select for preservation, Rieger (2007) identifies a number of processes that are required to enable uptake and implementation success: these include the identification of stakeholders and their involvement in the decisions concerning the selection of an IR model and its implementation; it also involves a needs analysis to determine what the IR should encompass. Most critically, it involves an understanding of the "existing human landscape", which involves understanding the organizational climate (culture, policies, governance issues, politics and goals). The points made above are critical: getting academics and scholars in universities to participate in IRs has been highlighted by a number of authors.

Getting academics to deposit their products or even to use the IR has been a challenge that has necessitated a number of activities aimed at informing and advocating for IRs. For example, Taylor (2009) states that the engagement of champions in each of the schools at the University of Bolton was carried out to try and ensure that the concept and practice of the IR were understood and that the IR would become "embedded" in the university research culture. Presentations were made again and again to faculty in order to ensure that awareness of the IR was maintained at all times.

Most authors writing about the creation and implementation of an IR highlight the need to establish a university-wide committee that will spearhead the whole process. Frequently it is assumed that the creation and implementation of an IR is the sole responsibility of academic librarians within the university; it is clear however that there is a need to involve other stakeholders, such as the IT staff in addition to academics and scholars. Such a committee should engage the university community on the need for an IR, and articulate the primary benefits of creating an IR, whether for published output or for all digital items created during the course of business of the institution.

The choice of software for implementing an IR is also important. There are a number of repository software packages that are available, and Laxminarsaiah and Rajgoli (2007) describe a few: Archimède, a Canadian software that supports multilingual implementations, developed by Laval University in Canada; CERN Document server software (CDSware, now known as CDSInvenio), which can handle large repositories; DSpace, software developed by the Massachusetts Institute of Technology (MIT) and Hewlett-Packard (HP), which facilitates the management of multidisciplinary content organized by community; E-Prints is developed by the University of Southampton and can be locally customized; it is also one of the most used software for repositories; and Greenstone, an open source software than can support multilingual documents.

Bailey et al. (2006), in a survey of 123 Association of Research Libraries (ARL) member libraries in the US, found that DSpace is the software of choice for many institutions because of the availability of technical support and its ability to support different formats of content. The fact that it has a large user group is also a major attraction.

Two approaches can be taken regarding the mode of output deposit into the IR: self-archiving and mediated archiving. Most institutions (for example, Bolton University and the University of Botswana) have opted to combine the two, realising that persuading academics to self-archive would always be a problem (as indicated in the literature about implementation of IRs at universities). Self-archiving is always preferred because of the advantages it offers:

- complete ownership of the process of depositing work;
- could become part of the research process;
- could mean that an item is represented in the way an author wishes it to be represented (Taylor, 2009).

However, there are also disadvantages because academics have to be trained on a number of aspects, including copyright and the correct version of a paper to load; the main challenge in the lives of busy academics, however, is that uploading their materials simply becomes one more thing to do.

Challenges

Contrary to what is believed, implementing an IR is not a matter of obtaining software and hardware, and waiting for content to flood in; it is more about the users and how they appreciate the need and use of an IR. Kingsley (2008) makes the point that IRs have not had as much success as discipline-based repositories because they are centralized systems where decisions about the implementation are imposed from the administration. Technical issues also come into play as a challenge and include matters such as the format of items to be deposited, and the fact that software versions change and may not allow retrospective use. This means that depositors may be asked to convert files to pdf format, which may be simple for some, but complex for most and definitely regarded as time-consuming. Learning to use the IR software, for both uploading and retrieving information can also present a significant learning curve.

A further challenge is the issue of copyright materials and the fact that authors who publish in journals usually sign copyright transfer forms that transfer copyright from the author to the publisher. Although publishers will allow depositing pre-prints or even the final print, many authors are never really aware of their rights and do not have the time to check what rights they have with regard to their published papers. According to Kingsley, very few academics know where and how to do this and it too may be considered as something extra in a busy academic schedule.

Yet another challenge in building IRs is identified in the literature as ensuring that the IR has content that grows. Problems that have been identified include the reluctance of authors to self-archive because of a number of factors: difficulties around intellectual property issues; learning to use the software; the fact that academics tend to see self-archiving as one more thing they have to do, especially if it involves checking on copyright, on the versions that they deposit, and getting the metadata complete and right. Other issues on the part of authors include a fear of plagiarism and having their ideas stolen, and confusion as to whether posting work on the IR is publishing (Davis and Connolly, 2007).

A lot of the challenges for the increasing participation of academics and researchers in depositing content are succinctly articulated by Fitzgerald and Austin (2008). They state that because academic progression of academics largely depends on their being published in high reputation, peer-reviewed journals, this tends to militate against their overwhelmingly positive response to depositing output on IRs. First is their belief that depositing output in the IR may reduce chances of further publication in reputable, peer-reviewed journals; secondly there is concern about possible re-use of their content, given the fact that IRs are generally open access; and thirdly, many authors and researchers are essentially ignorant about their

intellectual property rights and what they can negotiate with publishers: in general, authors tend to sign away their rights instead of negotiating for creative commons licenses. Fourthly, academic authors do not appreciate the impact that depositing in IRs will have on their academic profile and impact.

Clearly, in view of the above, there is great need to inform and educate academics and researchers so that they have a clear understanding of the role of IRs, how they relate to their rights and how they may advance their visibility. Donovan and Watson (2008) also point out that for repositories that accept all output, whether published or not, a challenge may be presented by authors who want to deposit everything and anything. Such researchers necessitate policies that control the intake of inappropriate, unwanted materials or content.

Sustainability of the IR is another important issue that may become a challenge: it may be easy to build an IR, but because it is based on technology that becomes obsolescent very quickly, an institution must bear in mind the costs that will be associated with long term preservation of research output.

Policy issues

In order to ensure that IR uptake and use is achieved, there are certain policy considerations that are necessary. First is the issue of interoperability, which ensures that outputs are discoverable. Most IRs have adopted the existing metadata protocol OAI-PMH. Second, it is necessary to ensure that the IR is available and accessible at all times, so as not to demotivate authors who deposit materials. This means that the technology must be robust and not prone to system problems from time to time. This is particularly crucial in Africa, where the technology does tend to be unstable. Third, the system used for deposit and searching for output must be simple and user-friendly, so as not to present too steep a learning curve.

Clearly policy is crucial for setting the parameters of the system. Rieh et al. (2008) caution that policies must take into account stakeholder needs, and existing research practices. The University of Botswana, which is in the process of creating an IR entitled the University of Botswana Research, Innovation and Scholarship Archive (UBRISA), has developed a policy as well as operational guidelines for depositors.

The creation of UBRISA at UB is closely tied to the goal set through the University Strategy to be a research intensive university by 2021. Clearly, if the University is to become research intensive, it must have something to illustrate that intensity. There are many ways, but having an institutional repository that records all research activities is one. The policy is that all research outcomes, published or not, should be deposited in the IR. The objectives of the IR are as follows:

- To promote and encourage the dissemination of research findings;
- To increase the level of African content in scholarly publications unduly dominated by Western academic discourse;

- To enhance socio-economic development through research that feeds into national systems of technology transfer and innovation;
- To strategically increase the visibility of UB nationally and internationally in scholarship and knowledge creation, application and exchange;
- To preserve the University's intellectual heritage for future use. (University of Botswana, 2009).

UB envisages that UBRISA will be the ultimate source of research outputs and will aim at encouraging staff to deposit their materials by offering a prize for a staff member who has the highest number of outputs recorded in the IR in any given year.

Policy issues that should govern the implementation of IRs include the following, also articulated in UBRISA:

Content, which states what content will be captured in the repository. As indicated earlier, different institutions have different content policies which may accept only published materials, theses and dissertations, or even unpublished material. Institutions such as UB will identify the elements that qualify the output to be deposited, which include: work produced, contributed, or sponsored by, or by association with, a UB faculty, centre, school or department; output must be produced by a member of the UB community; the work must be complete and ready for distribution prior to submission (University of Botswana, 2009).

Submission, which states the processes to be undertaken during submission of the output and also stipulates whether submission shall be made directly by the authors (self-archiving) or shall be mediated by designated individuals. The processes include quality assurance issues where only items that have undergone some form of refereeing within the institution will be accepted. Submission also states that authors should check copyright status before depositing their output.

Preservation, which indicates how long content will be retained in the IR. Some institutions provide a fixed term such as 10 years, and others will retain content in the IR in perpetuity. This means that migration to new formats will be carried out; software emulations will be provided to access materials that could not be migrated.

Withdrawal outlines the circumstances under which withdrawal of content may be possible. These include where there is doubt about the originality of the output dues to plagiarism, copyright infringement, falsification of research results, and similar circumstances.

Interoperability and Open Access stipulates that access to content will be open and metadata to be used will therefore be open archive compliant to ensure the worldwide visibility of research outputs.

Use and Benefits indicates the benefits of depositing and using material from the IR. These include enhanced citations, increased availability of research information, and advancement of scholarly research through quick and easy access.

Conclusion

Institutional repositories have been shown to be an important part of a university or research institution in that they make possible a central location for an institution's output and in the process enhance the visibility of the institution. IRs also provide an avenue for the preservation and archiving of an institution's research output. In the African context, implementation of IRs would go a long way towards making research from Africa and about Africa visible; it is therefore crucial for African universities and research organizations to develop institutional repositories as a matter of urgency. However, as much as IRs have advantages, there are challenges in implementing them that mainly hinge on their acceptance by the contributing authors. These challenges have been articulated in this paper and policy considerations have also been highlighted. For African universities and research organizations, there may of course be further challenges of a technical nature as well as attendant skills and resources required to make repositories a success. My conclusion, however, is that IRs are not a luxury but a necessity if Africa is to show the rest of the world that they do contribute to the worlds knowledge base.

References

Ashalatha Laxminarsaiah & Iqbalahmad U. Rajgoli (2007). Building institutional repository: an overview. OCLC Systems & Services: International digital library perspectives, vol. 23 no. 3

Bailey, Charles, W., Coombs, Karen, Emery, Jile et al. (2006). Institutional Repositories SPEC kit 292. Association of Research Libraries, Washington D.C. http://www.arl.org/bm~doc/spec292web.pdf Accessed 27/05/09

Davis, Phillip, M. and Connolly, Matthew (2007). Institutional Repository: evaluating the reasons for non-use of CU's installation of DSpace. D-lib Magazine, March/April 2007, vol. 13 no.3/4

Donovan, James, M. and Watson, Carol, A. (2008). White Paper: Behind a law school's decision to implement an institutional repository. Alexander Campbell King Law Library: University of Georgia School of Law. http://digitalcommons.law.uga.edu/law_lib_artchop/15 Accessed 27/05/09

Fitzgerald, Brian & Austin Anthony (2008). The OAK Law Project: issues for academic authors, institutional repositories, open access journals and end-users. Open Repositories Conference – 1 – 4 April 2008, Southampton.

Johnson Ian, M. & Copeland, Susan, M. (2008). OpenAIR: The development of the Institutional Repository at the Robert Gordon University. Library High Tech News, vol. 25, no. 4. May

Kennan Mary Anne & Wilson Concepcion (2006). Institutional repositories: review and an information systems perspective. Library Management, vol. 27, no. 4/5. 2006.

Kgautlhe, Calviniah (2009). Where is Research at UB? The Botswana Gazette, Wednesday 24-60, June.

Kingsley, Danny (2008). Those who don't look don't find: disciplinary considerations in repository advocacy. OCLC Systems & Services: International digital library perspectives, vol. 24 no. 4

Markey, Karen, Soo Young Rieh, St Jean, Beth, et al (2007). Census of IR in the US. Council on Library and Information Resources

Rieger, Oya, Y. (2007). Select for success: key principles in assessing repository models. D-lib Magazine, , vol. 13 no. 7/8, Jul/Aug.

Rieh, Soo Young, St Jean, Beth, Yakel, Elizabeth, Markey, Karen & Kim Jihyun. (2008). Perception and experiences of staff in the planning and implementation of IR. Library Trends, vol. 57 no. 2, Fall.

Rosenblum, Brian (2008). Developing new skills and expertise to support digital scholarship and scholarly communication. IFLA WLIC, Quebec, 74th IFLA General Conference and Council, 10-14 August 2008

Taylor, Sarah (2009). JISC final report. University of Bolton IR (UBIR). Final Report, April 2009. http://digitalcomment.bolton.ac.uk/ubiradmin/9 Accessed 27/05/09

University of Botswana (2009). Digital Repository Policy. University of Botswana Research, Innovation and Scholarship Archive (UBRISA).