Digital Library Perspectives
Digital Library Perspectives is a refereed journal which aims to provide wide-ranging coverage of developments in digital libraries and digital repositories, and the web-based delivery of cultural content.

EDITOR
Bradford Lee Eden, PhD
Dean of Library Services, Christopher Center Library, Valparaiso University, 702 E. Chicago Street, Indiana 46383, USA
E-mail brad.eden@valpo.edu

ACTING EDITOR
Dr Jeonghyun (Annie) Kim
Associate Professor, Department of Information Science, University of North Texas, USA

REGIONAL EDITOR
Pedro Hipola
Research Manager, International University of La Rioja, Spain

ISSN 2059-5816
© 2018 Emerald Publishing Limited

Guidelines for authors can be found at:
www.emeraldpublishing.com/journals/dlp/helpforauthors.htm
EDITORIAL ADVISORY BOARD

Toby Burrows  
Principal Librarian, Scholars’ Centre, University of Western Australia Library, Australia

Pascal V. Calarco  
University Librarian, University of Windsor, Canada

Dr H. Frank Cervone  
Director of Information Technology, College Information Security Officer and HIPAA Liaison, School of Public Health, The University of Illinois at Chicago, USA

Sally Jo Cunningham  
Associate Professor of Computer Science, University of Waikato, New Zealand

J. Stephen Downie  
Associate Dean for Research, Graduate School of Library and Information Science, University of Illinois at Urbana-Champaign, USA

Eugene Fieg Jr  
Cataloguer, Claremont School of Theology Library, USA

Robert Fox  
Library Information Systems – Hesburgh Libraries, University of Notre Dame, USA

Victor Herrero-Solana  
Professor, Information and Communication Department, Universidad de Granada, Spain

Barrie Howard  
Adjunct Faculty, Department of Library and Information Science, The Catholic University of America, USA

Norm Medeiros  
Associate Librarian of the College, Magill Library, Haverford College, USA

Plamen Miltenoff  
Professor, St. Cloud State University, USA

Panayiota Polydoratou  
Assistant Professor, Department of Library Science and Information Systems, ATEI of Thessaloniki, Greece

Joanna Richardson  
Library Strategy Advisor, Griffith University, Australia

Adam Sofronijeniv  
Deputy Library Director, University of Belgrade, Serbia

Bogdan Trifunovic  
Director, Public Library Cacak, Serbia

Frederick Zarndt  
Global Connexions, USA
Distributed digital preservation: preserving open journal systems content in the PKP PN

Bronwen Sprout
University of British Columbia, Vancouver, Canada, and
Mark Jordan
Simon Fraser University, Burnaby, Canada

Abstract

Purpose – This paper aims to discuss the public knowledge project (PKP) preservation network (PN), which provides free preservation services for eligible journals by collecting article content and preserving it in a network of (at the time of writing) eight “preservation nodes” using the LOCKSS system. The PKP PN was launched in June 2016.

Design/methodology/approach – This paper addresses the development and implementation of a free, distributed digital PN for open journal systems (OJS) content. It discusses challenges in developing the network, in particular relating to preserving content from a set of partners who have no formal business relationship with PKP. The paper examines data regarding journals that have opted in to the network to date and considers interface usability and other barriers facing those that have not joined.

Findings – Within 18 months of launch, more than 600 journals had opted to be preserved in the PKP PN. Many more journals are eligible to join the network; the paper explores potential strategies to increase participation and identifies and proposes methods to overcome technical and communication barriers.

Originality/value – This paper describes a highly collaborative, open-source preservation initiative which forms a unique part of the e-journal preservation landscape and preserves a particularly vulnerable portion of the scholarly record.

Keywords Partnerships, Open journal systems (OJS), Digital preservation, Open source software, Open access journals, Private LOCKSS network (PLN)

Paper type Case study

Introduction

In June 2016, the public knowledge project (PKP) announced the launch of a preservation network (PN) to digitally preserve journals that publish using the open journal systems (OJS) software. The vast majority of the estimated 10,000+ journals published using OJS are not preserved by a recognized digital preservation service. The PKP PN meets this need, providing free preservation services by collecting article content from any eligible OJS journal and preserving it in a network of (at the time of writing) eight “preservation nodes” based on the lots of copies keeps stuff safe (LOCKSS) system. After a trigger event – when the original journal is no longer accessible permanently – the content will be made publicly available in one or more “access nodes” running OJS.

The PKP PN, which is a free service provided by PKP and its partners, was launched in June 2016 and nearly 18 months later, almost 600 journals are being preserved in this worldwide PN (joining at the rate of about one per day). While this adoption is not insignificant, it represents only a small portion of OJS journals, many of which are open access and published by small publishers, and at particular risk of loss. Factors such as OJS...
user interface and workflow design as well as awareness of the service and of preservation risks all play a part in journal participation in the network. The authors discuss the challenges encountered in developing the PN as well as strategies for increasing participation.

**Context for the public knowledge project preservation network**

The PKP is a multi-university initiative based at Simon Fraser University in Vancouver, British Columbia, which develops free open source software and conducts research to improve the quality and reach of scholarly publishing. It is perhaps best known for developing the OJS software, which was launched in 2002 and is now the world’s most widely used journal management and publishing system, made freely available worldwide for the purpose of making open access, peer-reviewed publishing a viable option for more journals[1]. PKP was established in 1998 by John Willinsky, then at the University of British Columbia, and now at Stanford University. The central objective of PKP since that time continues to be “improv[ing] both the scholarly quality and public access to the body of research knowledge in a sustainable and globally accessible form” (Owen and Stranack, 2012, p. 138). In support of this objective, PKP invests in ongoing work to improve the software it develops – in addition to OJS, this includes Open Conference Systems and Open Monograph Press – and contribute to enhancing the technology for scholarly communication.

Testament to its popularity, the number of OJS journals has grown steeply since the initial release of this software, especially over the past 10 years. In 2016, more than 10,000 journals published more than 400,000 articles using OJS, with the highest concentration of usage in Latin America and the Caribbean, followed by East and Central Asia and Europe[2].

Using only the simple metric of the volume of scholarly content being produced and managed in OJS journals, their preservation status is worthy of concern. However, further data, based on a survey conducted by PKP in 2009, suggests that OJS journals are particularly vulnerable to loss. The results of this survey showed that OJS journals are primarily born digital, published by independent or scholarly publishers and predominantly open access. It also showed that they are operating on a budget, without the resources available to commercial publishers (Edgar and Willinsky, 2010). Among the use cases for e-journal preservation identified by Neal Beagrie in a (2013) DPC Technology Watch Report, two are particularly applicable to these journals: if the publisher ceases operating or their operations or servers suffer catastrophic failure, this content would likely disappear (2013).

As Seadle explains:

[...][f]ew librarians serious expect a large publisher like Elsevier or Springer to vanish overnight, but the danger for a publisher with only one or even fewer than ten titles is historically much larger (2011, p. 194).

Prior to the launch of the PKP PN, the available distributed preservation options were not practical for most OJS journals. The major, well-established e-journal preservation services – LOCKSS, controlled LOCKSS (CLOCKSS) and Portico – are not a good fit because of cost, membership requirements or content restrictions. LOCKSS is an open source, library-led digital preservation system that was developed and is based at Stanford University. The primary implementation of this software is the Global LOCKSS Network (or GLN) which preserves mainly subscription-based journal content. Publishers participate in the GLN at no cost, which would seem to align this preservation option with OJS journals. However, at the time of writing, the GLN preserves content from around 200 OJS titles (out
of approximately 10,000). This is in large part because OJS journals are generally not a top preservation priority for LOCKSS Alliance libraries who vote on the content that is preserved in the GLN[3]. And because OJS journals tend to be published by many different publishers in relatively small numbers, there is also an issue of economies of scale in working with publishers.

Private LOCKSS Networks (PLNs) [also referred to as Community LOCKSS Networks (CLNs)] like the PKP PN are used to preserve content in specialized areas; a small number of PLNs other than PKP’s preserve OJS journals hosted by their member libraries. However, these PLNs have membership requirements as well as associated membership fees, restricting access to a minority of OJS titles (Skinner and Schultz, 2010). Other established preservation services including CLOCKSS and Portico also have associated membership fees for publishers which, while relatively small for journals with lower revenue, are out of reach for many OJS journals.

While the preservation problem is particularly acute for OJS journals, two separate studies have found that only around one quarter of all e-journal titles with ISSNs currently collected have been preserved by a recognized archiving program (Burnhill and Otty, 2015; Cornell University Library and Columbia University Library LOCKSS Assessment team, 2011). The Keepers Registry, which maintains metadata recording e-journal archiving arrangements, suggests that this means “about three-quarters of titles (with ISSNs) could be said to be at risk” (Burnhill, 2013, p. 6). Additionally, the Mellon-funded Strategies for Expanding Ejournal Preservation study at Cornell University Library and Columbia University Library found that there was not much overlap between the content preserved by the various initiatives like Portico and LOCKSS. At the time of their analysis, “Cornell held 45,602 e-journal titles with an ISSN, ejISSN, or both[...]. Of the 45,602, both LOCKSS and Portico preserve 3,476 or 7.6 per cent” (Cornell University Library and Columbia University Library LOCKSS Assessment team, 2011, p. 15).

This problem is even greater for Open Access journals, the category that the vast majority of OJS titles fall into. A 2013 survey by the Directory of Open Access Journals (DOAJ) underscores how significant a problem is the preservation of open access journals more generally. In keeping with survey data collected by PKP about OJS journals, the DOAJ data show that there are a few open access publishers indexed in DOAJ with more than 50 journals, but thousands publishing one or two journals, many of which lack the resources to go beyond publishing the content, and most that had not addressed the issue of journal preservation (Bjornshauge, 2014). Indeed, based on their analysis of Keepers Registry data, Burnhill and Otty conclude that:

> The journals from smaller publishers — those in the so-called ‘long tail’—are less likely to be ingested by organisations with archival intent, and are therefore at much greater risk of loss. Ensuring that this part of the scholarly record is kept safe is perhaps the most difficult task facing the library community” (Burnhill and Otty, 2015, p. 5).

These findings highlight the need for additional and complementary options for preservation of electronic journals, in particular for open access journals.

These gaps in the e-journal preservation landscape combined with a few other factors to prompt PKP to develop the PKP PN as part of its service offerings. First, PKP is starting to develop coordinated services for the OJS community, and preservation efforts are consistent with that focus (its other services include aggregating OJS journals into a central index, article-level metrics, and DOI integration). As well, several PKP Development partners are members of the LOCKSS Alliance, so adding a LOCKSS box for the PKP PN would be a familiar technology and would not come with an additional cost for these members. A PLN
for OJS journals is consistent with LOCKSS’ “community” focus whereby “like-minded institutions work together to share the preservation responsibility (including governance and sustainability) of e-content important to the group”[5]. And finally, PKP has found ways to automate the work of getting the journals into the PN, meaning that the service is not time-intensive to maintain once launched, and that staffing costs are kept to a minimum.

One of PKP’s goals in developing the PN was to maintain a low barrier to entry to enable as much participation as possible. The resulting network provides free preservation services for any OJS journal that meets a few minimum requirements. At the time of writing, these requirements are that the journal has published at least one article, has an ISSN and is running a current version of OJS. The PN allows OJS journal managers to register their journals for inclusion in the PN by enabling a plugin in their OJS instance and agreeing to the terms of a preservation agreement. On registration of a journal, the PN automatically determines if the registered journal meets the minimum inclusion requirements, prepares the raw journal content generated by OJS’s existing export tools, harvests new content from registered journals and adds the content to the PN, thus eliminating manual staff intervention at PN member sites and centrally at PKP.

The PKP PN, like most PLNs, is a dark archive, meaning that end users will not have access to the preserved content until after a “trigger event” (Reich and Rosenthal, 2009). In the PKP PN’s case, the trigger event can be notification by the journal manager or cessation of deposits into the PN after a period of inactivity. If a potential trigger event is detected, PKP staff will contact the journal to confirm its publication status, and if it is confirmed, approve the importing of preserved content into a special OJS instance – an “access node” – operated by the PN where it will be openly accessible.

As Kenney asserts, the shift from print to electronic journals has brought a number of causes for concern related to preservation, not least of which is that the “well-defined relationships and roles of the print era continue into the digital era and are inhibiting collaborative approaches to solving these problems today” (Kenney and Wesley, 2016). Kenney calls on libraries, publishers and preservation repositories to develop cooperative approaches to e-journal preservation. The PKP PN is representative of an approach that involves cooperation by all of these parties.

**Terms of service**

Journal managers who would like their journal to be included in the PKP PN must accept the terms of service in the OJS plugin. This agreement currently consists of eight clauses:

1. I have the authority to include this journal’s content in a secure PN and, if and when necessary, to make the content accessible in the PKP PN and its successors.
2. I agree to allow the PKP PN to include this journal’s title and ISSN, and the email address of the Primary Contact, with the preserved journal content.
3. I confirm that licensing information pertaining to articles in this journal is accurate at the time of publication.
4. I acknowledge these terms may be revised from time to time and will supersede all previous versions. I will be asked to review them and to agree to them to continue to include this journal’s content in the PKP PN.
5. I agree not to violate any laws and regulations that may be applicable to the content.
6. I agree to make every reasonable effort to inform the PKP PN in the event this journal ceases publication. I acknowledge that the PKP PN will also employ
automated techniques to detect a potential trigger event and contact the journal to confirm their publication status.

(7) I agree that the PKP PN reserves the right not to preserve or make content accessible.

(8) PKP reserves the right to use the aggregated content in the PKP PN for research and reporting purposes and will adhere to the norms of standard research procedures.

Development of the terms was an iterative process driven by several principles related both philosophically and technically to the planned PN, namely: ingesting journal content must be completely automated from the journal’s point of view, and as automated as possible from PKP’s point of view (i.e. until such a point as manual intervention is necessary in deciding whether to make content available after a trigger event); joining the PN must be as simple as possible for journals managers, and getting content out of the PN for distribution after a trigger event must be as simple as possible for PKP staff; making content available after a trigger event must be as future proof as possible; that is, policies and processes must save effort for the people in the future who will retrieve content and make it available; and finally, the PN should be inclusive.

A number of methods were used to test the first and subsequent draft service agreements. Besides inviting several OJS journals to participate in early testing, the authors held a workshop focused on the terms at the 2015 PKP conference; this workshop was attended by OJS journal managers from around the world. PKP also received advice from lawyers based in Canada and the USA and from the PKP PN Advisory Panel.[6]

Early testing quickly identified a number of issues with a planned Creative Commons (CC)-BY[7] license stipulation: some journals were using other variations of CC licenses, others had changed licenses and intellectual property arrangements as they moved from toll to open access, and yet others had concerns about accepting an explicit CC license without consulting all affected authors (a daunting task for those journals which had been publishing for many years). Based on this feedback, the CC-related clause was removed from the terms, as was language around post-trigger event scenarios that might have legal significance.

In a literature review and survey conducted between 2002 and 2004, Ayre and Muir, (2004) identified a number of complicated rights issues and solutions associated with digital preservation including copyright legislation related to preservation copying and the ability for publishers to grant permission to preserve materials that have multiple rights holders. In the intervening 15 years, new issues and approaches to rights management in the digital environment – such as CC licenses – have arisen. These issues, along with content licenses, copyright and intellectual property laws, will inevitably continue to evolve; highly specific and legally binding terms may prove out of sync with long-term preservation goals. Further, at least for the PKP PN, the risk related to making post-trigger event content accessible is very low, regardless of the original license under which the content may have been published. The community that is served by the PKP PN is, for the most part, making content available under an open access license or model and is interested in the accessibility and persistence of their content. In the unlikely event that someone takes issue with content that has been made accessible, the simple recourse will be to take it down immediately. On a related note, the terms do not include dispute resolution clauses: the PKP PN is a collaborative and mutual network arrangement and “dispute resolution and other harshly binary clauses can abrogate any meaningful notion of a trust foundation for operating in collaborative ways” (Dwayne Buttler, personal email to author, July 27, 2016).
As work on the service agreement evolved, additional principles solidified: the terms should be as simple and “plain language” as possible so they could be easily understood by a global audience (and lend themselves to translation where necessary); the onus should be on individual journals to confirm they are both willing and able to participate in the PN; the terms should be jurisdiction-neutral; and they should minimize risk to the PKP partners and sponsors who are active participants in the PKP PN. Though development of the terms proved one of the more challenging aspects of launching the PN, identifying key principles and testing the terms with OJS users ensured that the resulting agreement is consistent with the technical and philosophical vision for the PN.

The public knowledge project preservation network’s user interface and network architecture
From a journal manager’s perspective, the terms of service for the PKP PN are the main user interface to the network. Once the journal manager accepts the terms within simple form within plugin’s Web interface, all of the journal’s published issues are deposited into the network automatically, and each new issue is deposited on publication. The network is innovative in this respect; in most PLNs, adding new content is a manual process (although some PLNs have developed various ways of making that task as simple as possible).

Because the addition of new OJS issues to the PN is automated, and because new journals join the network on a daily basis, PKP had to develop a network architecture and associated tools that would isolate the preservation nodes on the network (which are conventional LOCKSS servers) from the OJS instances, and that would also check the content produced by the OJS instances for potential threats such as viruses and spoofed content intended to overload the network. To do this, PKP inserted a staging server between the OJS journals and the LOCKSS network. The PKP PN plugin for OJS communicates directory with the staging server, and the staging server then communicates with the LOCKSS network. The lifecycle of a deposit is:

- On publication of the issue, the journal sends an automated message to the staging server informing it of the new issue.
- The staging server retrieves the issue content from the OJS instance, and runs some validation checks (including a virus check) against the content.
- The staging server then packages up the content in the form that will be preserved in the LOCKSS network and informs the preservation nodes that new content is available.
- Finally, the preservation nodes harvest the content from the staging server and add it to the dark archive.

If PKP staff recognize a trigger event, they will extract the journal’s content from the dark archive and load it into the access node described earlier.

At the time of writing, the preservation nodes in the PKP PN are hosted at Indiana University, the Italian National Library of Florence, the National Documentation Centre in Greece, the Ontario Council of University Libraries (Scholars Portal), Simon Fraser University, University of Alberta, University of British Columbia, University of Pittsburgh and the University of Victoria[8]. Whereas journals that register with the PKP PN simply submit a form containing the terms of use as described above, institutions that maintain preservation nodes agree to a Memorandum of Understanding that defines the relationships between PKP and the host institutions and the obligations of each party. The Memorandum also outlines contractual or legal issues that host institutions (such as universities) require.
However, consistent with the terms of use that journal managers accept, the preservation nodes’ Memorandum is intended to be simple, clear and as free of jurisdictional specifics as possible.

**Discussion and future work**

**Current usage of the preservation network**

The PKP PN officially launched in June 2016 when notifications were sent to a number of relevant Listservs, and details about the availability of the PN were added to PKP’s website. Nearly 18 months later, as of November 2017, more than 600 OJS journals have agreed to the terms of service and the response from the OJS and preservation communities has been overwhelmingly positive. Journals from 37 countries have joined the PN.

PKP worked with the LOCKSS team at Stanford to plan and launch the PN, and with the DOAJ to ensure that the PKP PN meets its criteria to be one of the resources that is explicitly mentioned as a preservation service in the DOAJ application. This development with DOAJ was made possible by the PKP PN’s acceptance in the Keepers’ Registry, which enables PKP to publicly share information on its archived titles. The Keepers Registry “acts as a global monitor on the archiving arrangements for electronic journal,” providing information about which journals are preserved by which archiving services and highlighting those journals for which no arrangements exist[9]. DOAJ checks the Registry to confirm that a journal’s archiving arrangements are as stated in its application. Thus, the PKP PN should have positive implications for the OJS-based journals which apply for inclusion in DOAJ, as well as those that wish to qualify for the DOAJ Seal (journals that are awarded the Seal are those that adhere to outstanding best practice[10]). Since 2014, when DOAJ expanded its review criteria and process, verifying a journal’s digital archiving practices is one method that DOAJ uses to ensure the journals it indexes are of a high quality; recent analysis by Marchitelli et al. confirms this approach (Marchitelli et al., 2017).

Despite the positive response to the PKP PN’s launch, the approximately 600 journals currently preserving content in the PKP PN is a fraction of the OJS journals that are eligible to do so. Using a script, the authors attempted to determine how many journals from the list used to compile the OJS usage statistics in Figure 1 met the PN’s following minimum

![Figure 1.](image)
requirements: that they be running a version of OJS at least as recent as OJS 2.4.8 and that they have an ISSN. Of the approximately 10,000 journals, more than 2,500 met both requirements.

**Improving the user interface**

Considering that the PN is a free service, technical and communication factors both likely play a role in this delayed adoption. Because the service relies on a plugin, it is currently only available for a specific release of the OJS software. Work is underway so that the plugin will be available for the most current version of OJS, but users who have not yet upgraded to the 2.4.8 version of OJS will not have access to the service. Further, in OJS 2.4.8, the location of the PKP PN plugin is not obvious; given its longer association with OJS, a preexisting LOCKSS plugin is more prominent. This challenge is being addressed in part by making the PN more prominent in the interface for OJS 3.1. (Because OJS downloads are not tracked or registered, PKP has no method of knowing or contacting all OJS users, or all OJS hosts, to notify them of the availability of new services like the PKP PN.)

Users must opt into the PKP PN by means of enabling a plugin in OJS. In the OJS 2.4.8 interface, plugins are grouped together in the “journal management” area, and appear as an item on a list of “management pages” called “system plugins”. Clicking on “system plugins” brings the user to a “plugin management” page (Figure 2); the PKP PN plugin is part of a group of plugins called “generic plugins” (Figure 3), located partway down this list. Though the pathway to the PKP PN plugin is documented, it is not intuitive; users must know the plugin exists and purposefully navigate to that page to enable it.

In contrast to the hidden location of the option for enabling the PKP PN plugin, enabling general LOCKSS harvesting for inclusion in the GLN and conventional PLNs in OJS 2.4.8 is relatively obvious. The option is included in the “Policies” settings within the “Journal Setup” section, along with the peer review and privacy options:

In OJS 2.4.8, “Enabling LOCKSS to store and distribute journal content” (Figure 4) is much easier than enabling the PKP PN plugin. This ease is substantiated by the number of journals that have enabled the LOCKSS option. Running another script against the list of journals used to compile the OJS usage statistics in Figure 1, the authors attempted to determine how many journals enabled this option by requesting each journal’s LOCKSS permission statement. They found that approximately one third of the journals tested provided a LOCKSS permission statement. In contrast, the total number of OJS journals archived in the GLN is around 200[11]. The authors speculate that the relatively high

**PLUGIN MANAGEMENT**

This page allows the Journal Manager to review and potentially configure the plugins that are currently installed. Plugins are divided into categories, according to their function. The categories are listed below, and within each category, its current set of plugins.

- Metadata Plugins
- Authorization Plugins
- Block Plugins
- Citation Management Plugins
- Citation Database Connector Plugins
- Citation Output Plugins
- Citation Extraction Plugins
- Gateway Plugins
- Generic Plugins
- Implicit Authentication Plugins
- Import/Export Plugins
- OAI Metadata Format Plugins
- Payment Plugins
- Public Identifier Plugins
- Report Plugins
- Theme Plugins
- Install A New Plugin

*Figure 2.*

OJS 2.4.8 plugin management
visibility of this configuration option, plus the brand recognition of “LOCKSS” as a preservation platform, account for the roughly 3,000 journals that have enabled this option without belonging to a LOCKSS network[12].

The PKP PN plugin was not included in OJS 3.0 as the plugin required upgrades that were not in place for the release date. As seen in Figure 5 though, an “Archiving” tab was introduced to the interface in this major release, as part of a group of “website settings.”

In OJS 3.1, the PKP PN plugin will be integrated with the other preservation plugins available in OJS: CLOCKSS, LOCKSS and Portico[13] (Figure 6). These other preservation options are hidden within an “Alternate Archiving Options” block that is collapsed by default. This approach should provide a usability improvement for the PKP PN as it serves to contextualize the PKP service as a preferred archiving option and as an administrative/website setting and is much easier for journal managers to locate than in the previous version of OJS. Figure 6 depicts the Archiving tab after the user clicks on the link to expand it.
Figure 5.
OJS 3.0 archiving tab
Figure 6.
OJS 3.1 archiving tab
Refining the network’s infrastructure

Around a year and a half after the launch of the PKP PN, its technical and infrastructural components have proven to be stable and reliable. In a small number of cases, the PN encounters issues with registering journals and with harvesting deposits. The two most common causes of these issues are nonstandard configuration of the Web server that the journal is hosted on and variations in the operating environment of the journal. PKP is improving the PKP PN plugin’s ability to detect these variations in a journal’s server environment and to report them to the journal manager.

Several other areas of development are on PKP’s roadmap for the PN:

- Harvesting and preservation of supplemental files: OJS allows authors to publish raw data files, videos, and other “supplemental files” with articles. Currently, the PKP PN only preserves the article text (PDF and/or HTML and embedded images); it does not preserve supplemental files. Work is underway to address this functionality.
- Automatic resolution of article DOIs (Digital Object Identifiers) to the access node: Many articles published in OJS have DOIs[14]. It should be possible to update the DOIs for articles that have them to point to triggered content.
- Feedback to the journal manager or editor about actions they can take to facilitate the preservation of their content: Currently, the only feedback that journal managers get about their journal’s content is the status of each deposit’s progress through the deposit/harvest/storage workflow. As the staging server runs validity checks on the harvested content, it should be possible for it to send information back to the journal manager or editor that may help them change local policies and practices to enhance the preservability of their content. One example of this type of information is the choice of preservation-friendly file formats for supplemental files.
- Automated tools for extracting content from the PN and ingesting it into the access node: Ideally, the process of making a triggered journal accessible again should be as automated as possible. Tools for doing this will leverage OJS’s ability to import journal content.

Understanding our users better

A significant obstacle to developing these new features is that PKP does not have a formal business relationship with the journals preserved in the network or with journals that wish to be preserved but are not for some reason (e.g. the journal manager does not have the authority to accept the terms of service). This lack of relationship results in incomplete knowledge about the journals’ requirements and awareness of preservation issues on the part of PKP. In other words, as PKP cannot communicate with the managers and editors of the journals it is serving (or could potentially serve), its ability to improve this service is hindered.

The large number of journals that may qualify for inclusion in the PN, and their broad diversity in terms of staffing resources, make formal relationships and effective communication between potential and current clients of the PN and PKP (which also has limited staff to devote to this level of communication) challenging. This lack of relationships is not necessarily negative. PKP has never required that anyone register or self-identify before downloading its software, and central to PKP’s philosophy is that anyone, anywhere, can download OJS and use it to publish a journal:
Researchers from all over the world can download and use our software; in a very real sense, journals from Indonesia and Sri Lanka can operate on the same field (or quality of platform) as journals from the United States and Germany” (MacGregor et al., 2014, p. 172).

John Willinsky, founder of PKP, elaborates:

[T]he original thinking behind the distributed model was to develop local capacities able to support this and other OSS [open source software] systems and tools with some autonomy and ownership, against a legacy of knowledge centers and peripheries, while we took on the additional responsibilities of providing a self-instructive system by which others could find their way (Willinsky, personal message to author, Oct 29, 2017).

Recently, PKP has added an opt-out feature to OJS that does send a minimal set of publicly available data back to PKP (Alperin et al., 2016, p. 60), which will help PKP improve its software and services while allowing OJS journals to remain autonomous and anonymous.

In addition to improvements to the OJS interface to make the PKP PN plugin easier to find and to provide contextual information about the different preservation options within OJS, defining a communication strategy will be key to increasing the number of OJS journals that are preserved in the PN. Although promotion of the service has not been a major focus thus far, once the PKP PN plugin is available in OJS 3.1, increased promotional efforts to OJS journals and their institutional hosts will be timely. Determining a communication strategy is complicated, however, both because there is no easy method of contacting journals or hosts, and because of a likely lack of awareness of preservation issues on the part of the journal publisher (for example, a question about the difference between back-ups and preservation has already arisen in relation to the PKP PN). In requiring that the journal manager accept the terms of service, the PKP PN shifts the onus for preserving journal content from libraries and librarians to journal publishers. Whereas libraries can be expected to be somewhat familiar with preservation issues, small journal publishers will not be as aware.

As part of their Strategies for Expanding EJournal Preservation Mellon-funded project, Cornell contacted over 350 small publishers, most of whom – much like OJS publishers – publish only one or two titles, in regards to their participation in a potential preservation project. Their key takeaway from this exercise was “not that publishers are resistant to preservation because of costs, or digital rights management concerns, but that there is a lack of understanding of digital preservation and its objectives” (Regan, 95). One of the central findings reached by the project holds great promise for the PKP PN and indeed, for other preservation efforts:

If we invest in efforts to educate publishers it will make great strides in expanding the preservation of e-journals […] [o]utreach and education have the potential to go a long way in securing scholarship (Regan, p. 95).

Conclusion

PKP considers 600 journals preserved in the PKP PN within the first year and a half of operation to be a success. However, given that there are over 10,000 OJS journals (2,500 of which could be using it now, provided they can accept the terms of service) that could be taking advantage of the service begs the question of why that number is not higher. PKP purposefully collects little information about OJS users, making PN feature development and communication with potential users challenging. Nonetheless, considering that many more journals enable the LOCKSS plugin within OJS than are preserved in the GLN or a PLN, and that evidence from a variety of studies demonstrates that small publishers are
interested in preservation, the authors anticipate that upcoming usability improvements in the OJS interface and a focus on promoting the benefits of inclusion in the PKP PN will greatly increase the number of journals that opt into the service.

From a more general perspective, the PKP PN demonstrates that it is possible to implement distributed digital preservation services that both preserve content from a large number of partners and that require little direct interaction with those partners. Determining an effective amount of interaction with a service’s users is probably best determined by understanding their needs and the amount of risk faced by their content but should also take into account the ability of all partners to manage that interaction.

Notes
1. https://pkp.sfu.ca/ojs/
2. Using the criteria of “at least 10 articles published that year” https://pkp.sfu.ca/ojs/ojs-usage/ojs-stats/
3. www.lockss.org/support/prepare-your-content/open-access-preservation-options/
4. Seadle, on the other hand, in a separate analysis of LOCKSS, CLOCKSS and Portico data, found “significant overlap among the archiving systems” though he notes that “[p]art of the overlap comes from a few large publishers, who use both systems” (2011).
5. www.lockss.org/community/networks/
6. At the time of writing, the PKP PN Advisory Panel was comprised of: Lars Bjornshauge, Directory of Open Access Journals (DOAJ), Sioux Cumming, International Network for the Availability of Scientific Publications (INASP), Alan Darnell, Ontario Council of University Libraries (OCUL), Geoff Harder, University of Alberta Libraries, Alex Mendonça, SciELO, Susan Murray, African Journals Online (AJOL), Victoria Reich, LOCKSS, Heather Joseph, SPARC, Dwayne K. Buttler, J.D., University of Louisville.
7. https://creativecommons.org/licenses/by/4.0/
8. A preservation node is a server running the LOCKSS software.

Table I. Geographic distribution of journals in the PKP PN[15]

<table>
<thead>
<tr>
<th>Country</th>
<th>No. of journals</th>
</tr>
</thead>
<tbody>
<tr>
<td>The USA</td>
<td>152</td>
</tr>
<tr>
<td>Canada</td>
<td>91</td>
</tr>
<tr>
<td>Brazil</td>
<td>84</td>
</tr>
<tr>
<td>Italy</td>
<td>74</td>
</tr>
<tr>
<td>Indonesia</td>
<td>37</td>
</tr>
<tr>
<td>Costa Rica</td>
<td>34</td>
</tr>
<tr>
<td>Ireland</td>
<td>23</td>
</tr>
<tr>
<td>Argentina</td>
<td>17</td>
</tr>
<tr>
<td>Germany</td>
<td>15</td>
</tr>
<tr>
<td>Russia</td>
<td>12</td>
</tr>
<tr>
<td>The UK</td>
<td>9</td>
</tr>
<tr>
<td>Singapore</td>
<td>8</td>
</tr>
<tr>
<td>Norway</td>
<td>8</td>
</tr>
<tr>
<td>France</td>
<td>7</td>
</tr>
<tr>
<td>Turkey</td>
<td>6</td>
</tr>
<tr>
<td>Malaysia</td>
<td>5</td>
</tr>
<tr>
<td>Cuba</td>
<td>5</td>
</tr>
</tbody>
</table>

Open journal systems content

259
Based on the Global LOCKSS Network’s titles database.

PLNs other than PKP’s preserve OJS titles; for example, the Council of Prairie and Pacific University Libraries (COPPUL) PLN preserves approximately 50 titles. However, these do not account for the number of OJS journals that have enabled LOCKSS.

The text relating to the Portico plugin was not finalized for the Archiving tab at time of screenshot.

Only countries that have five or more journals in the PKP PN are included in this table; not shown are 20 countries that have fewer than five journals in the PN. Table 1 is based on data in the list of journal issues at http://static.pkp.sfu.ca/pkppn/onix.csv at the time of writing.

References


Further reading

Corresponding author
Bronwen Sprout can be contacted at: bronwen.sprout@ubc.ca

For instructions on how to order reprints of this article, please visit our website: www.emeraldgrouppublishing.com/licensing/reprints.htm
Or contact us for further details: permissions@emeraldinsight.com
Linking historical collections in an event-based ontology

Qing Zou and Eun G. Park
McGill University, Montreal, Canada

Abstract

Purpose – This study aims to explore a way of representing historical collections by examining the features of an event in historical documents and building an event-based ontology model.

Design/methodology/approach – To align with a domain-specific and upper ontology, the Basic Formal Ontology (BFO) model is adopted. Based on BFO, an event-based ontology for historical description (EOHD) is designed. To define events, event-related vocabularies are taken from the Library of Congress’ event types (2012). The three types of history and six kinds of changes are defined.

Findings – The EOHD model demonstrates how to apply the event ontology to biographical sketches of a creator history to link event types.

Research limitations/implications – The EOHD model has great potential to be further expanded to specific events and entities through different types of history in a full set of historical documents.

Originality/value – The EOHD provides a framework for modeling and semantically reforming the relationships of historical documents, which can make historical collections more explicitly connected in Web environments.

Keywords Event, Basic Formal Ontology, BFO, Creator history, Event Ontology for Historical Description (EOHD), Historical collections

Paper type Technical paper

Introduction

Ontology technologies have greatly transformed separate information resources in such a way that they are now connected throughout documents within Web environments (Bizer et al., 2009; Berners-Lee et al., 2001). This is one step closer to building a semantically linkable document in digital collections. These technologies can have an impact on organizing, modeling, presenting and disseminating various types of information in historical collections. However, for humanities scholars and users who often use historical collections, to some extent, the Web still maintains existing problems related to unstructured, semi-structured and rarely structured ways of maintaining historical collections. Presenting historical documents is also complex and heterogeneous, usually involving a large amount of text.

Features of historical collections

Historical collections refer to sets of documented information that are unique primary sources with historical, cultural and informational value. Historical collections could be called digital collections, special collections in libraries, archival collections in archives, cultural heritage collections in museums or digital humanities, depending on their context. They are likely very context-specific as they have accumulated over many years. Documents in a historical collection are often locally organized in a non-standardized way in a regional repository; thus, historical documents are frequently created in silos. The metadata of historical collections is commonly limited to the collection level because they are
indexed by traditional text-based approaches without a rigid structure of information (Yakel et al., 2007). In addition, as information on the descriptions of historical collections is archival or bibliographic in nature, it is embedded implicitly in text form. Contextual information (as things) and their meanings in the documents are also embedded implicitly (Millar, 2002). Thus, the organization and description of historical collections seem to be disconnected due to limited technical, institutional and practical reasons, especially in the field of humanities. As a result, it is difficult for scholars and users to clearly organize and connect information on historical collections on the Web.

For research purposes, scholars tend to examine particular people or objects (or artefacts) at a particular moment, as well as over time at specific places, and then see those interactions in documents. To track such interactions, documents need to be specifically divided by things or topics represented by their relationships, and then a link to those things needs to be built internally in a document. That can then be connected externally to relevant things or topics in other documents within a collection.

Adoption of ontology technologies in historical collections

To make clear connections between documents, several projects have made an effort to apply ontology technologies to historical collections. In the field of archives, the Encoded Archival Context for Corporate bodies, Persons and Families (EAC-CPF) Project attempted to develop a link between authority data on individuals, families and corporate bodies in archival collections (Mazzini and Ricci, 2011; Society of American Archivists, 2010). In addition, the Social Networks and Archival Context Project (2013) is building further links, such as a social network of authority data, which has been identified from the EAC-CPF. In the UK, the Linked Open Copac Archives Hub (LOCAH) Project was initiated to create links between bibliographic and archival data of monographs and serials that have various external sources (Ruddock, 2011; Linked Open Copac Archives Hub Project, 2014). The project has transformed records into the Resource Description Framework (RDF) format and made them available on the Web as an open linked data hub.

Some projects use a linked data approach. For example, to describe contexts in digital collections, Lee’s (2011) context model provides a structure to describe context with nine classes to describe digital collections. The Europeana Data Model (EDM) developed a general model with basic entities to describe different types of digital objects from libraries, archives, and museums (EDM, 2013). International Council on Archives Experts Group on Archival Description (2016) proposed the Records in Context as a new standard with 13 classes and more than 700 relations for describing archival descriptions. In addition, the Linked Jazz project reformed the network of relationships made between jazz musicians by using linked open data technology (Thorsen and Pattuelli, 2016). In particular, this project analyzed transcripts of interviews containing oral histories. “Out of the Trenches,” by the Pen-Canadian Documentary Heritage Network, tested the application of linked open data to different forms of digital resources related to the First World War that had been collected from several institutions across Canada.

These example projects are good attempts to show the links between data within digital collections by building ontologies. In this way, relationships between selected pointers (called entities or events) and their semantic meanings in the documents can be reformed, represented and better re-discovered in digital collections. To provide clearer relationships between events regarding a person, thing, time and place and their interactions embedded in historical collections, it is necessary to examine events as entities that can define terms and hierarchical structures of documents.
To manage event information in documents, the concept of an event has long been studied along with history in the fields of history, philosophy, linguistics, artificial intelligence, library and information science and so forth. “Event” refers to something that occurs over a period of time at certain places and acts upon or with entities (Pustejovsky et al., 2003, p. 3). To historians, understanding and interpreting history is possible through historical documents, which are containers of historical information in historical collections. In philosophy, history is considered to be interconnected events (Galton, 2008), and historians build systems for describing collections as a history-as-practice. In philosophy, events are considered to basically be property instantiation (Kim, 1993) and then history consists of smaller and more inter-connected event instances. As events are related to time and place, events are considered a property of a spatio-temporal region (Lewis, 1986). With regards to time, there are two views of events: a three-dimensional view and four-dimensional view (Galton and Mizoguchi, 2009). The three-dimensional view is a normal view with space and absolute time. The four-dimensional view sees time as a three-dimensional view and the fourth dimension of time (Galton and Mizoguchi, 2009). In linguistics and artificial intelligence, where events are accepted as occurrences over intervals, an event can be associated with an interval when a predicate occurs (Bennett and Galton, 2004). Although an event may be defined and understood differently in various disciplines, it is agreed that an event is linked to certain agent(s), time, spatial region(s) and spatio-temporal region(s).

In the area of library and information science, a knowledge representation approach is the representation of event-related information; therefore, modeling the information in ontologies has been introduced. There are several traditional knowledge organization systems in the field, including a terms list, controlled vocabularies, subject headings, classification schemes, taxonomies, thesauri, ontologies and more. Ontologies refer to an “explicit specification of a conceptualization” (Gruber, 1993, p. 199), and there are two types of them, namely, lightweight-ontologies and heavyweight-ontologies. While lightweight-ontologies indicate taxonomies and thesauri, heavyweight-ontologies are description logic compliant ontologies (Uschold and Gruninger, 2004). Ontologies can also be grouped by upper ontologies, domain-specific ontologies and task/application ontologies (Guarino, 1998). In this study, when the term “ontology” is used, it indicates a formal ontology in a general sense and, more specifically, heavyweight ontologies.

**Ontology building approach**

To identify a feasible ontology for managing common events in documents, three relevant and popular ontology models have been considered: the International Council of Museum’s Committee on Documentation (CIDOC) Conceptual Reference Model (CRM) (CIDOC CRM Special Interest Group Working Group, 2011); ABC Ontology (Lagoze and Hunter, 2001); and Basic Formal Ontology (BFO) (Grenon et al., 2004). These ontologies are explained by how to define basic components (e.g. time, place, object, agent and event).

The first CIDOC CRM is domain-specific ontologies. The CIDOC CRM is designed to describe museum artefacts in cultural heritage institutions (CIDOC CRM Special Interest Group Working Group, n.d.). CIDOC CRM intends to make a semantic glue of artefacts and transform localized information to accessible resources on the Web (Doerr et al., 1999). The CIDOC CRM contains 140 classes and 144 properties in its latest version and some parts of it include event-related and class-based components. A subset of its classes and properties can be used for events. However, this ontology tends to be a rather complicated way to represent events in historical documents.
The ABC Ontology is a general event model. It was created for the Harmony Project, which aims to provide a conceptual basis for analyzing metadata vocabularies and instances developing descriptive vocabularies and mapping vocabularies (Lagoze and Hunter, 2001). The ABC Ontology can model the creation, evolution and transition of objects over time (Lagoze and Hunter, 2001). The ABC Ontology is a class-based approach with three primitive categories under the Class Entity (abc:Entity): abc:Abstraction, abc:Actuality and abc:Temporality (Lagoze and Hunter, 2001). ABC Ontology is more event-centric and simpler than CIDOC CRM. However, it is difficult to align the ABC Ontology with upper ontologies.

The BFO provides a genuine upper-level ontology with event models, which can be applied to specific domains (Grenon and Smith, 2004). The BFO can use both three-dimension and four-dimension entities (Grenon et al., 2004). It introduces the SPAN and SNAP concepts. SNAP is a specific type of ontology of events as continuants in time and SPAN is one of the occurrences. These two concepts nicely reconcile the three-dimensional with the four-dimensional views (Grenon and Smith, 2004). While continuants indicate entities that have continuous existence and persist in time, occurrences refer to processes, events, activities and changes (Grenon and Smith, 2004). In other words, continuants can be constant changes and occurrences depend on continuant objects. Thus, occurrences are four-dimensional and are all bound in time. BFO has been widely accepted and applied in the fields of biology and other communities. BFO is a more mature upper ontology. This aspect is advantageous to apply in the ontology of this study.

Making use of existing ontologies, when possible, is the general principle of ontology development. As described above, these event-related ontologies differ in several ways. To address the need of historical collections, a new ontology should capture the features of historical collections. We intend to adopt a domain-specific and event-centric ontology built on an upper ontology. In comparison, among the three ontologies, the ABC Ontology is relatively general and more event-centric and simpler than CIDOC and BFO. These two ontologies have more specific classes, properties and relations and an event model is a part of the entire ontology. BFO is a more mature upper ontology, which enables it to be represented in an elaborate way. To gain benefits from the chosen ontologies and have sufficient flexibility to fit them into historical collections, a BFO based model is adopted for building our domain specific and upper ontology.

Developing an ontology is a process that requires multiple phases, including:

- requirements and analysis;
- design and implementation;
- testing and validation; and
- maintenance (Jones et al., 1998).

Consequently, building an ontology involves many activities. Activities for building an ontology in general can be categorized into three types:

1. “ontology management activities” (e.g. scheduling, control, and quality assurance);
2. “ontology development-oriented activities” including pre-development (e.g. environment study, feasibility study), development (e.g. specification, conceptualization, formalization and implementation) and post-development activities (e.g. maintenance and use); and
3. “ontology support activities” (e.g. knowledge acquisition, evaluation, integration, documentation, merging, alignment and configuration) (Corcho et al., 2006, p. 10).
Among the three types of activities, the second type involves the actual development of an ontology. The third category, including knowledge acquisition, evaluation and documentation, should be performed during the entire life cycle of the development of the ontology. Each activity in the entire process should be carried out very carefully (Fernández-López et al., 1997), and the specific activities in the entire process may vary, depending on each type of research work.

Regarding the methods of building an ontology, to recognize the fundamental principles of ontology development, several approaches for building ontologies have been introduced and used. Popular examples include Methontology (Fernández-López et al., 1997), Model Driven Architecture (MDA) (Gašević et al., 2006) and the NeOn Methodology (Suárez-Figueroa, 2010). The MDA is a software design approach for developing an object model as a ground by use of software systems (e.g. Unified Modeling Language) and then systematically transforming it into a formal ontology (Keshk and Chambless, 2008). Among these approaches to building an ontology, methontology and NeOn are two systematic and comprehensive methods. Methontology has proven to be a comprehensive method for modeling a development process and building ontologies from the beginning (Fernández-López et al., 1997), and it has been implemented in several studies, for example, e-Government ontology (Dombeu and Huisman, 2011), information science ontology (Sawsaa and Lu, 2012), legal ontologies (Corcho et al., 2005), online learning (Guinebert et al., 2017) and others. The NeOn Methodology is an approach that supports making use of existing resources (e.g. glossaries, ontologies) (Suárez-Figueroa et al., 2012). Therefore, this study adopts the use of Methontology and takes into consideration the NeOn Methodology.

To solve the current challenges of historical collections, this study aims to adopt ontology technologies in historical collections and build an ontology to reconnect and represent the content and relationships in historical collections. This study is organized as follows. The second section in this study explains the methods and procedures for building the ontology. The third section in this study describes the results of the ontology that has been built and discusses relevant issues. The fourth section provides a conclusion, including plans for further studies.

Methods and procedures
This study aims to explore a way to organize historical collections by examining the features of historical documents, identifying events from the documents and building an event-based ontology model. Research questions are formulated as:

RQ1. What are the components of events in historical documents?

RQ2. Which procedures are used when building an event-based ontology model?

Sample data sets of historical collections are taken from the implementor listing site of the Society of American Archivists (SAA). The site contains multiple Encoded Archival Descriptions (EAD) records of historical collections in over 90 university archives in the USA and Canada. As one of the most authoritative institutions in this field, the SAA site is a valid source for locating archival institutions.

Historical collections are collections of information objects built by their creators (e.g. people, families and organizations) over a period of time. They contain recorded information, which is the result of their creators doing certain types of activities (or events). Among the ontology models explained in the previous section, the BFO model is selected in this study. Two views of SPAN and SNAP are introduced. In Figure 1 from top to bottom, an
individual, an organization and a family can be considered as a type of continuant (SNAP) and they show reflections (or results) of snapshots of a person, a family and an organization at a certain time in their existence. The person’s life or the history of a family or an organization is a type of occurrent (SPAN).

In Figure 1, File 1, File 2 and File n are information artefacts and blocks with different lengths denoting the duration that files were created and used. File creation and usage are two important events (i.e. activities) here.

As the representation information in historical collections is displayed as an aggregation of historical documents, they can be organized into object-oriented information about materials, namely, physical and topical information about the documents and types of histories (SPANs), such as the history of accumulated documents, history of creators, custodian history, administrative history, etc. Figure 2 shows the structures of descriptions in historical collections.

In Figure 2, historical collections are an aggregation. A resource map consists of three types of histories: administrative history, custodial history and creator history. Creator history includes individual, family and organization. To describe historical collections, information about documents can be organized using traditional approaches developed for bibliographical control. A history of a creator can be one of an individual, a family or an organization. The histories of accumulated documents and creators are considered to be the core creator information of the documents. The custodian history and administrative history are used to trace and reveal what has been done to the files, as they have been created and actively used. Descriptions of different types of history are inter-connected events. To present the structures of descriptions in historical collections, this study takes the Open Archives Initiative Object Reuse and Exchange (2008) as a working framework. To reflect the lives of individuals and the histories of families and organizations, the model in this study focuses on representing the nature of different types of histories and their linkages among the many types of histories.

As explained in the Introduction section, based on BFO, an Event Ontology for Historical Description (EOHD) is designed and developed by using Protégé, which is an open source ontology editor (Stanford Center for Biomedical Informatics Research, 2016). This model is intended to be a domain ontology for a particular purpose. The EOHD Ontology is designed in two layers: EOHD Core and EOHD Extended. The EOHD Core provides alignment with BFO. The EOHD Extended is built on the EOHD Core and provides flexibility for specific requirements and needs from host repositories. The EOHD should be flexible enough to have changes in events, and changes can be recorded in certain ways. In other words, the
changes are always associated with certain information if it is recorded. This model is unique and beneficial because it identifies certain types of events related to file creation and usage, which allow the EOHD model to connect the events to documents. In addition, the EOHD model is different from other existing event ontologies, such as the event ontology (Raimond and Abdallah, 2007), linking open descriptions of events (Shaw, 2010), and a simple event model ontology (van Hage et al., 2011). The EOHD model is useful by presenting not only spatio-temporal aspects of events but also the types of changes in events.

In the EOHD Core, the super class for processes, events and states is eohd:Occurrent. It is equivalent to bfo:Occurrent. Each eohd:Occurrent can include agents (eohd:hasAgent) and associate information objects as direct outcomes of eohd:Occurrent. It may also relate to the representation of information objects describing the eohd:Occurrent and associated information objects.

In BFO, events are considered a type of SPAN entity. The class bfo:event is a sub-class of bfo:process_boundary, which is a sub-class of bfo:processual_entity. Meanwhile, the class bfo:process is also a sub-class of bfo:processual_entity and disjoint with class bfo:fiat_process_part, bfo:process_aggregate, bfo:processual_context, and bfo:process_boundary – a parent class of bfo:event. However, bfo:event is a specific class that is defined for “the fiat and bona fide instantaneous temporal boundaries of processes” (Grenon and Smith, 2004, p. 154). To cover broader aspects of procedural entities, eohd:Event is defined as a sub-class of bfo:processual_entity. This approach is common with using inherited structure and the object-oriented concepts like other ontology developing methods in Unified Modeling Language (e.g. concepts, conceptual modelling, classifying ontologies, using notations for other symbolic representations, etc.). A typical eohd:Event occurs in a temporal region (eohd:Time) and at a spatio-temporal region (eohd:Place) as well.
Results and discussion

Defining events

When designing an ontology, the first step is to define events. To make use of existing vocabularies whenever possible, the Library of Congress (LC)’s event types (2012) are adopted because all of its historical events are well defined in the vocabulary of preservation events and preservation is a part of its operational activities. The preservation events include: capture, compression, creation, deaccession, decompression, decryption, deletion, digital signature validation, fixity check, ingestion, message digest calculation, migration, normalization, replication, validation and virus check (Library of Congress, 2012). Librarians or computer applications as agents are involved in events and deal with historical documents as objects. Administrative history and custodial history are the histories of those types in a series of events.

Defining event types

Defining event types is the next step. Human beings often pay attention to changes related to events. In Kaneiwa et al.’s study (2007), the authors formally defined the six types of changes, including state change, temporal existence change, spatial existence change, cardinality change, comparison and object identification change. The types of changes in the EOHD are adopted as the following:

- A state change event yields the change of states from one time to the next time.
- A temporal existence change event refers to an occurrence that changes the existence of an object according to a change in time (e.g. birth events).
- A spatial existence change event refers to occurrence changes to the existence of an object depending on movement through space (e.g. travel events).
- A cardinality change event changes the cardinality of objects.
- A comparison event refers to the attribute value of an object that is found to change when comparing that value with the attribute value the next time (e.g. temperature raise events).
- An object identification change event refers to when the essential property of an object is changed, and thus, the object cannot be recognized as the former object the next time (e.g. water to steam events) (Kaneiwa et al., 2007).

Figure 3 illustrates that under event, three classes, namely, eohd:Event, eohd:BioEvent, eohd:OrgEvent and eohd:ArchivalEvent, are defined with different kinds of changes.

In Figure 3, it is observed that an event can be a sub-class of one of the change events and also a sub-class of one of eohd:BioEvent, eohd:OrgEvent, or eohd:ArchivalEvent. There is no disjoint relation between the three sub-classes and the types of change events. For example, eohd:Birthevent is not only an eohd:TemporalChange event but also an eohd:BioEvent. Their relations are shown in Figure 4.

Similarly, the eohd:Death event can be defined. To demonstrate how to apply the EOHD Ontology to existing historical documents, Figure 5 displays one record, on Dale C. Thomson’s life as an individual, that is taken from the sample datasets. The biographical sketch of the selected document is in a typical free-text format. For the display purposes of the record, because of page limits, a partial view of five events is illustrated in Figure 5.

In Figure 5, the upper half consists of inter-connecting biographical events and the lower half is the “Research Activities” series containing many files (only four are listed here). In the upper part, the green oval shapes indicate events. The green boxes are temporal, spatial and
other important properties related to the events. Figure 5 includes events: “Dale C. Thomson’s Birth” event (i.e. eohd:Birth), “Dale C. Thomson’s Death” event (i.e. eohd:Death), “studied at University of Alberta” (i.e. eohd:StateChangeEvents) and others. The relations among events can connect events together. In Figure 5, eohd:hasNextEventRelation is applied to link events. Two events, “Studies at Université de Paris” and “Retried from McGill,” are connected, either through eohd:hasNextEventRelation or a more specific event relation. This way, a biographical sketch can be transformed from a free-text format into a structured file format. In the lower part, the green boxes are files linked with title and file number properties through creation events.

For the purposes of this study, an EOHD is built to overcome the existing challenges of historical collections and represent them more explicitly in web environments. Our approach is to design a general conceptual model for capturing the three types of history (e.g. administrative history, custodial history and creator history) and to build a domain-specific ontology to better describe events in historical documents. The proposed ontology has the following distinct features:
It is built on the BFO in a mature upper ontology.

It makes use of existing event-related vocabularies that are defined by LC.

It is built on types of history with six kinds of changes.

In this study, we demonstrate how to apply the event ontology to biographical sketches of a creator history to connect events and link them with event types. The EOHD can provide a framework for clearly modeling, representing and better understanding historical documents. It has great potential to be further expanded into a full set of historical documents in several ways. It enables unambiguous links to be made between events and entities through different types of histories and can connect other events in existing documents, such as the names of people, places, dates and so forth. It can be expanded to fully test other types of histories involving families and organizations. It can also be expanded to link to other events in more specific entities, such as authorized name-entities and subject-entities.

In addition, it is feasible to demonstrate the links that are together with, and adjacent to, the linked data approach. For example, by expressing events and their relations in RDF triples, the SPARQL Query Language and RDF Query Language (SPARQL) can be used to retrieve relevant information through queries. Furthermore, this ontology can be used for the more advanced capacity of searching for and browsing through historical materials and assisting in the exploration of these materials. Because the EOHD offers a better way of organizing and representing historical collections on the web, this approach enables
information on a creator and a history to be represented in a way that is more structured, linked, interconnected and user-friendly. This approach will move from the traditional representation practices in libraries, archives and museums (Zhang, 2012) to taking one step towards a more dynamic, intuitive and event-centered approach for knowledge representation.

**Conclusion**

Although a large amount of historical collections exist in text form on the Web, these collections are not sufficient to fulfill the needs of scholars and users. They bring up problems related to indexing, searching and retrieving. To move towards inter-connecting materials on the Web and make them more searchable, semantically linkable and usable for user groups and the public, this study has designed and tested an event-based ontology for historical collections. Our approach has taken into consideration the nature of historical documents and the relationships of events within the collections. This event ontology enables humanities researchers and users to make explicit links between events and entities to describe historical collections in a clearer and more explicit way.

In further studies, as the ontology developed was designed with events that are defined by LC, it may be too general in format. In a large-scale setting in which more diverse kinds of events are used, events could be divided into more specific types that are diverse enough to apply a greater number of sub-events to an aggregation. When using the six types of changes with events in the EOHD, a greater number of changes could be adopted in future works, such as to express the temporal and spatial relationships of events, as well as to present their structural and causal relationships (Westermann and Jain, 2007). For further study, an automatic or semi-automatic procedure needs to be developed to keep up with the accumulation of historical documents. In addition, when implementing ontologies in a large set of historical collections, it is necessary to use natural language processing and information extraction techniques with automatic and semi-automatic operations.

**References**


Further reading


About the authors
Qing Zou is a PhD Candidate at the School of Information Studies at McGill University. He is also System Librarian at Lakehead University Library, Thunder Bay, ON. His research interests are metadata, ontology, knowledge organization systems, archival description, and knowledge representation.

Eun G. Park is an Associate Professor in the School of Information Studies at McGill University, Montreal, Quebec, Canada. She holds a PhD from the University of California at Los Angeles. Her research interests include digital curation, ontology, metadata, trusted digital repositories, enterprise content management, and electronic health records management. Eun G. Park is the corresponding author and can be contacted at: eun.park@mcgill.ca
Master’s theses and open scholarship: a case study
Joachim Schopfel, Sylvain Vanacker, Eric Kergosien and Bernard Jacquemin
University of Lille, Villeneuve d’Ascq Cedex, France

Abstract
Purpose – This paper aims to show how Master’s theses can contribute to open scholarship and give reasons why this should be done.

Design/methodology/approach – The paper provides an overview of published studies and, based on the experience at the University of Lille (France), describes some essential aspects for the processing and valorization of these documents in the academic cloud, as a contribution of open scholarship.

Findings – Because of their number and diversity, collections of Masters’ theses in open repositories could be an excellent showcase for the universities’ Master programs and research. They could also offer interesting and large samples for content analysis, citation analysis and text and data mining (TDM). However, some issues need attention, above all intellectual property, quality and preservation. Quality is crucial, and the paper describes how the Lille project proceeds to assure sufficient quality and right clearance, and why the project shifted from students’ self-archiving to a digital library collection in the academic cloud, run by faculty and information professionals. The paper presents also some usage statistics to illustrate the potential, global impact of such a collection.

Practical implications – The paper provides helpful and empirical evidence and insight for those who want to develop the dissemination of Master’s theses via open repositories.

Originality/value – In the context of open scholarship, only few studies deal with Master’s theses, and this paper is the only recent reference that brings together a review of other papers and a case study with empirical evidence.

Keywords Digital libraries, Open repositories, Open access, Electronic theses and dissertations, Master’s theses, Open scholarship, Cloud

Paper type Case study

Introduction
Open scholarship has become the new paradigm of scientific communication. The term refers to “teaching and research practices that espouse openness (i.e.) a collection of emergent scholarly practices that espouse openness and sharing”, including open access, open publishing and open education (Veletsianos and Kimmons, 2012, p. 167). An ever-increasing number of books, journals and repositories provide free and unrestricted access to scientific content from all disciplines. Two third of the open institutional repositories contain more or less important collections of electronic theses and dissertations (ETDs), mostly PhD theses, considered as the most useful kinds of invisible scholarship and the most invisible kinds of useful scholarship (Suber, 2012).

The situation is quite different for Master’s theses. Even if their number is several times higher[1], they are much less well represented in open repositories than PhD theses. Less than 1 per cent of all repositories clearly indicate that they contain Master’s theses, and freely available Master’s theses represent less than 0.5 per cent of the scientific items retrieved by the Bielefeld Academic Search Engine. However, Master’s theses can be an excellent showcase for the academic Master programs and research. They are original work
of the mind, and at least for the best of them, they contribute to scientific and professional
to open scholarship.

Master's theses – an overview
The Master’s degree dates back to the origin of European universities. Today, it is usually a
second-cycle academic degree requiring previous study at the Bachelor’s level (or
version). It is awarded by universities or colleges upon completion of a one or two years
program demonstrating advanced disciplinary knowledge or professional skills. Master’s
degree names and the structure and duration of Master programs vary according to the
country and even the university; diversity is the rule.

In general, the granting of a Master’s degree requires some kind of thesis or dissertation
on a research or a professional project (internship), where the students have to demonstrate
their ability of independent and rigorous thinking and of complex problem solving, along
with field-specific methodology and knowledge. A recent report issued by the US National
Academies of Sciences, Engineering, and Medicine (2017) highlights the potential scientific
quality of Master’s theses based on substantial undergraduate research experiences defined
as “an inquiry or investigation conducted by an undergraduate student that makes an
original intellectual or creative contribution to the discipline”[2].

Alongside research capacities and originality, quality is assessed in terms of innovation
and discovery, refinement of experimental design, new research techniques, teamwork and
writing skills. “As Master’s theses aim to provide evidence of the skills learned, both
practical and theoretical skills are exemplified in theses” (Seliger, 2015, p. 131). Their quality
is impacted by the Master program and in particular, by the importance of research.
Master’s theses are a “complex genre”, considering both the set of problems related to its
configuration (structure, language, norms of reference) and the factors that constrain its
production (methodological procedures, student/supervisor relationship, time management
etc.) (Carvalho et al., 2017).

As mentioned above, few repositories contain Master’s theses. Out of the 3,519
repositories of the OpenDOAR directory, 57 per cent contain ETDs but only 0.5 per cent
clearly indicate that this means also Master’s theses. The international Registry of Open
Access Repository Mandates and Policies (ROARMAP) with 941 open access policies keeps
a record of 37 per cent mandates and instructions regarding ETDs but nothing about
Master’s theses. Many institutions do not consider undergraduate work as research output
worth open scholarship. Those which do have different options. Three vectors of open
access dissemination via repositories can be distinguished.

Institutional repositories: Some institutions disseminate their undergraduate students’
theses and dissertations in their own institutional repository, as part of its intellectual
output and along with articles, communications, working papers etc. For instance, in the
University of Amsterdam, Scripties repository Master’s theses represent 70 per cent of the
nearly 50,000 undergraduate works[3].

General aggregating repositories: These multi-institutional or consortial repositories
aggregate files and metadata from different document types and institutions. The Latin
America portal LA Referencia for instance provides access to nearly 1.5 million publications,
one third being Master’s theses, mostly from Brazil[4].

ETD repositories: As a specific variant of the multi-institutional repositories mentioned
above, they only contain ETDs from different institutions, like the OhioLINK Electronic
Thesis and Dissertation Center with over 90,000 undergraduate, Master’s and doctoral theses and dissertations from 32 universities and colleges[5] and the Scandinavian DiVA portal for academic publications and student papers produced at 44 universities, colleges of higher education, research institutes and museums, with more than 250,000 student papers from BA, Master and other diplomas of which 72 per cent are in open access[6].

Because of their number, diversity and quality, Master’s theses prove their usefulness for studies on the evolution and distribution of research subjects and methodologies. A synthesis of published research with samples of Master’s theses reveals three different approaches (for more details and further reading, see annex).

Content analysis: Nearly 20 recent papers from Iran, Slovenia, Brazil, China, Zambia, Turkey and other countries share an exploratory-descriptive approach and aim at assessing research paradigms, the diversity and trends of research topics, writing skills and information literacy, the students’ background and the choice of research methods like qualitative or action research. They also share the basic assumption that Master’s theses generally display sufficient quality and relevance for content analysis and scientometrics.

TDM: Some studies, especially from China, apply TDM techniques to Master’s theses, such as quantitative analysis of the frequency of key words, co-word matrixes, co-occurrence network analysis, factorial or cluster analysis, strategy coordinate analysis and multidimensional scaling analysis. Their objective is the assessment of research topics (“hot issues”), methodologies or funding trends through the automatic exploitation of Master’s theses.

Citation analysis: Citation analysis constitutes a third approach to Master’s theses. These studies are based on the references produced by the Master students and assess the use of scientific literature in terms of document types, preferred journals, geographic origin of cited sources, and so on.

Along with the potential for scientometrics and TDM, there are other reasons for open access to Master’s theses, such as enhancing accessibility, increasing the impact in scientific journals and conferences (Gillet et al., 2013), preventing plagiarism, strengthening visibility for Master’s institutions (Xia and Opperman, 2010), contributing to the curriculum quality and fostering the potential use of outcomes especially from less known programs and disciplines (Martens, 2011).

Quality is essential for the interest and potential of Master’s theses. Not all Master’s theses reflect substantial undergraduate research experiences or professional projects. Poor quality of abstracts can be a limit to TDM. Evaluation can be helpful to control the quality and to select the best papers. Even if the criteria are neither homogeneous nor applied in the same way everywhere, with a large diversity of written guidelines and evaluation grids, of the supervisors’ role and of oral exams to adjust the final mark (Bohn and Hasselgreen, 2011), the evaluation generally includes some similar features, such as background knowledge, description of methods, data and findings, discussion of the findings and the use of references. Sometimes, the Master assessment includes research skills and the relation with the supervising research team. And even if their quality may vary and should be controlled, their diversity and richness make them interesting and useful resources for citation analyses, for content mining and scientometric assessment. But this potential requires open accessibility and reusability. The Lille project MemorySID provides some insight how to proceed.

The Lille project MemorySID
The following section describes the project MemorySID at the University of Lille (France), with nearly 70,000 students the largest French University and one the largest French-
speaking universities worldwide. The project started in 2008, to contribute to open scholarship through open access publishing of Master’s theses in library and information sciences, via a national infrastructure in the public academic cloud. With the years, the project rationale shifted from self-archiving to a two-layer approach with two digital libraries, one in intranet and the other in the cloud.

**Phase one (2008-2015): self-archiving**

Ten years ago, from 2008 on, the Department of Information and Documentation Sciences (SID) at the University of Lille (France) started encouraging the Master students to deposit their theses on a national repository for Master’s theses library and information sciences called *mémSIC*. The objective was to publish a selection of good and non-confidential works in open access and to guarantee long-term findability, accessibility and preservation through a national OA infrastructure, to valorize the students’ work and to increase the Master programs’ visibility (Chauvin *et al.*, 2010; Mann, 2010). The three guiding principles for this first project phase were:

1. validation by the jury (faculty);
2. double authorization by the student and by the project tutor (intership); and
3. self-archiving by the students.

The academic cloud was preferred to a local, institutional solution to optimize visibility and usage. The success was mixed. Between 2008 and 2015, 70 theses were deposited on *mémSIC*, with scores equivalent to the US academic A and B grades (for French readers: 14/20 or more). The impact in terms of usage was more than satisfying, with statistics going up to several thousand downloads per paper (see below). However, the main problem was the lack of motivation for self-archiving and, in some cases, the lack of quality of metadata and files. Another problem was that *mémSIC* does not allow the creation of an institutional collection or portal, thus reducing the impact for the Master program.

**Phase two (from 2016 on): digital library**

To resolve these issues, a follow-up project was launched in 2016 called *MemorySID* (Vanacker, 2017). For this second phase, the priorities were different:

- retrodigitization of all print Master’s theses still in the academic library and department holdings;
- integration of all digitized Master’s theses into the campus-wide Nuxeo document management system (DMS), accessible for students and scholars via their intranet workspace; and
- creation of a new collection on the national repository DUMAS, with a selection of the best theses (“showcase”).

In other words, we privileged a two-layered approach with two digital libraries, one as a folder in the academic intranet and the other as a portal in the cloud, on the national ETD repository DUMAS (Figure 1). DUMAS is hosted by the French Center for Direct Scientific Communication[8] which operates also the French national open repository HAL, and it contains more than 19,000 Master’s theses. Only library and faculty can deposit files and create metadata. There is no self-archiving. Each deposit needs validation by the repository’s administrator. Insofar, DUMAS is a kind of crossover between an open repository and a digital library.
After removing duplicates, 765 print theses were digitized, 583 from the library holdings (1996-2004) and 227 from the SID department (2005-2016). The text files (PDF) were uploaded in the DMS and indexed in a Dublin Core (DC) compliant metadata format. To the 765 print works from 1996-2016, 9 native digital Master’s theses were added in 2017. From these 774 Master’s theses, 715 are searchable and available for all students and scholars on the campus, via the academic virtual workspace, while the others remain in the “dark” part of the intranet because of their confidential content. The digital library (folder) is administered by the faculty (SID Department), while the content of the DMS is maintained and backed-up by the university IT Department.

From the 715 Master’s theses in the public DMS, 87 per cent have a sufficient score (grade A or B) and are authorized for online dissemination, according to the administrative records (i.e. defense minutes). So far (July 2018), 523 theses from the DMS have been deposited on the DUMAS repository and are online. Additionally, the metadata from 64 self-archived theses on mémSIC (first project phase) were transferred to DUMAS. This means that at present, the new Lille LIS Master’s theses collection on the DUMAS repository provides unrestricted access to 587 theses from 1996 to 2017.

**Usage of the DUMAS collection**

The following usage statistics are from July 19, 2018. They assess the download figures for all 587 Master’s theses available on DUMAS, for the period from January 2015 to June 2018.

Cumulative figures: up to now, the 587 items have been downloaded 296,074 times (Figure 2). The median download statistics for the last 12 months are 3,855 per month, and they are continuously increasing.

Usage statistics per item: the analysis of the usage statistics reveals that all deposited Master’s theses, without exception, have been accessed and downloaded (Figure 3).

Figure 3 shows that the differences of usage among the Master’s theses are very important, ranging from 10 to more than 60,000 downloads, with a median of 43. At least two factors may explain these differences:

1. The date of deposit and publishing: the items with more than 10,000 downloads have all been deposited between 2009 and 2011, while the items with low download figures are all from 2017 and 2018.

2. The topics: probably, theses with “hot issues” will be more downloaded. But the real impact of the content on usage statistics is difficult to assess and surely affected by other variables like the quality of indexing and referencing.

Information of interesting and recent deposits is published via social media, e.g. Twitter, Facebook and LinkedIn; metadata, links and the full text can be discovered on sites like Scoop.it, CiteULike or portals for students. However, without Digital Object Identifier

![Figure 1. Two-layered approach of MemorySID](image)
minting and a more sophisticated altmetrics tool it is impossible to assess the real impact of the deposits inside and outside of the academic communities.

Where do the successful download requests come from? The DUMAS repository identifies the country of nearly 80 per cent downloads.

In total, 30 per cent successful download requests come from France, another 13 per cent are from the three French-speaking Maghreb countries Morocco, Algeria and Tunisia. The USA represents 3 per cent, Germany and Canada 2 per cent each, the UK, South Africa and China 1 per cent each. The member states of the so-called “Francophonie” represent 60 per cent of the download requests, i.e. 75 per cent of all geographically identified downloads. However, the collection’s outreach is global and not limited to the French-speaking world: the complete download list contains 186 countries from all continents and regions, including a long list of countries and states like Mongolia, Trinidad and Tobago, Surinam and the Vatican.
Discussion

MemorySID is a faculty-based project in social sciences and humanities, launched and undertaken by the SID department staff to increase the outreach and impact of the Lille LIS Master program. From the beginning, it raised questions and problems that had to be fixed by the faculty, in partnership with the academic library, the IT department and the host of the national open repository. Four issues appear crucial for the future development of open and unrestricted dissemination of Master’s theses.

IP clearance

Intellectual property (IP) is an important issue. Open access to Master’s theses via open repositories means uploading and publishing, and both need authorization from the rights holder and, at least in France where dissertations are considered as administrative proof for the diploma, also from the institution.

Once the student has left the Master program, this IP clearance is difficult to obtain and the risk is high that time produces an increasing number of “orphan theses”. The best moment is at the defense (viva), when the student presents the results of her or his research work and internship. This is the opportunity for the student and the jury to sign the authorization for open access publishing, as part of the minutes of the thesis defense. The signed minutes are archived by the department and serve as proof for the dissemination via the DMS and the open repository.

But this is not enough. Some theses contain sensitive information or personal data while others include material protected by third party rights (photos, maps etc.). Also it may be appropriate to request a third authorization from the professional tutor of the internship. If the student cannot guarantee full rights clearance for included material, it will be prudent either to drop this material or to exclude the thesis from open access.

At present, of all theses preserved in the DMS, four out of five are freely shared on the internet, while 12 per cent are restricted to on-campus access via the intranet and 8 per cent are in the DMS but not available at all (Figure 4).

The key factor for the IP-related process is the faculty support, i.e. the personal and professional commitment to open scholarship which implies two “upstream actions”:

(1) A clear and explicit communication about open repositories, open scholarship and sharing.

![Figure 4. Status of Master’s theses in the DMS (N = 774)](image-url)
Guidelines about how to write a Master’s thesis, how to use protected third-party material, how to handle confidential and sensitive information, personal data (privacy issues), etc.

Sometimes, fears are expressed that unrestricted dissemination of Master’s theses may increase the risk of plagiarism. Indeed, this “collateral damage” cannot be excluded but on the other hand, open access to Master’s theses will improve the performance of anti-plagiarism software.

Quality, metadata and identifiers
The second issue is quality. While all Master’s theses should be preserved in the intranet DMS, only the best ones should be disseminated in open access, for two reasons:

1. Insofar, as the collection provides an excellent showcase and high visibility for the Master program, the dissemination of poor papers should be avoided.
2. To foster the potential for TDM, particular attention should be paid to the quality of metadata.

The quality issue has a paradoxical effect: if the deposit is limited to the best (“outstanding”) Master’s theses, only a small part will be available, reducing their diversity and representativeness. On the other hand, a selective “institutional quality label” will increase their value for the human reader and the machine.

At Lille, the threshold for online dissemination is set at 14 marks out of 20, which is similar to the US A and B grades. This score is the result of collegiate evaluation by at least two scholars, applying the usual criteria, e.g. quality of the analysis, independent thinking, presentation, structure, language and style and references. Nevertheless, the score is only one criteria and the final decision is taken at the defense (viva), by the jury.

The format and quality of metadata remains a challenge. Standard metadata improve findability and interoperability. Both systems, the DMS (Nuxeo) as well as DUMAS, make use of the Dublin Core (DC) elements. To limit inconsistent indexing, some elements were refined, such as rights (level of availability), source (Master program, level, date of the viva), description (author’s abstract) and filename.

The DUMAS metadata schema follows the Text Encoding Initiative format of the French national HAL repository with some specific elements for Master’s theses, e.g. the name of the Master program with its subdivisions, specialties and options. The problem is that the DMS and DUMAS follow different schemas, with different refinements and values, and that their metadata silos are not connected. This means the creation of two different metadata sets for the same document, one for the DMS and the other for the repository. The lack of interconnection and interoperability is a source of inconsistency and errors, causes double workload and requires more post-deposit curating than necessary. A three-level routine reduces the metadata heterogeneity and error level:

1. Inconsistencies and anomalies are rejected by the DUMAS administrator during the validation.
2. The index of authors and directors’ names is periodically controlled by the local administrator.
3. The tag cloud is checked from time to time for synonymy, variants etc.
A last comment on unique identifiers: DUMAS systematically attributes a specific HAL identifier to each new deposit. To track impact on social media (altmetrics), it may be interesting to consider additionally digital object identifier minting.

**Text and data mining**

TDM tools and methods are already applied to Master’s theses, especially to assess research topics, methodologies and funding trends. In Lille, a research team started to mine PhD and Master’s theses in exploratory research projects, to identify specific topics in different disciplines like Agronomics, Cultural Studies or Information Sciences, and geographical information related to the concept of territory (Kergosien et al., 2018).

These studies confirm that the diversity and volume of Master’s theses bear interesting potential for TDM and can produce relevant new knowledge about research topics, methods, people, etc. As the output of TDM is conditioned by the quality of the corpus, this means that special attention should be paid to the quality of titles, abstracts, key words and the structuration, annotation, indexing and open formats, whenever possible.

In an uncertain legal environment of TDM like in France, it would be appropriate and helpful to publish Master’s theses under a liberal open license (CC-BY or CC-BY-SA) to foster their reuse and exploitation; however, this remains an exception up to now.

**Preservation**

Master’s theses are part of what is called gray literature, not easy to identify, not always well controlled by catalogs and databases and often at risk of disappearance (Schöpfel and Farace, 2010). In France, Master’s theses must be conserved for at least five years, as a proof of graduation. MemorySID adopts a three-level strategy to guarantee long-term preservation at least for the best Master’s theses:

1. Remaining print copies are preserved by the faculty administration during the obligatory five-year period, in the facilities of the SID Department.
2. All digital copies are preserved in the university DMS and maintained by the IT Department for at least 5-10 years, the usual lifespan of documents in intranet.
3. As for the selection of best and non-confidential Master’s theses, they are preserved in the DUMAS open repository. DUMAS is part of the French national open scholarship infrastructure HAL which is backed up by the French National Computing Center for Higher Education[9] specialized in long-term preservation of digital library resources, research data etc.

Thus, the DUMAS solution includes outsourcing of the long-term availability of Master’theses, in a dark archive in the public academic cloud which can be used as a failsafe during disaster recovery.

**Conclusion**

The environment of open scholarship provides excellent opportunities for a new approach to the management of Master’s theses. Based on the Lille project and other studies, the following matrix sums up some key factors for the contribution of Master’s theses to open scholarship:

- Strength: their diversity, richness and representativeness make Master’s theses interesting for reading, content mining, citation analysis, etc. which is an important feature of the emerging landscape of open science. Also, they contribute to the visibility and promotion of the institution and the Master program.
- Weakness: the main problem is related to their unequal quality. Sensitive content (personal data, corporate information etc.) is another problem.
- Opportunity: the European and French open science policy creates a favorable environment for the dissemination of Master’s theses on open repositories; this is emphasized by the institutions’ interest to promote and valorize their Master programs.
- Threat: similar to PhD dissertations (Schöpfel and Prost, 2013), the major threat is access restriction because of intellectual property, i.e. no authorization by students. Restrictive policies of institutional repositories (“scientists only”) and lack of interest for undergraduate work are other barriers.

A collection of Master’s theses in the public academic cloud is a unique showcase for the academic excellence of Master programs, of teaching and research. The download statistics show that its impact and visibility are real and significant. It can also produce a rich and representative corpus for TDM and scientometrics. However, this potential is conditioned by a couple of key variables, such as a selective approach, long-term preservation and free dissemination on open repositories. It is only when these conditions are met, that Master’s theses fully contribute to open scholarship.

Notes
1. In France for instance, Master students represent 35% of all university students, compared to 4% PhD students, in other words, the relation is 10 to 1. Statistics from 2017, covering all disciplines including medicine, produced by the French Ministry for Higher Education and Research www.enseignementsup-recherche.gouv.fr/cid126315/reperes-et-references-statistiques-edition-aout-2017.html
2. Council of Undergraduate Research www.cur.org/
3. Scripties www.scriptiesonline.uba.uva.nl
4. LA Referencia www.lareferencia.info
5. OhioLINK ETD Center https://etd.ohiolink.edu/
8. CCSD www.ccsd.cnrs.fr/

References


Further reading


Corresponding author
Joachim Schopfel can be contacted at: joachim.schopfel@univ-lille.fr

For instructions on how to order reprints of this article, please visit our website: www.emeraldgrouppublishing.com/licensing/reprints.htm
Or contact us for further details: permissions@emeraldinsight.com
Awareness, knowledge and attitude of lecturers towards institutional repositories in university libraries in Nigeria

Francisca Nwakaego Okoroma
University of Ibadan, Ibadan, Nigeria

Abstract

Purpose – Content recruitment is a very crucial aspect of institutional repository (IR), and yet, a major challenge in the implementation of an IR system. Literature highlighted the numerous issues associated with IR implementation in the Sub-Saharan Africa, including Nigeria, and the necessity to review and tackle the divergent challenges of institutional repositories. This study aims to address the issues of awareness, knowledge and attitude of lecturers towards institutional repositories in university libraries in Nigeria.

Design/methodology – The study adopted the descriptive survey research design. A multistage sampling procedure was used to sample 751 lecturers of the universities in Nigeria and eight heads of the digitization section of the libraries. Data generated were analyzed using descriptive statistics.

Findings – The findings revealed that majority of the lecturers in Nigeria are either unfamiliar with the term IR or have very little level of knowledge on the aims and objectives of IR, and therefore have low disposition to submitting their work. The study identified various options to resolve the problem of awareness, knowledge and pessimistic attitude of lecturers towards IRs, including massive awareness and enlightenment programmes on the subject of IR and on copyright issues within the university communities in Nigeria.

Originality – The study will enhance the awareness and advocacy of IR in Nigeria. The awareness and knowledge of IR will bring about positive attitudinal change of lecturers towards investing into IRs.

Keywords Awareness, Nigeria, Knowledge, University libraries, Institutional repositories, Attitude of lecturers

Paper type Research paper

Background to the study

The essence of establishing universities is to advance the society through education, research and manpower training. The university can hardly achieve these goals and make any impact on the society without its intellectual output being well accessed and utilised by the citizens and decision makers. Universities are, therefore, expected to facilitate the documentation, preservation and dissemination of the intellectual output of their faculty members, students and other staff for optimal access and utilisation in order to realise their goals. The branch of the university responsible for the management, provision and dissemination of information to support the effective and expeditious attainment of the objectives of the university is the university library. In this era of resource constraint, it is evident that many university libraries are beset with dearth of information resources, especially in the local content, which includes theses, dissertations and faculty research works. There is a huge glaring divide between the explosive output of literature in the universities and the users of
information, including decision makers. A survey conducted by Shulenburger (2007) on strategies for disseminating the intellectual products of researchers indicated that only a minute fraction provided any kind of affirmative response. The implication of the result is that the dissemination of scholarship is left to the individual researcher.

Most African countries spend much resource on research, yet only a few individuals access the results, especially the grey materials (Aina, 1995; Okoroma, 2011). For instance, the National Universities Commission (NUC, 2016) reported that Nigeria has 141 universities, more universities than any country in Sub-Saharan Africa, and these universities function as focal points for academic research in the country. This makes the volume of research output originating from academic institutions and addressing local problems in Nigeria very high. Yet, many of these research outputs addressing issues endemic to the region are dwelling in obscurity. As a result of restriction of access to knowledge, there is a restriction in the development of science and technology which has severe effects on the society. Open access (OA) and open content emerged as an intervention in the area of electronic journals.

The idea of OA was first noted in the mid-1990s. The main objective of the OA initiatives is to remove the prevailing barriers and promote unhindered access to electronic articles and scholarly communication. Institutional repository (IR) was an outgrowth of the OA Initiative. IR is a system which facilitates the capture, storage, preservation and dissemination of an institution’s intellectual outputs in an electronic form (Rosenblum, 2008). Dhanavandan and Tamizhchelvan (2015) define IR as an online archive for the intellectual output of a particular institution for collecting, preserving and disseminating the research community which can be viewed as a set of services that institutions offer to members of the academic community to maintain and provide the digitised materials. Therefore, IRs play an important role in the preservation and dissemination of institutional research outputs, which in turn becomes a constituent part of a global research output (Crow, 2002). IRs not only act to preserve an institution’s intellectual work product but will equally contribute to a fundamental, albeit long-term change in the structure of scholarly communication.

According to Foster and Gibbons (2005: 11), “libraries build institutional repositories because they provide an institution with a mechanism to showcase its scholarly output, centralise and introduce efficiencies to the stewardship of digital documents of value, and respond proactively to the escalating crisis in scholarly communication”.

With the evident innumerable benefits of IRs, universities and other academic institutions all over the world are embracing IR as a means of bridging the gap between the authors, scholars, researchers’ output and the various users of information, as well as to preserve their wealth of knowledge. For instance, the concepts of OA and IR have come to the forefront of Turkish information management landscape with the formation of the Consortium of Anatolian University Libraries (ANKOS) in April, 2004. The Middle East Technical University Library Electronic Theses and Dissertations Archive was established in 2003 and became the first Turkish member of the Networked Digital Library of Theses and Dissertations. Ohio State University’s knowledge bank provides another example of a project that addresses the objectives of IR while serving broader digital resource goals. Growing out of the University’s Distance Learning/Continuing Education Committee, the knowledge bank plans to include all of the digital assets and information services available to the Ohio State University (OSU) community, whether created by OSU constituents or not.
Christian (2008) reported that institutions in South Africa seem to be making greater progress in terms of development and deployment of institutional repositories, while their counterparts in Nigeria are bugged down by a complex combination of problems. The Open DOAR (2014) also ascertained a dismal record of adoption of IRs in Nigeria. Out of 95 OA IRs in Africa, Nigeria has only eight (7.9 per cent), while South Africa has 28 (31.5 per cent). Eke (2011) highlighted some of the challenges associated with the deployment of IRs in Nigeria as lack of awareness, attitude of academics, legal aspect/copyright, funds, technological, infrastructural issues and policy development. Content recruitment is nevertheless a fundamental aspect of IR (there cannot be IR if the content is not there in the first place), yet a survey of ARL libraries identified content recruitment as the number one challenge in the implementation of an IR (Bailey et al., 2006). Earlier studies consistently reported that recruiting content is difficult (Ware, 2004; Rowlands and Nicholas 2005; Lynch and Lippincott, 2005; Heery and Anderson, 2005; Ware 2006; Davis and Connolly, 2007; Salo, 2008). Harnad (2009) also indicated difficulties in populating IRs and disciplinary repositories through voluntary submissions have led to proselytizing for mandates. Yang and Li (2015) collaborated this view in his study that “Despite their [TAMU faculty] positive attitudes towards OA publishing, they are not so positive towards OA mandates”. Salo (2008) overviews of the state of IRs, painted an abysmal picture of lecturers disinterest in populating IRs. If there is a nonchalant attitude towards IRs by the lecturers who are the major contributors to IRs, how then can the system be sustained? IRs cannot continue to exist without the resolute and persistent support of the lecturers, hence the need to look into the attitude of the lecturers towards their IRs.

There is no doubt that the disinterest of lecturers in an IR has to do with their mindset. Attitude has been defined by scholars in various ways. Brown defined attitude as “a combination of beliefs, thoughts and feelings that influence an individual to react in a positive or negative way to objects, people, processes or institutions”. Attitude is a kind of mind-set, it can be changed due to the fact that it was presented by particular stimuli; thus, the orientation and reorientation of the mind can change negative attitude to positive and vice versa. It is clear that content recruitment for IRs is contingent on attitudinal change of lecturers and other authors towards IRs. This also depends on the reorientation of IR contributors through awareness and enlightenment programmes on IRs. To this end, the awareness and advocacy on IR area prerequisite in building trust with lecturers and other researchers who play a key role in populating the repositories. Covey (2011) collaborated that “recruiting a critical mass of content for the institutional repository is contingent on increasing awareness, aligning deposit with existing workflows, and providing value-added services that meet needs not currently being met by other tools”.

Nwokedi (2011) evaluated University of Jos lecturers’ knowledge of the existence of IR and willingness to submit research works and found that majority (79 per cent) of the respondents did not have any idea of OA IR, and only 21 per cent of the respondents claimed to be aware of the existence of IR in their institution through a seminar on IR by the library. However, when the benefits of IR were explained to them during the research, 91.6 per cent of them then indicated that IR was very useful. This underscores the role of awareness and knowledge of IR in the attitudinal change of faculty members towards IRs. This study seeks to investigate the awareness, knowledge and attitude of lecturers towards IRs in university libraries in Nigeria.
Objectives of the study
The main objective of the study is to examine the awareness, level of knowledge and attitude of lecturers towards IRs in university libraries in Nigeria.

Research questions
The following research questions are identified based on the objectives of the study.

RQ1: What is the level of awareness of an IR by lecturers in universities in Nigeria?

RQ2: What is the degree of lecturers’ knowledge of IR?

RQ3: What is the attitude of lecturers towards submitting their scholarly research results to the IRs?

RQ4: What are the solutions to the problem of awareness, knowledge and attitude of lecturers towards IRs in university libraries in Nigeria?

Methodology
The study adopted a survey research design. It involves quantitative and qualitative methods of data collection. The study adopted a multistage sampling procedure. The purposive sampling technique was used in the selection of eight institutions that participated in the study. These institutions were Ahmadu Bello University, Covenant University Ota, Ogun State, Federal University of Technology, Ondo State, Obafemi Awolowo University, Osun State, University of Ibadan, Ibadan, Oyo State, University of Jos, Plateau State, University of Nigeria, Nsukka, Enugu State and the Federal University, Oye Ekiti, Ekiti State. These institutions were purposively selected because they were the only universities adopting an IR in Nigeria. The study also adopted the purposive sampling technique to sample three (3) faculties of the institutions. These faculties were the faculties of sciences, social sciences and humanities, respectively. The study also used the simple random sampling techniques of the balloting type to identify three departments, each that represents each of the selected faculties in the study. The study used the simple random sampling techniques to identify 751 lecturers who participated in the study.

The survey questionnaire adopted for the study was a self-designed structured questionnaire to find out the awareness, knowledge and attitude of lecturers to an IR. This instrument consists of four sections: A, B, C and D. Section A of the instrument elicited demographic information of the respondents. Section B has four questions asking about the familiarity of the term “IR”, awareness of the aim and objectives of IR and a total of 14 statements on a four-point scale. Section C has a total of 12 statements on a four-point scale probing the full agreement of lecturers regarding the sustenance of IRs and their willingness to submit their research output to IRs, while section D measured the proposed solution to the problem of awareness, knowledge and attitude of lecturers to an IR.

Data collected were analysed using descriptive statistics which involves frequency counts and simple percentages.

Results
Demographic representation of the respondents
Figure 1 reveals that, amongst the faculties used in the study, lecturers from the faculties of the humanities were more 281 (37.4 per cent) followed by those in the social science 253 (33.7
per cent), and then sciences 217 (28.9 per cent). Humanities has more participants, with 66 from University of Nigerian (UNN), 59 from Amadu-Bello University (ABU), 45 from University of Jos (UNIJOS), 45 from University of Ibadan (UI), 12 from Federal University, Oye-Ekiti (FUOYE), 22 from Obafemi Awolowo University (OAU), five respondents from Federal University of Technology, Akure (FUTA) and 21 respondents from Covenant University (CU). For social science, ABU has 66 sampled respondents; UNIJOS has 55 respondents; and UNN has 40 respondents. While for the science, UI has 37 respondents, FUOYE 37, ABU has 35 respondents; UNN 34 respondents and 26 respondents were from FUOYE. Therefore, more of the respondents that participated in the study were from humanities, followed by those from social science and then those from science department.

*RQ1*: What is the level of awareness of IR by lecturers in universities in Nigeria?

Data on the level of awareness of institutional repository by lecturers in universities in Nigeria are presented on Table I.

![Figure 1. Distribution of respondents by faculty](image)

<table>
<thead>
<tr>
<th>Statement</th>
<th>Strongly agree</th>
<th>Agree</th>
<th>Disagree</th>
<th>Strongly disagree</th>
</tr>
</thead>
<tbody>
<tr>
<td>I am aware of the existence of my university IR</td>
<td>190 (26%)</td>
<td>337 (46%)</td>
<td>135 (18%)</td>
<td>79 (10%)</td>
</tr>
<tr>
<td>I am aware of the benefits of IR</td>
<td>165 (22%)</td>
<td>338 (46%)</td>
<td>159 (22%)</td>
<td>77 (10%)</td>
</tr>
<tr>
<td>I am aware of the content of my university’s IR</td>
<td>135 (18%)</td>
<td>264 (36%)</td>
<td>227 (31%)</td>
<td>111 (15%)</td>
</tr>
<tr>
<td>I am aware of my university’s IR policy</td>
<td>132 (18%)</td>
<td>238 (32%)</td>
<td>242 (33%)</td>
<td>123 (17%)</td>
</tr>
<tr>
<td>I am aware of the publishers’ policy on OA</td>
<td>116 (16%)</td>
<td>257 (35%)</td>
<td>245 (34%)</td>
<td>107 (15%)</td>
</tr>
<tr>
<td>How did you originally learn about IR</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I read about IR</td>
<td>121 (17%)</td>
<td>250 (35%)</td>
<td>199 (29%)</td>
<td>123 (18%)</td>
</tr>
<tr>
<td>From information provided by other institutions</td>
<td>111 (15%)</td>
<td>231 (33%)</td>
<td>231 (33%)</td>
<td>129 (18%)</td>
</tr>
<tr>
<td>Information about IR through my colleagues influenced my awareness</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The IR awareness programmes on mass media</td>
<td>117 (16%)</td>
<td>237 (33%)</td>
<td>234 (32%)</td>
<td>124 (17%)</td>
</tr>
<tr>
<td>My departmental meetings consistently remind me</td>
<td>90 (13%)</td>
<td>186 (26%)</td>
<td>264 (37%)</td>
<td>175 (25%)</td>
</tr>
<tr>
<td>Our university library presentations on IR</td>
<td>89 (13%)</td>
<td>179 (25%)</td>
<td>262 (37%)</td>
<td>182 (26%)</td>
</tr>
<tr>
<td>The university workshops on the importance of IR in scholarly communication</td>
<td>125 (17%)</td>
<td>224 (31%)</td>
<td>223 (31%)</td>
<td>141 (20%)</td>
</tr>
<tr>
<td>Librarians in my institution brought to my attention IR publishing initiatives</td>
<td>137 (19%)</td>
<td>183 (26%)</td>
<td>235 (33%)</td>
<td>155 (22%)</td>
</tr>
<tr>
<td>I have not heard about it</td>
<td>164 (23%)</td>
<td>209 (29%)</td>
<td>191 (27%)</td>
<td>146 (21%)</td>
</tr>
<tr>
<td></td>
<td>122 (17%)</td>
<td>129 (18%)</td>
<td>230 (32%)</td>
<td>229 (32%)</td>
</tr>
</tbody>
</table>

Table I. Awareness of IR by lecturers in universities in Nigeria
In determining the level of awareness of respondents on IR, descriptive statistics were computed and the result reveals that a total of 72 per cent of the respondents opined that they are aware of the existence of their universities' IRs, 67 per cent are aware of the benefits of IR, while 53 per cent are aware of the content of their universities' IR. Also, 50 per cent are aware of their universities' IR policy, and nearly half of the respondents (51 per cent) report that they are aware of their universities' IR policy and publisher's policy on OA.

Table 1 also shows how the respondents learnt about IR. The result reveals that only 52 per cent of the respondents have read about IR, 50 per cent got information about IR through colleagues, 55 per cent indicated that the awareness was gotten through the universities' workshops on the importance of IR in scholarly communication, while 53 per cent got their awareness through librarians in their institutions who brought to their attention the IR publishing initiatives. About 51 per cent got to hear about IR from the information provided by other institutions, 63 per cent of the respondents received IR awareness through the mass media, whereas 62 and 51 per cent attested that it was through departmental meetings and university library presentations on IR, respectively. The implication of the result is that the level of awareness of IR in Nigeria is inadequate because as many as (251) 35 per cent of the respondents ascertained that they have not heard about IR at all. The rating of the level of awareness of IR by lecturers in universities in Nigeria is represented in Figure 2.

The results in Figure 2 highlighted the respondent’s self-evaluation of their awareness of IR. Only 17 per cent of the respondents report that their level of awareness of IR is highly adequate, whilst almost half of the respondents (56) report that their level of awareness of IR is not adequate. This implies that, amongst the Nigerian university lecturers, the rate of awareness of IR is still poor. This finding agrees with Nwokedi (2011) who found that the majority (79 per cent) of their university community did not have any idea of OA IR. If poor awareness is still existing amongst lecturers in the universities in Nigeria that have already started IR implementation, what then will be the situation in other universities that have not started at all? This accounts for low adoption and sustenance of IR in universities in Nigeria, as an institution’s lecturers staff supposed to be a key contributor to IR sustenance, but where there is lack of awareness by the lecturers’ of their own universities’ IRs, how can adequate and appropriate support towards IR sustainability be made possible?

**RQ2:** What is the level of lecturers’ knowledge of IR?
<table>
<thead>
<tr>
<th>Universities in Nigeria</th>
<th>Completely unfamiliar</th>
<th>I have come across this concept but know nothing about it</th>
<th>I have come across this concept but know just a little about</th>
<th>I have come across it and know quite a bit about it</th>
<th>I am very knowledgeable about IR</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>UNN</td>
<td>26</td>
<td>18</td>
<td>51</td>
<td>25</td>
<td>20</td>
<td>140</td>
</tr>
<tr>
<td>Federal University, Oye-Ekiti</td>
<td>5</td>
<td>6</td>
<td>12</td>
<td>26</td>
<td>29</td>
<td>78</td>
</tr>
<tr>
<td>Federal University of Technology, Akure, Ondo</td>
<td>8</td>
<td>6</td>
<td>8</td>
<td>12</td>
<td>9</td>
<td>43</td>
</tr>
<tr>
<td>UNIJOS</td>
<td>6</td>
<td>18</td>
<td>43</td>
<td>42</td>
<td>17</td>
<td>126</td>
</tr>
<tr>
<td>ABU</td>
<td>29</td>
<td>59</td>
<td>26</td>
<td>27</td>
<td>19</td>
<td>160</td>
</tr>
<tr>
<td>OAU</td>
<td>4</td>
<td>15</td>
<td>14</td>
<td>8</td>
<td>15</td>
<td>56</td>
</tr>
<tr>
<td>UniBadan</td>
<td>12</td>
<td>20</td>
<td>31</td>
<td>12</td>
<td>29</td>
<td>104</td>
</tr>
<tr>
<td>Covenant University</td>
<td>1</td>
<td>7</td>
<td>16</td>
<td>13</td>
<td>5</td>
<td>42</td>
</tr>
<tr>
<td>Total</td>
<td>91</td>
<td>149</td>
<td>201</td>
<td>165</td>
<td>143</td>
<td>749</td>
</tr>
</tbody>
</table>
Table II shows the level of the respondents' knowledge of IR. About 12 per cent of the respondents reported that they never heard of the term “institutional repository”, 149 (20 per cent) of the respondents have come across the concept of IR but do not know anything about the concept, whilst 91 (12 per cent) are completely unfamiliar with the term. The respondents include UNN 26 and ABU 29. Again, 201 (27 per cent) of the respondents have come across the IR concept but know just a little about it. The respondents include 51 lecturers from the University of Nigeria and 31 lecturers from the University of Ibadan. About 165 persons (22 per cent) ascertained that they know quite a bit about IR, such respondents include lecturers from the University of Jos 42, University of Nigeria 25, Federal University Oye-Ekiti 26 and Amadu-Bello University 27. Only 143 (19 per cent) agreed to be knowledgeable on the topic of IR. Such universities include the University of Ibadan and Federal University of Oye-Ekiti 29, respectively, UNN 20, and ABU 19.

The study shows that 59 per cent of the respondents are either unfamiliar with the term IR or have a little knowledge of the topic, as many have come across the concept of IR, but have little or no knowledge about the subject. By implication, lecturers in Nigerian universities are yet to come to the full awareness and knowledge of IR. For instance, Table III shows that, out of 140 respondents in UNN, only 20 were knowledgeable on the subject of IR, for Federal University Oye-Ekiti, only 29 out of 78 responses. Federal University of Technology, Akure, Ondo, recorded only nine lecturers out of 41 who are knowledgeable on IR. The University of Jos has only 29 out of 126 responses, ABU has 19 out of 160 persons, OAU has 15 out of 56 respondents, UI has only 29 out of 104 respondents who are knowledgeable on IR and Covenant University has just five persons out of 42 respondents. It could be deduced that only a few lecturers in Nigerian universities 143 (19 per cent) are knowledgeable on IR, whilst the majority of the lecturers 604 (81 per cent) are not.

The respondents’ level of knowledge of the objectives of institutional repository are presented in Table III
Table III shows the level of knowledge of the aims/objectives of an IR by the lecturers. 22 per cent of the respondents reported that they are completely unfamiliar and 31 per cent of the respondents ascertained that they have come across the aim of the IR but know nothing about it, such universities are the University of Jos 45, University of Nigeria 32, Federal University, Oye-Ekiti OAU and ABU 23 and University of Ibadan. About 201 (27 per cent) know just a bit of the aims of IR, such include UNN 21, federal university Oye-Ekiti 31. Also, 170 (23 per cent) of the respondents agreed on completely unfamiliar with the aims of IR, such respondents include UI and UNN (44) and ABU (47). This implies that some universities’ lecturers in Nigeria may have had little knowledge about IRs but are not yet familiar with the benefits and policies of its operations in the universities. Furthermore, 79 (11 per cent) have come across the aims of IR, but know just a little about it such universities are UNN 38, University of Jos 52, ABU 50, University of Ibadan 21 and Covenant University 23, whilst only 72 (9 per cent) out of 749 persons agreed to be very knowledgeable about the aims of IR in the university libraries. These are more tangible in Covenant University, University of Jos and Amadu Bello University. Hence, the study reveals that most of the Nigerian universities’ lecturers are not yet familiar with the aims of IR. From the study, 677 (91 per cent) lecturers out of 749 that responded to the questions are not knowledgeable on the topic. This is indeed quite dismal, considering the fact that the universities used in the study are those institutions that have already adopted IR. If the awareness and knowledge of IR are too low on these institutions that have adopted IR, what is the fate of those institutions that are yet to adopt IR?
<table>
<thead>
<tr>
<th>University</th>
<th>Level of knowledge of the objectives of IR by lecturers</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>UNN</td>
<td>Completely unfamiliar</td>
<td>I have come across this concept but know nothing about it</td>
</tr>
<tr>
<td>Federal University, Oye-Ekiti</td>
<td>38</td>
<td>117</td>
</tr>
<tr>
<td>Federal University of Technology, Akure, Ondo</td>
<td>11</td>
<td>111</td>
</tr>
<tr>
<td>UNIJOS</td>
<td>13</td>
<td>52</td>
</tr>
<tr>
<td>ABU</td>
<td>44</td>
<td>50</td>
</tr>
<tr>
<td>OAU</td>
<td>4</td>
<td>17</td>
</tr>
<tr>
<td>UI</td>
<td>44</td>
<td>21</td>
</tr>
<tr>
<td>Covenant University</td>
<td>2</td>
<td>23</td>
</tr>
<tr>
<td>Total</td>
<td>169</td>
<td>229</td>
</tr>
<tr>
<td></td>
<td>I have come across this concept but known little about it</td>
<td>32</td>
</tr>
<tr>
<td></td>
<td>23</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>21</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>23</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>79</td>
</tr>
<tr>
<td></td>
<td>I have come across this concept and known quite a bit about it</td>
<td>21</td>
</tr>
<tr>
<td></td>
<td>31</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td>79</td>
<td>72</td>
</tr>
<tr>
<td></td>
<td>I am very knowledgeable on the objective of IR</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td>72</td>
<td>72</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>140</td>
</tr>
<tr>
<td></td>
<td></td>
<td>78</td>
</tr>
<tr>
<td></td>
<td></td>
<td>43</td>
</tr>
<tr>
<td></td>
<td></td>
<td>126</td>
</tr>
<tr>
<td></td>
<td></td>
<td>160</td>
</tr>
<tr>
<td></td>
<td></td>
<td>56</td>
</tr>
<tr>
<td></td>
<td></td>
<td>104</td>
</tr>
<tr>
<td></td>
<td></td>
<td>42</td>
</tr>
<tr>
<td>Statement</td>
<td>Strongly agree</td>
<td>Agree</td>
</tr>
<tr>
<td>--------------------------------------------------------------------------</td>
<td>----------------</td>
<td>-----------</td>
</tr>
<tr>
<td>Lecturers are yet to come to full agreement regarding the sustenance of IR</td>
<td>215 (29%)</td>
<td>376 (51%)</td>
</tr>
<tr>
<td>Lecturers are reluctant to submit their work in IR</td>
<td>214 (29%)</td>
<td>374 (51%)</td>
</tr>
<tr>
<td>I would consider having a copy of my pre-published research outputs (e.g. thesis) in IR</td>
<td>29 (4%)</td>
<td>104 (14%)</td>
</tr>
<tr>
<td>I would consider having a copy of my article(s) previously published in IR</td>
<td>43 (6%)</td>
<td>122 (16%)</td>
</tr>
<tr>
<td>There should be a national policy explicitly directing or supporting the development of IRs in universities in Nigeria</td>
<td>28 (4%)</td>
<td>118 (16%)</td>
</tr>
<tr>
<td>Technical support is a challenge</td>
<td>212 (29%)</td>
<td>363 (50%)</td>
</tr>
<tr>
<td>There is a security issue</td>
<td>201 (28%)</td>
<td>304 (42%)</td>
</tr>
<tr>
<td>Technophobia is a limitation</td>
<td>278 (38%)</td>
<td>282 (38%)</td>
</tr>
<tr>
<td>There is inadequate users’ education on IR</td>
<td>247 (34%)</td>
<td>305 (42%)</td>
</tr>
<tr>
<td>Submission process is not certain</td>
<td>294 (41%)</td>
<td>312 (43%)</td>
</tr>
<tr>
<td>Withdrawal services are issues</td>
<td>205 (28%)</td>
<td>301 (42%)</td>
</tr>
<tr>
<td>Access control and rights management: to restrict access to the information when OA is premature or not desirable is not certain</td>
<td>240 (33%)</td>
<td>309 (42%)</td>
</tr>
</tbody>
</table>
RQ3: What is the lecturers’ attitude towards submitting their scholarly research results to the IRs in university libraries in Nigeria?

The lecturers’ attitude towards submitting their scholarly research results to the IRs is confirmed by both the lecturers and IR managers in Table IV and the interview responses.

Table IV presents the responses of the lecturers on their attitude towards IRs. The result reveals that 80 per cent of the lecturers are yet to come to full agreement regarding the establishment of an IR, and 80 per cent of the responses were of the opinion that lecturers are reluctant to submit their works in IR, whilst 80 per cent disagreed on the idea of publishing either their preprint or post-print work on IR, respectively. Again 79 per cent of them agreed on technical support as a challenge, whereas 76 per cent established that technophobia is a limitation to lecturers. Furthermore, 75 per cent ascertained there is inadequate users’ education on the subject of IR, 70 per cent agreed that there are issues on withdrawal services, whilst 75 per cent said that access control and rights management: to restrict access to the information when OA is premature or not desirable, is not certain. Finally, 84 per cent of the lecturers asserted that the submission process is not certain, and 70 per cent of the respondents agreed that there is a security issue. The table has revealed so many factors that cumulated to the pessimistic attitude of lecturers to IR.

The findings from the open-ended questions for lecturers and structural interview (with the heads digitisation) on lecturers’ attitude towards submitting their scholarly research result to the IRs revealed difficulties in convincing and getting the lecturers to populate the IRs in their institutions with their works. It was further established that majority of the lecturers are pessimistic on the issue of IR and the submission procedures involved. Others reported not being sure of their conviction to upload their intellectual works to the IRs. They were indecisive and nonchalant. According to the interview respondents, some of the lecturers make comments like “what will be my benefit for the academic rigour? Do you want me to give away my work just like that? “People will copy my work freely”. Some others expressed concern for the publication of works that have not been peer-reviewed. Therefore, university management authorities may need to consider incentives for academics and a team to review the institutions’ works before uploading, to gain faculty compliance to the submission of their works. The thesis and dissertations may not need further review considering that each of the works has been subjected to a thorough and intensive review and scrutiny by the author’s supervisors and other faculty members before approval.

RQ4: What are the solutions to the problem of awareness, knowledge and attitude of lecturers towards IRs in university libraries in Nigeria?

This research question was addressed by both the lecturers and the heads of digitisation units of the universities surveyed.

The recommendations made by lecturers on solutions to the problem of awareness, knowledge and attitude of lecturers towards IRs are itemised in Table V. Majority of the lecturers ascertains that awareness through various media, including seminars, workshops/training for stakeholders, flyers and brochures and social media is highly needed to boost the level of awareness and knowledge of IR in Nigeria. For instance, 84 per cent agreed on workshops/training for stakeholders, 78 per cent opined on awareness through social media, 78 per cent on the use of flyers and brochures, 82 per cent agreed on libraries newsletters and 84 per cent agreed on publicity through the universities’ websites. As regards to copyright issues, 83 per cent of the lecturers
<table>
<thead>
<tr>
<th>Solutions to problem of awareness, knowledge and attitude of lecturers towards IRs in university libraries in Nigeria</th>
<th>Strongly agree</th>
<th>Agree</th>
<th>Disagree</th>
<th>Strongly disagree</th>
</tr>
</thead>
<tbody>
<tr>
<td>Awareness through seminar</td>
<td>345 (47%)</td>
<td>276 (38%)</td>
<td>68 (9%)</td>
<td>38 (5%)</td>
</tr>
<tr>
<td>Awareness through workshops/training for stakeholders</td>
<td>321 (44%)</td>
<td>287 (40%)</td>
<td>90 (12%)</td>
<td>27 (4%)</td>
</tr>
<tr>
<td>Awareness through flyers and brochures</td>
<td>276 (38%)</td>
<td>288 (39%)</td>
<td>118 (16%)</td>
<td>42 (6%)</td>
</tr>
<tr>
<td>Awareness through social media</td>
<td>279 (38%)</td>
<td>289 (40%)</td>
<td>115 (15%)</td>
<td>44 (7%)</td>
</tr>
<tr>
<td>Advocacy through interpersonal communication</td>
<td>256 (35%)</td>
<td>343 (47%)</td>
<td>93 (13%)</td>
<td>34 (5%)</td>
</tr>
<tr>
<td>Advocacy through brief lectures (emphasising benefits)</td>
<td>305 (42%)</td>
<td>295 (41%)</td>
<td>84 (12%)</td>
<td>43 (6%)</td>
</tr>
<tr>
<td>IR presentations at the faculties</td>
<td>297 (40%)</td>
<td>320 (44%)</td>
<td>88 (12%)</td>
<td>22 (3%)</td>
</tr>
<tr>
<td>Libraries newsletters</td>
<td>313 (43%)</td>
<td>283 (38.9%)</td>
<td>103 (14%)</td>
<td>29 (4%)</td>
</tr>
<tr>
<td>Publicity on universities’ websites</td>
<td>312 (43%)</td>
<td>298 (41.0%)</td>
<td>79 (11%)</td>
<td>38 (5%)</td>
</tr>
<tr>
<td>Institution-specific participation incentives for contributors</td>
<td>286 (39%)</td>
<td>305 (42%)</td>
<td>81 (11%)</td>
<td>51 (7%)</td>
</tr>
<tr>
<td>As regards to copyright issues, authors should be educated on their rights in relation to their intellectual output</td>
<td>315 (44%)</td>
<td>286 (40%)</td>
<td>81 (11%)</td>
<td>42 (6%)</td>
</tr>
<tr>
<td>A team of reviewers to review work before added to IR content is necessary</td>
<td>287 (40%)</td>
<td>297 (41.1%)</td>
<td>104 (14%)</td>
<td>34 (5%)</td>
</tr>
<tr>
<td>IR success is dependent upon mandates (given to researchers to submit their works)</td>
<td>279 (39%)</td>
<td>290 (40%)</td>
<td>105 (15%)</td>
<td>46 (6%)</td>
</tr>
</tbody>
</table>
opined that authors should be educated on their rights in relation to their intellectual output. Again, 83 per cent of the respondents concurred that the knowledge of IR can be enhanced through brief lectures (emphasising benefits) and through interpersonal communication. About 82 per cent are of the view that institution-specific participation incentives for contributors should be introduced to encourage authors’ full participation, 81 per cent agree that there should be a team of reviewers to review works before they are added to IRs. Furthermore, 79 per cent of the respondents are of the view that IRs’ success is dependent upon mandates (that is lecturers/researchers within the universities should be mandated to submit their works).

The extracts from the open-ended questions responses by the lecturers identified the following suggestions on the solutions to the problem of development of IRs in university libraries in Nigeria:

- Everybody should be carried along and involved in the publicity of IR (both old and young), including students, business people and academics.
- Incentives in the form of research grants, conference support, etc. should be given to academics to encourage them to contribute to IRs.
- The university system should make contributions to IRs as part of their policy to access research grants in the universities.
- Motivation through reward system, such as compensations and awards for the contributors to IRs, is necessary because the works are the achievement of the authors, and so they need benefits.

As regards to copyright issues, the open-ended questions' responses from lecturers ascertained that:

- There should be a workable policy guiding copyright matters and its implementations in Nigeria.
- Enough legal framework need to be put in place.
- A commission needs to be instituted to monitor and regulate copyright issues within and outside the universities.
- There should be sanctions for those who violate copyright laws, and copyright law observance should be sacrosanct, it should be taken very seriously and enforced to the later.
- Authors are to retain their copyright on submitting to their universities IRs.
- At the point of publishing, authors should request for permission from their publishers to upload their works on their institutions’ IRs.
- Plagiarism and piracy which are fears of lecturers should be discouraged and eliminated by the government through national policy.
- To prevent piracy and plagiarism, only the abstract should be uploaded with the CV of the authors so that whoever wants to make use of the full text will contact the author.
- There is need to ensure security of the work uploaded on the IR.
- The IR management should check against plagiarism before a work is uploaded to the IR.
- A department should be set up within the campus, solemnly to handle and document matters relating to copyright.
The extracts from the responses of the respondents from the heads of digitisation units of the universities further noted (as part of the solution to copyright issues) that: the universities do not take the ownership of the works, authors still maintain their ownership as the sole owners of their works. However, some journals do require formal permission, whilst others have already given out permissions to authors to upload their works on their institutions’ IRs. For those journals that require formal permission, they should be sorted out by the library, persuading faculty members to obtain formal permission from the journals they have given out their copyright to, before those works can be uploaded. This is to ensure lack of bridge of contract and sanctions. Generally, authors are to avoid giving away their copyright to the journal publishers.

Discussion
The study reveals that the awareness and knowledge of IR in universities in Nigeria is still low. Majority of the respondents have very little level of knowledge on IR. The study reveals that most of the lecturers in Nigerian universities are not yet familiar with the subject, including the aims and benefits of IR.

Most of the respondents were either unfamiliar with the term IR or have a little knowledge of the topic. It is worthy of note that many of the lecturers who responded to the questionnaire requested explanation on IR before they could effectively answer the questionnaire questions, and these respondents are situated within the universities that have already established an IR. The question now is what will be the level of awareness on the campuses that have not started the processes of IR? It will definitely be very dismal. This calls for massive advocacy and enlightenment on the subject of IR within the university communities in Nigeria. This finding collaborates the survey conducted by Yang and Li (2015), which revealed a low level of awareness of the TAMU IR by the faculty.

Furthermore, the finding partly agrees with Nwokedi (2011) who evaluated the University of Jos lecturers’ knowledge of the existence of IR and willingness of researchers to submit research works and found that majority (79 per cent) of the respondents did not have any idea of OA IR, and only 21 per cent of the respondents claimed to be aware of the existence of IR in their institution. In contrast to Nwokedi’s findings, 71 per cent of the respondents claimed to be aware that IR is existing in their universities, but the majority are not knowledgeable on the topic at all. The study infers that only 9 per cent of lecturers are knowledgeable on IR, whilst 91 per cent are not.

The digitisation staff reported poor and nonchalant attitude of lecturers towards the submission of their works to IR. Most of the lecturers also reported that they are yet to come to full agreement regarding to the sustenance of IRs and their reluctance to submit their work in IR. This collaborates with Earwage (2008) and Yang and Li (2015) who noted that faculty members have been reluctant in contributing to IRs. Also, Mark and Shearer (2006) opined that, in some universities, faculty members are yet to come to full consensus regarding the establishment of an IR. This attitude can be partly linked to the lack of in-depth knowledge of IR, especially on the benefits of IR, as it does not really make sense for someone to support a course he/she does not understand. Therefore, everybody should be carried along and involved in the publicity of IR, including students, business people, teaching and non-teaching staff in the universities. Enlightenment programmes for the stakeholders on IRs through seminars, fliers, publicity on universities’ websites are very imperative to enhance input and use of the system. There is an urgent need for IR
presentations at the faculties, and authors should be well informed on their rights in relation to their intellectual outputs.

Again, there is need for the faculties to discuss on the establishment and sustenance of IR, with the view to strengthening certain cloudy areas, e.g. plagiarism issues, who owes the copyright? Quality assurance of the contents of IRs through peer review; what will be the compensation of researchers for the academic rigour? As well as the numerous benefits and rewards they stand to gain for the investment they make in the IRs. This agrees with the model by Davis (1985) which postulated that the attitude of a user towards a system will, to a large extent, determine whether the user will actually use the system or not, and the attitude of the user is influenced by perceived usefulness of the system.

**Conclusion**

Awareness, knowledge and attitude of lecturers play a key role in the population and the sustenance of institutional repositories in universities in Nigeria. Therefore if awareness that emphasizes the benefits of institutional repository is intensified in the universities, coupled with incentives for the contributors, lecturers attitude will not only be positively affected, but they will become the key drivers of IRs in their institutions.

**References**


Ware, M. (2006), Scientific Publishing in Transition: An Overview of Current Developments, Mark Ware Consulting, Bristol.


Further reading


Corresponding author
Francisca Nwakaegho Okoroma can be contacted at: frankaonyeka@yahoo.com
Appendix. Research questionnaire on the awareness, knowledge and attitude of lecturers to institutional repositories

Dear Respondent,

Please this questionnaire is designed to elicit information that is purely for research purpose.

Kindly assist in answering the questions objectively. Your confidentiality will be ensured.

Thank you.

Okoroma, F. N.

A. DEMOGRAPHIC INFORMATION

Please complete the spaces where necessary and tick (✓) where applicable.

1. Name of Institution:
2. Name of Faculty:
3. Sex: Male ( ) Female ( )
4. Age: Age of Respondent: 20-30 years ( ) 31-40 years ( ) 41-50 years ( ) 51-60 years ( ) 61 years and above ( )
5. Highest Educational Qualification: (a) First degree
   (b) Master degree ( ) (c) Ph.D. in view ( ) (d) Ph.D ( )
   (e) Others (Please specify) ........
6. What is your designation in your university?
   a. Lecturer II
   b. Lecturer I
   c. Senior lecturer
   d. Professor
   e. Others (Please specify) __________________________

B. Awareness and knowledge of institutional repository (IR) by lecturers:

1. Are you familiar with the term “Institutional Repository”? 
   a. Completely unfamiliar—I have never heard of this term before .
   b. I have come across this concept but know nothing about it .
   c. I have come across this concept and know a little about it.
   d. I have come across this concept and know quite a bit about it.
   e. I am very knowledgeable about institutional repository

2. Are you aware of the aims of IR?
   a. Completely unfamiliar
   b. I know little about it .
   c. I know quite a bit about it.
   d. I am very knowledgeable about the aims of institutional repository

3. Do you know the advantages of IR?

(continued)
a. Completely unfamiliar
b. I know little about it.
c. I know quite a bit about it.
e. I am very knowledgeable about the aims of institutional repository

4. Please indicate your level of agreement on the statement below about your awareness on IR. (Strongly agreed = SA; Agreed = A; Disagreed = D; Strongly Disagreed = SD)

<table>
<thead>
<tr>
<th>Statement</th>
<th>SA</th>
<th>A</th>
<th>D</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>I am aware of the existence of my university IR</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I am aware of the benefits of IR</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I am aware of the content of my university IR</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I am aware of my university IR policy</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I am aware of the publishers’ policy on open access</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

How did you originally learn about institutional repository?

I read about institutional repository

From information provided by other institutions

Information about IR through my colleagues influenced my awareness

The IR awareness programmes on mass media

My Departmental meetings consistently remind me

Our university library presentations on IR

The university workshops on the importance of IR in scholarly communication.

Librarians in my institution brought to my attention Institutional Repository publishing initiatives

I have not heard about it

Others (please specify)……………

5. How can you rate your level of awareness about IR? Please tick the appropriate

(i) Not adequate [ ]
(ii) Barely adequate [ ]
(iii) Sufficiently adequate [ ]
(iv) Highly adequate [ ]

C. Lecturers attitude towards IR

Please indicate your level of agreement on the statement below:

<table>
<thead>
<tr>
<th>Statement</th>
<th>SA</th>
<th>A</th>
<th>D</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lecturers are yet to come to full agreement regarding the sustenance of institutional repository.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lecturers are reluctant to submit their work in IR</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I would consider having a copy of my pre-published research outputs (e.g. thesis) in IR</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I would consider having a copy of my article(s) previously published in IR.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>There should be a national policy explicitly directing or supporting the development of institutional repositories in universities in Nigeria.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(continued)
Technical support is a challenge
There is security issue
Technophobia is a limitation.
There is inadequate users’ education on IR.
Submission process is not certain
Withdrawal services are issues.
Access Control and Rights Management: to restrict access to the information when open access is premature or not desirable, is not certain.

Please specify other issues related to IRs in your university which could be of interest to share with others: 

D. Solutions to the problem of awareness, knowledge and attitude of lecturers towards institutional repositories in university libraries in Nigeria by lecturers.

Please rate the extent to which you agree with each statement ticking (✓) whichever applies:

<table>
<thead>
<tr>
<th>Solutions to problem of IR in university libraries in Nigeria need to include:</th>
<th>Strongly Agree</th>
<th>Agree</th>
<th>Disagree</th>
<th>Strongly Disagree</th>
</tr>
</thead>
<tbody>
<tr>
<td>Awareness through seminar.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Awareness through workshops/training for stakeholders.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Awareness through Flyers and brochures.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Awareness through social media.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Advocacy through interpersonal communication</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Advocacy through brief lectures, (emphasizing benefits).</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IR presentations at the faculties</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Libraries newsletters</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Publicity on Universities’ websites</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Institution-specific participation incentives for contributors.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>As regards copyright issues, authors should be educated on their rights in relation to their intellectual output.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A team of reviewers to review work before added to IR content is necessary.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Institutional repository success is dependent upon mandates (given to researchers to submit their works).</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Please specify other solutions to issues in IRs.

Thank you for filling this questionnaire.
Ensuring access to theses in the Arab speaking world – University of Jordan initiatives

Dina Tbaishat

Library and Information Science, University of Jordan, Amman, Jordan

Abstract

Purpose – This paper aims to examine some of the digitization projects at the library of the University of Jordan. As the library acts as the deposit centre for theses from the Arab world universities, an effective management policy for deposit, management, and access is required. The paper illustrates some of the existing workflows and considers the challenges and possible improvements.

Design/methodology/approach – Interviews (n = 8) with library staff in the Information Division provided details about the methods used for the collection and management of print and electronic theses in the University of Jordan, in addition to some digitization projects. In addition, very recent unpublished brochure was collected from one member of staff in the computer applications and databases section to learn about theses deposit statistics from various countries in the Arab world. A business process modelling technique (Riva) (Ould, 2005) using role activity diagram illustrated some of the existing workflows, the challenges and possible improvements.

Findings – Some Arab universities do not send theses on regular basis; some do not collaborate at all. Workflows vary according to the format of the thesis. The number of digitized print theses has reached 77,885 titles, out of total of 84,043, but it is a slow process. All newly submitted theses from Jordanian universities come in print and on a CD, but other external institutions may use different methods (CD only by mail, or direct uploading to the webpage provided by the University of Jordan). The current online system for uploading theses was last updated in 2017. Deposit rates through the system are very low. Access to theses is relatively limited; print theses can be accessed within the library premises. The online theses repository provides full text and all academic staff, students and researchers can access these from any workstation within the university campus. Only registered users may access theses databases (read only) outside the network. Resourcing for digitization is limited, and the digitization itself is currently outsourced although other processes are done in-house.

Practical implications – The paper encourages the use of a practical online theses’ repository to deposit theses from the higher education within the Arab world.

Social implications – Establishing theses repository and ensuring proper and smooth deposit process by Arab universities libraries, would enhance communication and collaboration amongst them.

Originality/value – The paper supports the theme of the primacy of global access to information for learning and scholarship. The paper adds value in the context of theses access amongst Arab universities. As the library acts as the deposit centre for theses from the Arab world universities, an effective management policy for deposit, management, access and preservation is required. The paper provides a set of recommendations that would contribute to raise the visibility of research produced by Arab higher education, through a single source of access to theses.

Keywords Academic libraries, Digitization, Theses, Digital collection preservation, Theses repositories, University of Jordan

Paper type Research paper

1. Introduction

The past two decades have witnessed great growth in the number of information resources. As part of keeping up with the rapid changes in the information resources environment,
many academic libraries around the globe are adopting digitization projects to provide better and wider access of information resources. Besides the “born digital material”, they have been creating digital sources from the original non-digital material. Digitization is a hot topic. Academic and research libraries have become increasingly interested in using electronic information and resources on the internet. Hughes (2004) defines digitization as “the conversion of an analogue artefact into a binary representation”. Digitization has become a big business nowadays. Most museums, large libraries, and organizations are currently at some stage of making their materials available in the digital form (Hughes, 2004). This indeed has also arisen the need to preserve digital objects to strengthen the trend of the digital library environment and to be able to survive in it.

According to Sutherland (2008), digitization can be at different levels. It can start with capturing images of book pages using scanner or digital camera, which is similar to storing books on microfilm except that the digital form allows more functionality. Alternatively, it can be improved to include optical character recognition, which is used by Google to search easily through page images. Finally, there is an advanced phase of digitization called semantic coding, which can identify whether a particular instance of a word, say for example “Petra” refers to a person or a city. There are various courses available online to explain detailed steps of digitization (Cornell University Library, 2003; Canadian Council of Archives, 2002). Digitization can be done in-house or can adopt outsourcing.

Dissertations and theses play a major role in the educational experience. They represent research conducted under supervision of academic staff. However, much of this research has been treated as “grey literature” and has been hard to handle for easier access. Many institutions now believe it is essential to make this research available to other scholars. Handling theses is a very important process, as theses are now getting into universities’ digital repositories.

Digitization and thesis handling overlap at University of Jordan library, as thesis are now being digitized. The aim of this paper is to examine some of the digitization projects at the library of the University of Jordan. In addition, the paper investigates the process of theses handling, as the library acts as the deposit centre for theses from the Arab world universities. To understand how the library handles theses, the paper illustrates existing workflows using role activity diagram (RAD), and considers the challenges and possible improvements.

2. Electronic theses and dissertations initiatives

In many universities, it is now mandatory for students to submit an electronic copy of their thesis. Haddow’s (2008) study concludes that mandating academic staff to deposit is the only way to get stuff into a repository. Regardless that some disciplines are better at self-archiving than others, and some disciplines run subject-based repositories. Digital repositories are important as they provide a digital archival record of research outputs carried out by the institution.

There has been a trend towards electronic theses and dissertations (ETDs) internationally. Baro and Otiode (2014) investigated the extent of ETDs adoption in university libraries in Africa. All 56 universities under study have already adopted ETDs in their institutional repositories. D-Space was found the most widely used software to manage electronic theses.

A similar study was conducted in eight government-funded universities in South Eastern Nigeria. Ezema and Ugwu (2013) found that only three out of the eight universities initiated ETDs projects. The three universities revealed benefits, including better global visibility and enhancement of scholarly communication in Nigeria. In 2005, the Indian National
Theses Database was created to allow mandatory online submission of metadata sets of a PhD thesis (Khaparde and Ambedkar, 2014). The Canadian Association of Research Libraries (CARL) has been promoting the use of institutional repositories in Canada since 2003. This was achieved through the CARL institutional repositories programme (Khaparde and Ambedkar, 2014).

In the UK universities on the other hand, much attention is given towards accessing theses online. WRAP (Warwick Research Archival Portal, 2011) is an example of the university of Warwick institutional repository which is part of the growing movement towards open electronic access. Academic libraries in the UK are increasingly becoming interested in moving the workflow and management of theses into the digital realm. JISC supports academic institutions in the UK by providing innovative programmes to bring about original and effective solutions, aiming to fully demonstrate and exploit potential of information and communication technologies. This would in turn provide access to highly quality research sources (Jacobs, Thomas and McGregor, 2008). JISC launched the Digital Repositories programme in 2005 and completed it in 2007. The programme embraced three e-theses projects as listed in JISC (2010), and these are:

- EThOS (Electronic Theses Online Services) project, co-supported by the British Library and the Consortium of University Research Libraries (CURL).
- Evaluation of options for UK electronic thesis service, an independent evaluation of the EThOS prototype service.
- Repository Bridge – automated linkage of national and institutional repositories. The project examined the interaction between a regional theses repository based at the National Library of Wales, and pilot institutional repositories (Aberystwyth and Swansea universities).

The White Rose Consortium is another example on partnership between three Yorkshire’s leading research universities. It offers White Rose E-theses Online to provide doctoral theses awarded by the White Rose universities (Sheffield, York and Leeds) (White Rose University Consortium, 2009).

ASLIB (Association of Special Libraries and Information Bureau) Index to Theses is another tool of searching online for theses accepted for higher degrees by the universities of Great Britain and Ireland (Index to Theses, 2011). Subscription is necessary to be able to view theses, and it is granted through the IP address.

3. Digital content preservation

Digital preservation is accorded high priority in many libraries. Digitization is not preservation. It offers instead a new preservation paradigm. Although digitization is being carried out at many academic institutions, many of them do not consider how long the digital content should remain available or what procedures should be taken to update content. Therefore, IFLA (2002) recommended some preservation strategies to ensure long-term maintenance of data and access of resources within the rapid changing environment, these are:

- Associate preservation and access as organizational objectives;
- Set digital preservation policies before you begin scanning;
- Adhere to international standards and adopt current best practice;
- Avoid dependence on proprietary software;
- Assign administrative, descriptive, structural and preservation metadata to all;
• digital objects; and
• Identify a trusted digital repository committed to the long-term management of your digital resources.

JISC appreciated the need for digital preservation. As part of JISC digitization programme, JISC looked at the digital preservation element of some projects undertaken. The findings revealed that external examination is necessary. In addition, the following issues were recommended to avoid putting the long-term usability, authenticity, discoverability and accessibility of the digital collection at risk (JISC, 2009):

• providing a well-written preservation policy;
• defining a collection and content management procedures; and
• maintaining digital collections on a suitable digital preservation infrastructure.

For managing digital collections, DuraSpace collaborated with the Bishoff Group to learn about the status of digital content management and preservation in the non-ARL (Association of Research Libraries) academic library community. The survey results revealed that 81 per cent of the 145 respondents are currently acquiring digital content. In all, 72 per cent have institutional repositories that are either locally managed or hosted. While 66 per cent noted that they do have some commitment to digital content preservation. Data backup won ground as the most used digital preservation policy. Followed by selecting outsourcing to an externally managed preservation repository (Bishoff and Smith, 2015).

4. University of Jordan library experience

4.1 Data collection methods

Data were collected through \( n = 8 \) interviews with library staff involved in thesis handling and digitization processes. At university of Jordan library, the Information Division comprises three sections: The Computer Applications and Databases section, the Theses section and the Archives and Microfilm section. Table I illustrates the number of interviews conducted to collect data about the processes.

Permission was gained from the library’s director to conduct semi-structured interviews with members of staff involved in the digitization and theses handling processes, to learn about the detailed activities (from members of staff) and policies (mainly from sections’ heads). The schedule of the interviews was smoothly arranged and the interviewees were

<table>
<thead>
<tr>
<th>Information division departments</th>
<th>No. of staff</th>
<th>No. of staff interviewed</th>
<th>Roles</th>
</tr>
</thead>
</table>
| Computer applications and databases section | 9 | 2 | Two members of staff interviewed. Those involved in the digitization process — they upload theses onto the system.
| Theses section | 3 | 3 | All of them interviewed. They receive theses and they also scan print theses (prepare them to be digitized).
| Archives and microfilm section | 3 | 3 | All of them interviewed, including the section head. They are responsible for digitization projects, archiving and preservation. |

Table I. Staff interviewed at the information division to collect data about theses handling and digitization processes.
highly cooperative. A list of topics was prepared (as a guide). The interviewees had great way of leeway specially that the researcher comes from computer science background, not librarianship. Therefore, a lot of information and insights were proposed in some matters. The list of topics and related questions covered are:

- **Stakeholders involved** – who does what, and the number of people involved in the digitization and theses handling processes.
- **Digitization projects** – How many projects proposed? How many were actually carried out? What sort of materials the library digitize/wish to digitize? Do they do digitization in-house or it is usually outsourced?
- **Theses handling process** – What are the detailed activities? What are the roles? What are the challenges? How does local students’ submission to theses differ from the process of theses submission from Arab world universities?
- **Preservation** – What sort of archiving/preservation policy do you have for digital content?
- Statistical information regarding theses submissions from Arab universities.

In addition to interviews, very recent unpublished brochure was collected from one member of staff in the computer applications and databases section, to learn about theses deposit statistics from various countries in the Arab world (Figure 3).

### 4.2 Digitization at University of Jordan library

The Archives and Microfilm Department at University of Jordan Library is responsible of digitization projects. It comprises three members of staff including the head of department. According to the head of department, they have been using microfilms and microfiches as a way of preserving old materials. Examples of such materials are manuscripts (there are around 2,500 titles in various disciplines); newspapers and magazines, which go back to the nineteenth century; documents of legitimate spiritual courts aged five hundred years back from the Ottoman period; Parliament’s documents since 1926 (very important source for interested researcher as such documents hold the country’s political and social story); and, finally, rare books and municipality records.

Many digitization projects were initiated at University of Jordan library. However, some of these projects were halted for lack of fund. The project that was successfully completed was digitizing Arabic and foreign bounded newspapers collections. Those newspapers go back from 1908, until the 1960s. They were available in print, and used to occupy large space. The project started in 2009 and was successfully completed in 2012. The bid division hired an outsourcing company to digitize the material within the library premises. The material was archived digitally, making 204 newspaper titles, that is 5.5 million pages.

Regarding the printed material, the library decided to get rid of it. The project was supervised by University of Jordan library and was continuously followed up in 2013 and 2014 to ensure it was implemented in the right way. Although University of Jordan library has been assigning digitization to an outsourcing company, the library still supervises the work and performs all tasks related to cataloguing, classification, titles’ separation and auditing. In addition, it is worth mentioning that the above material is not available online, it can only be accessed from the PCs in two halls in the library, for copyright issues.

The library has been trying to move from microfilms and microfiches to digitization. The library tendered hiring an outsourcing company to digitize all microfilm material. This
collection includes new newspapers from 1972 onwards, and also manuscripts and other documents that go back from 1870 to 1950. Some of these documents are available in print format, while the rest is preserved using microfilms. It was estimated that this project will take five years. The library is negotiating with one specific governmental institution in Jordan to set the deal. Through this collaboration with the second party, the library plans to reach an agreement that will result in obtaining the digitization devices once the project ends. In hope to continue, digitize other material in-house, such as old print books, old periodicals and special collections.

In attempt to making progress towards redefined workflow, an interesting question would be; how is the library dealing with new copies of newspapers? Most of them are still deposited in print! There is no collaboration from the newspapers industry regarding uploading electronic copies, except for one newspaper, that has started depositing its copies electronically, since 2017.

4.3 Theses handling at university of Jordan library
University of Jordan library has been the Theses Deposit Centre since 1986. The library embraces large number of submitted theses from local students and from member universities in the Association of Arab Universities, also referred to as Union of Arab Universities. The aim of this organization is to support and connect universities in the Arab World, and enhance collaboration among them. The association is based in Amman-Jordan.

To understand how University of Jordan library has been dealing with the process of theses deposit, a graphical tool called RAD is used. RAD is part of the Riva method proposed by Ould (2005), created to illustrate processes, their activities and the roles undertaking those activities. The next two figures illustrate the process of handling theses at University of Jordan library. Figure 1 deals with University of Jordan students submitting their theses, while Figure 2 deals with theses submitted by the Arab World universities (including other Jordanian universities).

RADs have been used in the literature as a modelling tool to visualize and improve processes in various contexts. Dawkins (1998) applied RAD for safety process definition. RAD was used to investigate decision and communication processes, in attempt to link strategic decision making with information system provision in an insurance company (Beeson et al., 2002). Cox and Phalp (2003) used RAD to derive problem frames in a business process. Khan et al. (2006) investigated the scientific publishing process for digital libraries to bridge the gap between business models (presented by RAD) and grid computing. Tbaishat (2010) illustrated periodical acquisitions process and derived possible improvements through RAD. Shukla et al. (2014) used the RAD to model magnetic resonance (MR) scanning service-delivery process in the radiology department of a large hospital. Suggested guidelines are presented to visually analyze the resulting model for identifying process vulnerabilities.

It is important to model thesis handling process at University of Jordan library to understand the current workflow. In addition, the library is the thesis deposit centre from Arab world universities. Therefore, attention should be given to enhance the process and ensure accessibility and content preservation. Lakos and Phipps (2004) stress that unexamined workflow is a major challenge to overcome. Evaluation of processes in academic libraries for efficiency may exist, but there is little published in the peer-reviewed literature (Urquhart and Tbaishat, 2016). RAD proved to be a feasible method used to model processes (as shown earlier in other work in the literature), for this paper, it is found that:
• RAD comprises a rich set of notations to specify detailed workflow (Appendix).
• Riva is quick and easy to learn, it is relatively straightforward to enact, as stated by Green and Ould (2004) and claimed by Khan et al. (2006). Once it is learnt, one can draw diagrams quickly and easily, the tool used adopts the drag and drop facility.
• Not only does Riva save time but also saves paper work. It reduces the paper work needed to describe processes when written in natural language. Diagrammatic tools act as a substitute for long and too wordy documents. They are considered systemized way of presenting and documenting processes. The resulted RADs were reviewed with one member of staff for process validation, and it was easy to comprehend and follow.
• RAD acts as a searchlight to uncover problems and reveal bottlenecks, allowing possible improvements. It was mentioned earlier that Theses section comprises three members of staff. From Figure 1; it is clear that they are responsible for most of the activities (receiving theses and also putting them back on shelves). This reflects workload. It was also learnt form the interviews that acquisitions used to
be responsible for receiving theses and the computer applications and databases section used to be responsible for scanning print thesis. However, assigning roles to functions has changed, so that both functions are now the responsibility of Theses section. Ould (2005) suggests many types of improvements through RAD; one of them is called “restructuring roles, which recommends investigating the tasks performed by each role and then moving tasks between roles accordingly. In this case, some activities can be assigned to other roles to reduce Theses section burden.

The process models also lack any digital content preservation activities. The performance of academic libraries was frequently assessed in terms of service quality expectations. The emphasis is on outcomes rather than internal processes. Academic libraries are in a period of great change, and the advantage of a tool such as RIVA is that traditional functions can be re-assessed.

To improve the process and to keep up with the rapid technological developments, University of Jordan library launched a system for electronic these deposit in 2013. However, due to some deficiency in the system, Arab World university libraries continued sending their theses on CDs by post. In 2017, the system was updated and
therefore improved, and the libraries were encouraged again to submit their theses online. Unfortunately, electronic deposit rate via the system is still very weak. According to the computer applications and databases section at University of Jordan library, only one university in Lebanon actually uploads their theses online through the system.

The system allows uploading a pdf and word document of theses. The university of Jordan library staff then enter the bibliographic records of theses (supervisor name, year, university, etc.) into the repository. In some cases, where the classification number is not yet assigned, theses still get uploaded and can be viewed via the repository.

In regard to theses access, users can access print theses only in the reading hall but they cannot borrow them. There is also the online theses repository which provides full text for theses content online. All academic staff, students and researchers can access theses from any workstation within the university campus; however, only registered users have access to theses databases (read only) outside the network. In addition, all members of the Union of Arab Universities can access the repository as read only, except for their theses which can be uploaded.

4.4 Preservation at University of Jordan
University of Jordan library follows some procedures in attempt to preserve and maintain their digital collections; they refer to it as “post digital archiving development”. These procedures include:

- continuous auditing of the work performed and acting if something goes wrong; for instance, correcting any mistakes noted by researchers, or separating some different resources that are entered as one;
- various enhancement on searching and browsing facilities; and
- providing backup copies of the digitized material; two DVDs copies, one saved in the archives department in the library, and another stays with the computer centre.

4.5 Results
University of Jordan library follows a certain process for handling theses. The process is illustrated using RAD in Figures 1 and 2. Not only do the models provide understanding of the activities and roles involved in the process, but also act as a searchlight for possible improvements. The diagram denotes workload on members of staff in the Theses section. The diagram also lacks any digital content preservation activities or adherence to preservation policy. To sum up the results found through the interviews with the involved library staff, the current situation can be described as follows:

- The total number of theses (in all formats print and archived electronically) reached 84,043 titles in 2018.
- The computer applications and databases section used to be responsible for scanning theses (to digitize back copies). Recently, this job has been assigned to theses section. Archiving old print theses electronically has started few years ago and still in process. Bounded theses are cut, scanned, then catalogued and classified. Around 77,885 titles were archived electronically, out of total of 84,043.
Since 2012, students at the University of Jordan are mandated to deposit an electronic copy of their theses along with the print copy.

Other universities in the Arab World still send their theses on CDs, although a validated electronic system for theses deposit was launched properly in 2017. Only one University library in Lebanon, along with all public and private universities in Jordan, upload their theses directly via the system. Those who send CDs, either provide all theses on one CD, or send a CD for each title.

Total number of employees dealing with theses at the University of Jordan library is eight. Three receive theses, two classify them, two in binding, and one in the theses hall.

Regardless the way theses are deposited (electronically or by post), three Syrian universities (mainly Damascus university library) win ground after Jordanian universities, where the number of deposited theses reached 14,770 from Syria only. However, Universities in Syria have stopped depositing their theses since 2013, due to the war and the current political situations. Palestine comes next, with 2659 theses. Moreover, there is 1981 theses submitted from Egypt, but mostly considered old collection as Egyptian universities have also stopped depositing their theses. The same applies on Iraq. The political situation in the Middle East played a negative role in Arab World university libraries participation. It was learnt that at the moment, the committed university libraries that deposit there theses regularly are public and private universities in Jordan, AlNajah University and University of Jerusalem in Palestine and the Islamic University in Lebanon. The next Figure 3 demonstrates the total number of theses for top six countries.

4.6 Challenges associated
University of Jordan library faces some challenges associated with digitization projects; these are mainly tied to tight budgets as stated by the archives and microfilm department in the library. The library sparked ideas on digitization but none of the projects suggested was implemented, due to lack of funding. One example of these

![Figure 3. Number of theses deposited by top six Arab countries](chart.png)
projects is converting all microfilm material into digitized collections, as mentioned in Section 4.2.

It is not a surprise that the library faced some difficulties while digitizing material, as making decision about digitization projects is costly, and requires a detailed plan to ensure continuous use of material. IFLA (2002) states that institutions in the developing countries in specific should consider whether the costs and time involved will be commensurate with the benefits. According to IFLA (2002), the greatest challenge lies in migration of data formats and devices. Storage devices change over time which might make them no longer accessible on emerging new platforms, a clear example at university of Jordan library is the tendency of moving towards digitization as a substitute of microfilms, and moving from submitting theses on CDs to direct deposit online.

There are also some challenges associated with theses handling at University of Jordan library. The biggest challenge is the Arab universities’ commitment towards sending their theses. Some universities do not send theses on regular basis; most of them do not send at all; others have been going through difficult political conditions that affected their collaboration with the university. Arab universities are frequently reminded in the regular meetings for the higher education institutions, arranged by the Association of Arab Universities; still, the collaboration is not as it should be.

4.7 Recommendations

As University of Jordan library is the Theses Deposit Centre for all Arab World universities, it is highly recommended that the library takes strict action regarding how theses are deposited, managed and accessed. As learnt from the head of archives and microfilm department, theses are still deposited by most Arab World university libraries by regular mail, in a form of a CD. The online theses deposit system (launched in 2017) can solve this problem and save much time and effort, inspired by the UK experience of EThOS project mentioned earlier in section 2. According to Russell (2009), EThOS – which was implemented in 2009 – has helped to raise the visibility of UK research since it has been providing a service to the research community to access UK theses. It offers a single source of access where researchers can access most theses produced by UK higher education, and it also supports the movement towards e-theses. The same can be achieved for the Arab World University libraries through the online theses repository, if only they can be encouraged to make deposit online. This is not an easy task! Implementing an institutional repository is not easy, as it is difficult to convince academics to deposit their research work into repositories. Ferreira et al. (2008) testify this:

[...] the task of convincing researchers to deposit their publications in the institutional repository is, by far, a repository’s manager most demanding task. A great deal of research and imagination are needed to attempt to counter the initial reluctance of researchers to begin depositing their research materials in the institutional repository.

To tackle the problems of low deposit rates, the authors devised a strategy – at the University of Minho/Portugal – composed of four components, these are developing a promotional plan, developing value-added services for the authors such as providing user guides, and useful statistics (e.g. how many times it has been downloaded), the third component concerns further involvement in the international repositories (e.g. allow comments and recommendations) and, finally, mandating academics to deposit their research with offerings of financial incentives to promote deposit.
As for University of Jordan library, it is essential that other universities in the Arab World are continuously reminded about theses deposit. It was mentioned earlier that Arab World Universities are frequently reminded in the regular meetings for the higher education institutions, arranged by the Association of Arab Universities; still, the collaboration is not as it should be. Therefore, it is important to stress this point and exploit other means of communications such as social media.

For University of Jordan theses, it is recommended that students upload their own theses online. This will reduce the workload on the library’s staff. It is worth mentioning that the library is considering cancelling receiving theses in print from university of Jordan students.

The third recommendation is providing a well-written preservation policy that ensures persistent archive management. One of the challenges suggested by IFLA (2002) is that the concept of long-term access is not supported by the IT industry; this implies that continuous technical support is needed, along with a frequent updated preservation policy. University of Jordan library has some sort of a similar policy called post archive development, which was referred to in Section 4.4. However, more attention to a detailed well-written preservation policy should be given.

Finally, some improvements can be highlighted through the RAD provided. Ould (2005) suggests that some improvements can be achieved by restructuring roles. There is workload on the theses section role where only three members of staff deal with the large number of theses submitted (receiving and scanning).

5. Conclusions
Managing research within universities is essential especially in such highly competitive and collaborative environment. This paper presents a selection of digitization projects in the literature, along with some measures taken to protect the investment in these projects – focussing on UK institutions’ experience. The paper then investigates the University of Jordan library initiatives in digitization. It also examines workflows of theses handling process. Analyzing theses handling process at University of Jordan is essential, as their library is considered the Theses Deposit Center from Arab World Universities. Therefore, attention should be given to enhance the process and ensure accessibility and content preservation. Data was collected through interviews (n = 8) with all library staff involved in digitization and thesis handling. RAD was found easy to enact for illustrating the workflow of theses handling process. It was also used to searchlight some possible improvements. It is highly recommended that the library takes strict action regarding how theses are deposited, managed and accessed. It is also essential to maintain a well-written and updated digital collection preservation policy. Digitization challenges are inevitable; therefore, providing an optimum solution is a hard task; however, some strategies can meet some of the challenges.

References


Haddow, G. (2008), “Self-archiving to institutional repositories is improved by assisted and mandated deposit; disciplinary culture is not a factor”, Evidence Based Library and Information Practice, Vol. 3 No. 2, pp. 52-57, available at: https://doi.org/10.18438/B8SK5N


<table>
<thead>
<tr>
<th>Figure A1.</th>
</tr>
</thead>
<tbody>
<tr>
<td>RAD notations, from Visio; the modelling tool used in this research</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>A role. Example: acquisitions team, or library director</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>An action. Example: prepare presentation</td>
</tr>
<tr>
<td></td>
<td>Start another role. Example: start new designer</td>
</tr>
<tr>
<td></td>
<td>Driver part-interaction. Example: order book</td>
</tr>
<tr>
<td></td>
<td>An interaction.</td>
</tr>
<tr>
<td></td>
<td>Interaction between two roles. Example: negotiate price</td>
</tr>
<tr>
<td></td>
<td>A state description. Example: order complete</td>
</tr>
<tr>
<td></td>
<td>A state.</td>
</tr>
<tr>
<td></td>
<td>A trigger. Example: start of academic year</td>
</tr>
<tr>
<td></td>
<td>Case refinement: alternative paths depending on the condition. Example: budget exceeds 2000? Ye/No</td>
</tr>
<tr>
<td></td>
<td>Part refinement: concurrent paths (to represent things done in parallel).</td>
</tr>
<tr>
<td></td>
<td>Replication. Example: for each book</td>
</tr>
<tr>
<td></td>
<td>Don’t care what happens then.</td>
</tr>
</tbody>
</table>

**Source:** Microsoft Office (2010)
Emerald is excited to announce a recent partnership with Peerwith, a platform that provides authors with a variety of services.

The Emerald Peerwith site can be found here: https://authorservices.emeraldpublishing.com/

Peerwith connects academics seeking support for their work with a relevant expert to get their research submission-ready. Peerwith experts can help with the following: language editing, copy editing, scientific editing, translation services, statistical support, funding application support, visuals, video, publication support, literature search, peer review and indexing services. Authors post their assignments on the Peerwith site, experts provide a quote, and the fee and conditions are then agreed upon directly between the author and the expert.

While we are not, of course, guaranteeing publication upon use of Peerwith, we hope that being able to direct academics to this resource either before submission or during the peer review process will help authors further improve the quality of their papers and increase their chances of positive reviews and acceptance.

Academics with relevant expertise can sign up as an expert on the Peerwith system here: https://www.peerwith.com/services/offer
eBook Collections

Research at your fingertips

Bringing together over 1,600 eBooks, the Emerald eBook collections are a cost-effective way of instantly expanding library holdings and increasing usage through an award winning platform alongside journals.

Visit emeraldinsight.com

Get eBook Collections for your library

Recommend eBooks to your librarian today. Find out more:
emeraldpublishing.com/ebookcollections
Digital Library Perspectives

Number 4

245 Editorial advisory board

246 Distributed digital preservation: preserving open journal systems content in the PKP PN
   Bronwen Sprout and Mark Jordan

262 Linking historical collections in an event-based ontology
   Qing Zou and Eun G. Park

276 Master’s theses and open scholarship: a case study
   Joachim Schopfel, Sylvain Vanacker, Eric Kergosien and Bernard Jacquemin

288 Awareness, knowledge and attitude of lecturers towards institutional repositories in university libraries in Nigeria
   Francisca Nwakaego Okoroma

308 Ensuring access to theses in the Arab speaking world – University of Jordan initiatives
   Dina Tbaishat

www.emeraldinsight.com/loi/dlp