The Role of Institutional Repository in Digital Scholarly Communications

P. Jain\textsuperscript{1}\textsuperscript,* G. Bentley\textsuperscript{2} and MT Oladiran\textsuperscript{3}

\textsuperscript{1}The Department of Library & Information Studies
\textsuperscript{2}Library Services
\textsuperscript{3}Department of Mechanical Engineering
University of Botswana
P/Bag 0022, Gaborone, Botswana

Abstract

In this electronic publishing age, academic institutions including universities have increasingly recognised that an institutional repository (IR) is an essential infrastructure of scholarly dissemination. An IR is a digital research archive consisting of accessible collections of scholarly work that represent the intellectual capital of an institution. It is a means for institutions to manage the digital scholarship their communities produce, maximise access to research outputs both before and after publication and also to increase the visibility and academic prestige of both the institution and authors. This paper discusses the benefits and obstacles of setting up an IR, and librarians’ and authors’ roles in the successful management of an IR. As the strength and advantages of DS can only be maximised through effective IRs, the paper recommends that an IR be considered as the principal benchmark of digital scholarship.

Key words: Institutional Repository, Academic Institution, Digital Scholarship, Librarians & authors in IR.

*: Corresponding author

Introduction

Information and Communications Technology (ICT) continues to transform the scholarly environment and management of higher education institutions. For example, ICTs are core resources required for digital publishing and online teaching and learning. ICT has created platforms and opportunities for scholars to work collaboratively through extensive infrastructures, with access to resources and knowledge services in borderless environments. The rapid growth of digital assets creates challenges in the use, management, archiving and application of digital information and datasets. That is why digital scholarship is the fastest growing academic phenomena today. Digital Scholarship (DS) has been popularly defined as “any element of knowledge or art that is created, produced, analyzed, distributed, published, and/or displayed in a digital medium, for the purpose of research or teaching” (Kirsten Foot, Assistant Professor, Department of Communication, University of Washington). However, this definition does not seem to capture the totality of digital activities in any tertiary institution. Therefore DS can be defined as “an integrated collaborative blended environment which embraces cutting-edge technologies in learning, teaching, research, professional and administrative services” (Task Group on UB and Digital Scholarship, 2008).

Recently academic institutions have been grappling with how to manage the digital intellectual output they produce including journal articles, conference papers, reports, theses & dissertation, teaching materials, artwork, research notes, and research data. Clearly, technology has made it easy to create, store and access digital material. Paradoxically however, while there is potential for instantaneous access, all too often many materials are not usually made accessible to many users and they remain marooned in the authors’ computers. About 80-85% of digital intellectual output of universities is never made accessible to the public (The Open Citation Project, 2004). Also the escalating costs of online journals prohibit subscription and it is becoming more unrealistic and challenging for libraries to subscribe to all, or even most of the online academic journals (Warren, 2003).

In response to the above mentioned conditions, Massachusetts Institute of Technology (MIT) announced a research project titled DSpace “to build a stable and sustainable long-term digital storage repository that provides an opportunity to explore issues surrounding access control, rights management, versioning, retrieval, community feedback, and flexible publishing capabilities” (DSpace Project, 2000).
DSpace is not a completely new concept because ‘Preprint’ archives in the sciences were already in existence as a form of innovation managed by the data creators (Smith, 2003). The research archive for physics, mathematics and computer science ‘arXiv’, is now 14 years old, and holds around a third of the global physics research output. Thus scientific communities of knowledge developed and managed relevant information nodes on the Web to speed communication in time and across space. Because of open archive initiatives it no longer matters whether documents are stored in one central discipline based archive or in many distributed ones. They are all interoperable and harvestable into one virtual global archive, in which all contents are seamlessly navigable and retrievable (Smith, 2003).

The enterprise of DSpace and the other technical solutions that came to prominence around the same time is that they are ‘institutional’ in scope. Individual institutions have accepted ownership rights and responsibility for management and preservation of their local scholarly corpus. Universities also want “to address the issues of repatriating their scholarly work from commercial publishers and providing long term, secure and open access” (Westell, 2006). Since the launch of DSpace, institutional repositories have sprung up at academic institutions across the world and more are appearing regularly.

Various authors have defined Institutional Repositories (IRs) but the frequently cited definition is that of Lynch (2003), viz. “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 institution and its community members. It is most essentially an organizational commitment to the stewardship of these digital materials, including long-term preservation where appropriate, as well as organization and access or distribution”. Johnson (2002) defines an IR simply, as “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”. The essential characteristics of an IR are that it is institutionally defined, scholarly in scope, cumulative and perpetual, open and interoperable (Crow 2002). Deposit within an IR is carried out in order to maximize the visibility and accessibility of comprehensive, local research. IRs are beneficial to both the researcher and the researcher's institution (Lynch 2003). An IR sits firmly within the DS landscape, which includes “building digital collections, creating tools for collecting, analyzing, and authoring digital information, and using digital collections and analytical tools to generate new intellectual products” (American Council of Learned Society, 2006). Therefore, IR can also be considered as a benchmark of Digital Scholarship.

IR as a major benchmark of Digital Scholarship

IR proponent Lynch (2003) regards institutional repositories as essential infrastructure for modern scholarship. He argues that “the development of institutional repositories emerged as a new strategy that allows universities to apply serious, systematic leverage to accelerate changes taking place in scholarship and scholarly communication”. Also, Markey et al (2007) remarked that, “a considerable portion of the scholarly record is born digital, and some scholarship is produced in digital formats that have no physical, in-the-hand counterparts. The proliferation of digital scholarship raises serious and pressing issues about how to organize, access, and preserve it in perpetuity. The response of academic institutions has been to build and deploy institutional repositories (IRs) to manage the digital scholarship their learning communities produce”.

All the above definitions of an IR agree that it is an enabling component of digital scholarship and this paper works on the premise that IR is currently a cornerstone of DS and an important tool to manage institutional intellectual output. This seems to suggest that the success of DS can at least partially, if not fully depend on and be measured by the success of IRs.

IR concepts and digital scholarship

There are specific concepts integral to digital scholarship that an IR does or potentially can characterise, both in concept and functionality. These are interoperability, research management and new forms of publishing. They further support the assertion that an IR is currently a major benchmark of digital scholarship.

Interoperability

The Open Archives Initiative (OAI) designed a shared code for metadata tags (e.g., ‘date’, ‘author’, ‘title’, ‘journal’ etc.). So full-text documents may be in different formats and locations, but if they use the same metadata tags they become interoperable. It means that their metadata can be harvested and all the documents can then be jointly searched and retrieved as if they were in one global collection, accessible to everyone (Open Archives Initiative 2002). Signing up to OAI compliance (by choosing for example open access, open source platforms) and committing to interoperability with other repositories indicates an ability and openness on behalf of the institution to contribute to global scholarship. Opening the
repository to cross repository searching tools (OAI and other harvesters as well as the generic indexing
intools like Google Scholar) exposes scholarship in a new way and puts it in an international context
(Westell 2006). The institutional repository may be the only vehicle to attract an international audience to
the institution.

Another type of interoperability is integration of the IR with course management tools. There is potential
to extend the objective and content of an IR beyond the ‘purist’ content, viz: copies of peer reviewed
research articles, to embrace classroom, and distance teaching and learning materials and offer open
educational resources. Arguably these developments have evolved from the open access culture
proselytised first by digital research repositories.

**Linkage with digital research management systems**

Some developers see potential for developing IRs beyond their current capacity, even believing that the
current repository concept is flawed (Stuart, 2008). They posit that institutions should provide a system
for ‘work in progress’ complete with descriptive data to associate it with the appropriate research grants,
researchers, departments and funders. The current conceptual ‘institutional repository’ should be a slice
in this data-corpus, as a logical front end for a system which manages the entire research process,
rather than an independent initiative. In effect IR functionality would be part of a current research
information system (CRIS), bringing together the information that underpins the complete process of
research, from grant application up to and including peer-reviewed publications.

**New publishing and communications models**

Discussions concerning IRs as a possible alternative publishing model split IR adherents into different
camps, between innovators and purists. One model frames deposit in a repository as an adjunct, and
complementary to, the traditional publication process (Hunter, 2007). Lynch (2003) for example firmly
believes “it underestimates the importance of institutional repositories to characterize them as
instruments for restructuring the current economics of scholarly publishing rather than as vehicles to
advance, support, and legitimize a much broader spectrum of new scholarly communications”. The other
group sees repositories as the beginning of new forms of academic publishing (Hunter, 2007). For
example, in an interview about the University of California (UC) eScholarship repository, Candee (2006),
the director of publishing and strategic initiatives in the Office of Scholarly Communication, agreed that
the eScholarship Repository had effectively become a publishing platform for the University of California
Press. She further added, “I think it is unfortunate that the term institutional repository has come to mean
something narrower... ultimately I envision a very different arrangement between universities and
publishers than we have now”. The UC vision is that universities will be in control of the publishing
process, with all content hosted and managed by universities themselves. It is possible that in future IRs
will be regarded as the principal route for the dissemination of research papers.

From the ongoing discourse, it is apparent that there is at least complete agreement on how an IR can
complement analogue and e-traditional forms of publication. All parties accept that the revolution in
scholarly communications means that traditional forms such as journal articles now frequently include
supplementary datasets and analysis tools (Lynch 2003). Then there are products of what Smith (2003)
calls ‘new model scholarship’: ‘digital resources documenting contemporary actions and ideas, wholly
new types of information resources, so novel that no common term except ‘digital objects’ or ‘sites’ can
describe them, and which fall outside all traditional modes of publications”. Scholarship has become
data intensive; it is supported and documented by data and tools that complement interpretive works of
authorship. Institutional repositories can support these new manifestations of scholarship that
emphasize data as an integral part of the record and discourse of scholarship (Lynch 2003).

**Benefits of an IR**

The benefits of repositories to institutions and individuals are numerous and can be grouped into the
following categories (Pickton & Barwick (2006)):

**Specific to the University, an IR offers:**

- increasing visibility and prestige. A high profile IR may be used to support marketing activities
to attract high quality staff, students and funding,
- centralisation and storage of all types of institutional output, including unpublished literature,
- support for learning and teaching. Links may be made with the virtual teaching environment
and library catalogues,
- standardisation of institutional records. The compilation of an ‘Institutional CV’ and individual
online dossiers linked to the full text of articles becomes possible,
- ability to keep track of and analyse research performance,
- breaking down of publishers’ costs and permissions barriers.
• alleviation of requirement to trust publishers to maintain information in the long term, without any commercial benefit for the authors.
• promotion of a philosophy of wider communication.

Specific to authors, an IR enhances:
• dissemination and impact of scholarship. Some studies have estimated that open access articles are cited 50% to 250% more than non open access articles. In some disciplines, online files receive on average 300% more citations than materials available only in paper format (The Open Citation Project, 2004). Also, Google Scholar gives preferential treatment to materials in IRs; a paper picked up from an IR would appear higher up on the Google results list (Ashworth 2006).
• storage and access to a wide range of materials. Many authors lack time, resources, or expertise to ensure preservation of their scholarly work. Through an organizationally based IR strategy, long-term accessibility and greater security of work is assured. Research items get a permanent URL compared to a personal or departmental web site.
• feedback and commentary from users. Authors are able to receive and respond to commentary on ‘pre-prints’.
• added value services; such as hit counts on papers, personalised publication lists and citation analyses.
• a central archive of a researcher’s work.
• a researcher’s profile.
• benefits to researchers and their institutions in terms of prestige, prizes and grant revenue.

Specific to society/community:
As scholarship is shared, society at large benefits. Maximising public access to research findings online, in turn maximises its visibility, usage and impact. It also maximises its benefits to research itself (and hence to the society that funds it) in terms of research dissemination, application and growth, research productivity and progress. (ePrints website FAQ, n.d.). Also more sponsors of funded research now have mandates for authors to deposit their articles and other research outputs as a condition for funding. Some policies promote Open Access for funded research. These requirements are intended to increase readership, re-use and dissemination of research outputs. The message to researchers is that research is incomplete until the output is widely disseminated.

To summarise, the potential uses of an IR are: scholarly communication; management and storage of learning materials, electronic publications and research collections; preservation of digital research work; building university profile by showcasing academic research work; providing an institutional leadership role for the library; research assessment; encouraging open access; and housing digitised collections (Barton and Waters, 2004).

Challenges of an IR

Despite the numerous benefits of an IR, there are implications and potential barriers to its success as summarised below (Pickton & Barwick, 2006):
• **Cost:** The initial financial cost for an open source software adopted by most institutions for creating IRs is not high but the recurrent costs, especially staff costs (e.g. time spent drafting policies, developing guidelines, publicising, training, supporting users and creating metadata, specialist IT consultancy) may be significant. This is further discussed below.
• **Difficulties in generating content:** A successful IR depends on the willingness of authors to deposit their work voluntarily and there may be local barriers and hindrances to be overcome. There are acknowledged difficulties in generating content, especially at the beginning. Unless the value of an IR can be demonstrated quickly, the organization’s long-term commitment to the project may begin to wane. The best way to prove the enduring value of the IR and to ensure its long-term survival is to quickly populate it (Gibbons, 2004).
• **Sustaining support and commitment:** Far too often, it is difficult to sustain continuous support and commitment from the management and academic staff. Lynch (2003) has succinctly described this obstacle: “Stewardship is easy and inexpensive to claim; it is expensive and difficult to honour, and perhaps it will prove to be all too easy to later abdicate”. There is a need for institutions to think seriously before launching institutional repository program as it may disintegrate rapidly if not properly managed.
• **Rights management issues:** Sometimes researchers are apprehensive about infringing publishers’ copyright and lack adequate awareness about their own intellectual property rights. They may be uncertain about making their work available online before it is published by a traditional publisher.
• **Working Culture issues**: Contributing content to user-generated or ‘self-service’ sites is time-consuming; and time is something which academics often lack. They may be willing to contribute content but reluctant to do it themselves. This calls for mediated deposits service for them.

• **Policy Issues**: Experiences suggest that an IR will only function to its capacity when a mandate is in place to populate it but clearly researchers can react negatively to any suggestion of compulsion. Lynch (2003) has cautioned that an IR should not become a tool for enforcing administrative control over academic work.

• **Lack of incentives**: In the absence of any incentive academics feel reluctant to provide even bibliographic details of their scholarly output especially when they know that incentives are available in other institutions.

The challenge therefore in implementing an IR is to promote the benefits it offers while allaying stakeholders’ concerns.

**Role of librarians in an IR**

Pro-activity and responsibilities relating to IRs are assumed by different people in various institutions. Largely they will be undertaken collaboratively by officers within the library in partnership principally with research and development, and information technology sections. Stimulating engagement for buy-in is crucial in the early stages of an IR when efforts are made to build a critical mass of material. Nixon (2002) rightly observed that “Reference librarians are a library's eyes and ears. They understand users needs and perceptions. They know what's working and what's not. When they act as subject selectors, they are the library's primary liaison with faculty in their subject areas and its most visible representatives. They know how to help, inform, persuade, and teach users. For an IR to succeed, it is essential that they be involved in its planning, implementation, and operation.” So librarians have critical roles to play in both establishing and maintaining an IR through:

- **Advocacy.** Librarians need to know all about the IR, its principles, benefits and operational processes in order to promote it and act as ‘IR evangelists’ (Ashworth 2006). Librarians will need to develop advocacy programs, publicise IR through institutional news media and respond to questions by the stakeholders.

- **Building content.** Librarians can employ advocacy and marketing strategies to promote engagement with faculty members and help to generate content. They can also assist by proactively searching for content independently.

- **Collection administrators and metadata specialists:** Librarians have potential roles as collection administrators and metadata specialists. For effective implementation of IR, libraries will need to recruit or train librarians with digital collection management and provide a mediated deposit service for reluctant ‘self-archivers’.

- **Training:** Librarians should be able to train staff and students to use the IR and help them prepare their digital products.

**Role of authors in an IR**

Digital scholarship is the most dynamic academic phenomenon today. However IRs are comparatively new to much of the academic world, particularly in developing countries. According to Westell (2006) the concept of archiving scholarly outputs of a university and making them available in the context of the institution is one that scholars and administrators are still coming to terms with. Equally then, the real challenge is not the technical implementation of the IR but rather the cultural change necessary for it to become embedded in the activities and normal behavioural pattern of researchers (Chan et al. 2005). The roles of authors include:

- Uploading of research output.
- Responding to questions and comments posted by readers.
- Updating materials especially pre-print articles and work in progress.
- Ensuring high quality and standards of materials.
- Negotiating copyright issues with publishers. This should be done before publication of papers.

**Quality assurance**

Establishing a functional IR can facilitate growth in research, scholarly communication, academic collaboration, strategic alliances, and teaching and learning. To achieve these laudable goals, quality issues must be prominent from the design stage, through installation and operation of the IR. Comprehensive procedures must be established, well understood and also reviewed periodically in line with changing DS platforms. Quality standards must cover various processes including system and
services procurement, and contents collection, submission, preservation, retrieval and re-use. Also copyright is a quality benchmark for IRs and repositories must operate in legally sound environments.

The three main quality players are the authors, IR administrators and senior management. Populating an IR is vital in assessing its performance and it is easier for authors to upload their research outputs directly into the system. In so doing they will be responsible for the quality assurance and copyright of their materials. Pre-print articles would have been subjected to the usual peer review publication process of journals. The IR administrator is responsible for the system performance and would have to facilitate drafting of policies to guide operation of the IR. A repository policy may cover a number of issues e.g. self – archiving and handling of non-refereed research outputs. Without a firm policy to guide authors repositories may remain virtually empty but with a mandatory policy they will be populated more readily (Swan, 2006; Sale, 2006). The administrator may also organise mediated deposit services for some authors to encourage buy-in. The top management of universities must be committed to matters of IRs by approving policies and operating procedures and providing resources for operating and updating the IR. Issues of Open Access must be decided at the top management level.

Economics of setting up an IR

Current literature indicates that IRs are relatively quick and inexpensive to establish (OpenDoar 2008, DRIVER 2008). There are however some fiscal requirements for starting and maintaining an IR.

- **Software** – Although many IRs employ free Open Source software e.g. Eprints software and Dspace, provision must be made for modifications to the repository software.
- **Hardware** – A dedicated server with sufficient capacity must be included in the budget.
- **Staffing** – Administrator & other supporting personnel for populating the IR and time for developing several policies to manage the system.
- **Running (i.e. O & M) costs** – this element captures how the IR would be operated and maintained. Issues will include system maintenance, upgrades and training of staff in situ or elsewhere.
- **Marketing & Advocacy** – the success of an IR may depend on the robustness of advocacy and public relations strategies. A well developed program of activities for buy-in of academics and senior management must be an ongoing event. The momentum of setting up the IR must not fizzle out soon after commissioning.

The Primary Research Group (2007) recently carried a study of 56 institutional digital repositories from eleven countries. Findings show that on average $78,802 (US) was spent as start up costs and about 41% of survey participants purchased software to develop their digital repositories. However most of the US institutions bought commercial software instead of open access ones which are becoming popular in other regions. For institutions especially in developing countries planning to establish research repositories, the concept of “think big but start small” should be highly considered.

Institutional repositories in Africa

The worldwide development of institutional repositories has been phenomenal in the last three years. Fig 1 shows the annual growth of IRs in different regions of the world. Europe and North America have the highest concentration of IRs and it seems that any institution of repute in the two regions will also have an IR. Fig 2 shows the proportion of IRs in various continents. In spite of the above scenarios, what is the state of IRs within the African context? Figures 1 and 2 illustrate that development of IRs is very low in Africa. The question therefore arises, if institutional repositories are not evenly widespread can they be rightly and fairly considered as a major benchmark of DS? Is an IR simply an aspiration rather than a realistic benchmark in the African context? Since African repositories have an extra regional imperative, namely to get African research on the international scene, arguably IRs should be an even more important benchmark in the African context. Currently, thirteen out of the nineteen existing repositories are maintained by South African institutions, with the others in Egypt, Kenya, Namibia, Uganda and Zimbabwe. Of them all, the University of Pretoria is the only institution with a well established repository, judging by the number of item records (over 2000). The other repositories have less than 1000, mostly below 500. This may be an indication that due to a lack of resources Africa has been slow in IR initiatives and implementation.

However, a useful refinement of this data is to look at the growth rate of repositories within Africa as shown in Fig 3. From fig. 3 it is clear that Africa has moved from a single repository prior to year 2006, 13 by the end of 2007, 19 in 2008 and there are others in the pipeline. This significant growth
demonstrates that Africa is embracing the IR concept, and shows that IRs are strategically important for institutions, making them a benchmark for their digital scholarship initiatives.

Fig 1. Growth of Institutional Repositories in different regions
Lynch (2003) emphasised that at the most basic level, an IR is “recognition that the intellectual life and scholarship of our universities will increasingly be represented, documented, and shared in digital form, and that a primary responsibility of our universities is to exercise stewardship over these riches”: both to make them available and to preserve them. IR technical platforms were created by universities to fulfil precisely the scholarly requirements of academic institutions. Part and parcel of digital scholarship is to exploit the available technology and moving with the times to create an infrastructure to support and develop teaching, learning and research. The very currency of an IR makes it a major benchmark of Digital Scholarship.

Figure 3: Growth of repositories in Africa.

It will be interesting to monitor the extent of growth of IRs as well as their place in the digital scholarship over the next several years; whether they will settle into the niche originally envisaged for them, and in which they fit well, or expand to extend their compass towards the totality of the research process, research analysis and performance management, and as a new model for scholarly publication and peer review. Both models have the potential to permanently and radically alter the landscape of scholarly communication.

References


