THE IMPORTANCE OF REPOSITORIES IN SUPPORTING THE LEARNING LIFECYCLE

Joanna Richardson, Malcolm Wolski

Griffith University (AUSTRALIA)
j.richardson@griffith.edu.au, m.wolski@griffith.edu.au

Abstract

Traditionally repositories have been seen as either supporting a researcher's activities throughout what is traditionally referred to as the "research lifecycle" or storing learning objects to support learning and teaching. In research and e-science workflows, data is analysed with one or more of the outputs typically resulting in a peer-reviewed publication, e.g. journal article. Depending upon the licensing conditions of the publisher, a version of that same publication may be uploaded to an institutional repository for open (public) access.

In addition a lecturer in a course may choose to utilise this same publication as a required or recommended reading normally by either linking to the journal website, uploading a copy of the published version or leaving the students to source a copy. Frequently it does not occur to the lecturer to investigate whether a version is already held in their institutional repository.

This paper examines the current literature on the topic. It discusses a number of initiatives, particularly in Australia, which incorporate repositories within the student lifecycle. The paper concludes with an analysis of how repositories could be better utilised to support the needs of students in higher education.

Keywords: repositories, digital repository, student lifecycle.

1 INTRODUCTION

The focus worldwide on generating knowledge and innovation to drive economic and social progress has created challenges for national higher education systems. Government funding and policy guidelines are placing pressure on universities to increase their research impact. An important element in the equation is improving the quality of graduate students, especially those who may progress to become early career researchers.

At the same time in many countries there is a push to have publicly funded research available via open (public) access. It is open access (OA), which has driven the creation of repositories, especially institutional repositories. There is also an impetus to make learning objects more accessible, shareable and reusable. Historically learning management systems have either provided their own internal storage capacity or have offered a third-party solution for storing learning objects. However neither solution has been particularly effective in promoting discoverability and reuse.

This paper provides an analysis of ways in which repositories can add value to the learning / student lifecycle, with particular reference to content which has been created as part of the research lifecycle.

2 LITERATURE REVIEW

The Joint Information Systems Committee [1] makes the distinction between digital, institutional and open access repositories. Conventionally a [digital] repository is any collection of resources that are accessible via a network without prior knowledge of the structure of the collection [2].

Crow [3] defines institutional repositories as "digital collections capturing and preserving the intellectual output of a single or multi-university community". In a recent report [4] on "Redefining the Academic Library", the authors note that, as part of rethinking the scholarly publishing model, a number of US colleges and universities are investing in a range of open access initiatives. These include disciplinary repositories, which are hosted by libraries but run by faculty and scholarly societies, and institutional repositories.

Xia and Opperman [5] have examined the use of repositories to support what could be considered the prequel to the official research lifecycle: student outputs. They surveyed non-doctoral (master's and

baccalaureate) institutions and reported that students contribute nearly half of all repository content in the form of theses and dissertations, honours papers, and student journals.

In the learning lifecycle environment, Lehrmann et al [6] have examined the specific issues associated with learning resources, particularly those delivered through a learning management system (LMS). They discuss the specific instance of the adaptation or re-authoring of the learning resource as part of its re-use in a different context, and the subsequent consideration of that re-authoring to be the creation of a new object. The authors suggest that it is important to capture "lifecycle information" that is generated as part of the re-authoring which could enrich a user's understanding of both versions, i.e. the original and the re-authored. This highlights the need for an information architecture which does not store content within the LMS.

It is really the Content Lifecycle Integration Framework (CLIF) Project [7] which has sought to embed repositories in the content lifecycle and "prevent them from becoming yet another content silo within the institution". It has looked at how two UK institutions proposed to deal with the interaction of the authoring, collaboration and delivery of materials using three systems: the Fedora Commons repository software, Microsoft SharePoint and the virtual learning environment, Sakai.

While the CLIF Project is very useful for those institutions who wish to integrate –from a technical perspective– the entire digital content lifecycle, the authors of this paper will focus instead on less tightly coupled approaches.

3 IMPORTANCE OF REPOSITORIES IN THE RESEARCH LIFECYCLE

Institutional repositories (IRs) developed from the practice of academics and researchers of uploading their publications to their staff / departmental website or disciplinary repositories [3]. Librarians—as experts in information management—realised that the convergence of new digital publishing technologies, increased use of metadata, development of interoperability protocols, and the focus on global networking could provide better solutions to increasing access to institutional scholarly outputs. Institutional repositories were promoted as having a dual role: (1) capturing and preserving an institution's collective intellectual output and (2) enhancing scholarly communication by providing greater access to that output.

IRs are now part of the scholarly communication landscape, which formerly described the process of publishing a research monograph or an article in an academic journal but which has been subsequently expanded to include issues such as preservation, copyright, intellectual property, and open access (OA). The latter concept aims to promote access to scholarly material that is open to all and free of technological and economic restraints. The OA objective is to encourage the dissemination of knowledge broadly and freely across the internet in a timely fashion. It is underpinned by the precept that taxpayers should not have to pay twice: first to fund the research itself and then again to gain access to the research results. Providing open access to information through the internet enables academia from disadvantaged countries to use scholarly research that they otherwise could not access because of the cost of journal subscriptions. Granting equal learning opportunities to these researchers and academics encourages social unity and cultural advancement. It also increases the possibility of vital discoveries that could have an important impact in the global community [8].

Adding to these motivating factors is that funding bodies and governments are now seeking an improved return on investment for funded research. In several Commonwealth countries, such as the United Kingdom, New Zealand and Australia, accountability is measured among universities by means of a research assessment exercise. Institutional repositories are integral to collecting and making accessible research outputs for these types of government requirements.

Funding agencies are also promoting freely available, publicly-funded research findings. Mandates by funding bodies such as the National Institutes of Health (NIH), the Medical Research Council, the Wellcome Trust, the National Science Foundation and—in Australia—the Australian Research Council (ARC) and the National Health and Medical Research Council (NHMRC) demonstrate the recent change in funding rules based on new research paradigms. In addition many major research funders worldwide either currently have or are implementing policies that require grant holders to submit data management plans for formal approval and to manage their data in accordance with those plans. The National Science Foundation [9], for example, has mandated that data management plans will be subject to peer review.

Scholarly information services providers within the university context are currently exploring ways in which they can assist in the data collection stage, for example, as well as a much more extensive interpretation of "research outputs". In a paper presented at EDUCAUSE Australasia, Wolski and Richardson [10] proposed the following diagram (Fig. 1) to illustrate the shift in focus from traditionally published works to new publishing paradigms.

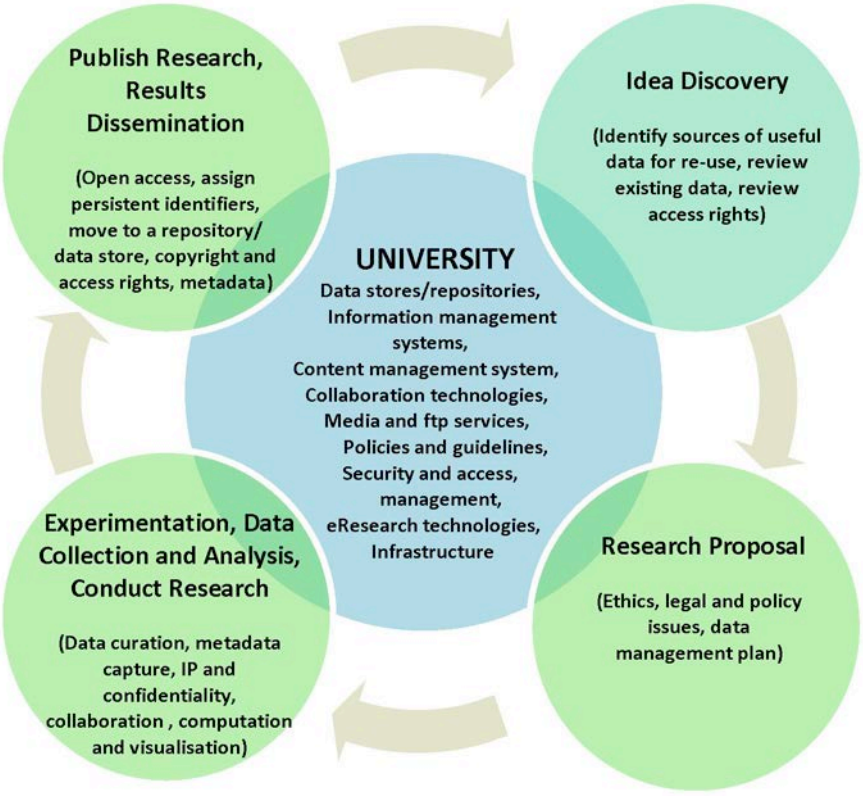

Fig. 1: Principal Stages of the Research Lifecycle

Until recently the scholarly output that institutions have focused on capturing has tended to be limited to traditionally published works. More recently new publishing paradigms are emerging, with data—supporting journal articles—as the focus. In reviewing major data management lifecycle models, Ball [11] mentions the role of repositories at several stages, e.g. preservation. The necessity of obtaining a Digital Object Identifier (DOI) prior to publishing implies data needs to be in a managed storage system such as a repository. Repositories will continue to have an important role to play in supporting the research lifecycle as that support now moves to encompass more than just print research output [12].

4 ROLE OF REPOSITORIES IN THE LEARNING LIFECYCLE

In the following diagram (Fig. 2), Lyon [13] describes how she envisages the integration of research into the learning lifecycle. As research is undertaken and results are written up in publications, the latter are self-deposited by their authors into the institutional repository. This is then linked to by the learning management system (LMS). In addition to the IR, the LMS also links to content in other organisational repositories: data generated by research as well as traditional learning objects. In this model repositories are critical to learning and teaching workflows.

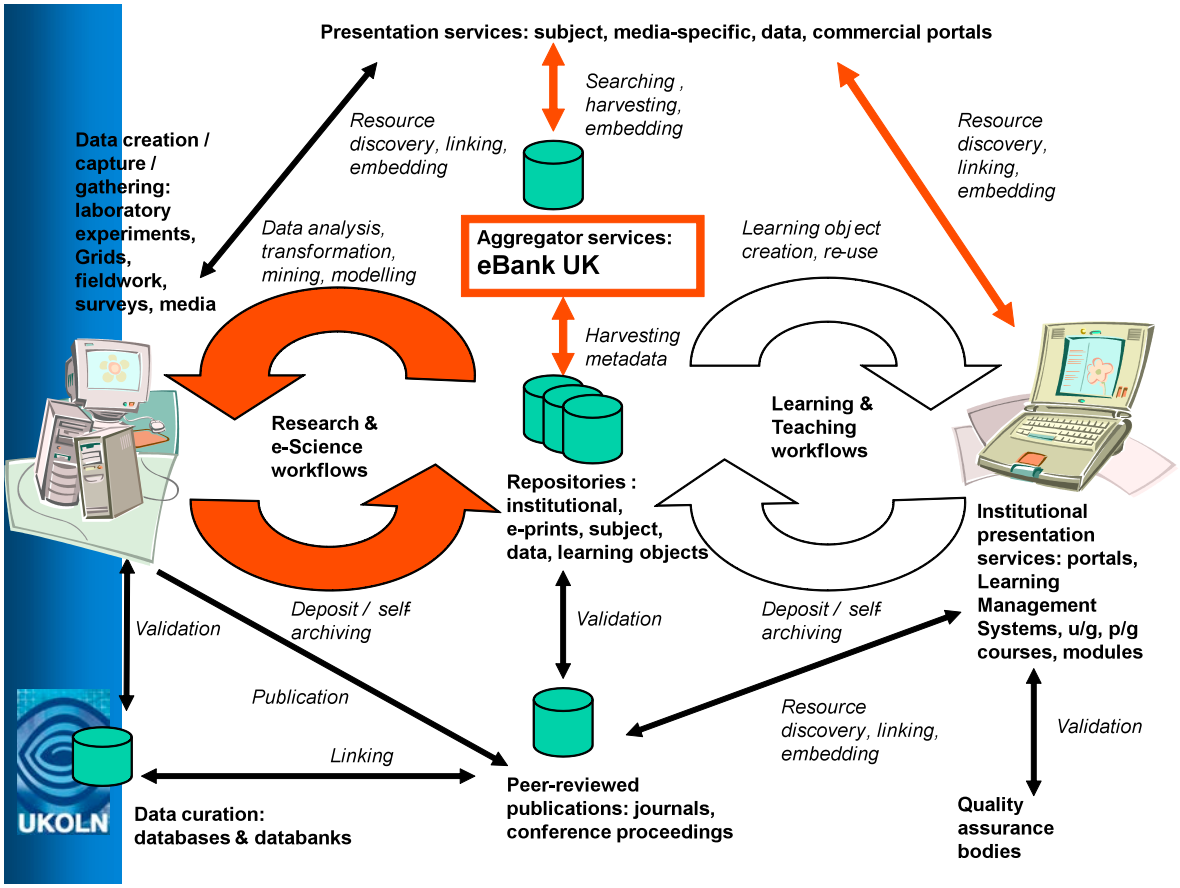

Fig. 2: Integration of Research Content in the Learning Lifecycle

In current practice the author of a research publication may upload a version of that same publication to an institutional repository for open (public) access, depending upon the licensing conditions of the publisher. A lecturer in a course may then choose to utilise this same publication as a required or recommended reading normally by either linking to the journal website, uploading a copy of the published version or leaving the students to source a copy. Frequently it does not occur to the lecturer to investigate whether a version is already held in their institutional repository.

Publications are not the only research content utilised in coursework. The importance of research data is not only now coming to the fore in the publishing world but also in the learning environment. Scholarship has become data intensive [14] and that data is as important as the resultant publication in the learning lifecycle.

Microscopes, for example, have changed from interactive tools into instruments that capture digital microscopy images that, by their nature, generate large data storage needs. The next generation of instruments, already being deployed in Australian research institutes, have the potential to generate Terabytes of data on a daily basis. This enormous volume of data must be stored, indexed, managed, manipulated and analysed. Researchers at Griffith University are investigating a complete solution for the enterprise-wide management of microscopy imaging, high content screening and medical imaging data to meet their research needs, including image analysis capability and research data archival. The proposed solution provides an institutional repository for microscopy image publishing ensure that not only current best-practice and future mandated supplementary materials publishing requirements are met but also that the visibility and profile of Griffith imaging expertise and the advanced tools are available to both researchers and students.

Research in the creative and performing arts also generates non-text-based outputs. Increased multimedia content, for example, requires streaming, increased storage, and new types of metadata. Multimedia content demands the capture of more extensive descriptive metadata than that traditionally used to describe a book or journal article. Metadata for a film, for example, may include a plot summary, cast and crew, and technical information such as the type of film used. Other research outputs could include an artist's exhibition, archaeological finds or the performance of a musical event, Support for multimedia requires greater digital storage than that required by text-based content. The

repository may require the installation of a streaming service so that students can download multimedia files.

5 DISCUSSION

A review of some of the seminal early writings about the potential of institutional repositories reveals that they were envisaged in fact as containing potentially all of an institution's digital assets. Boundaries would logically be determined by institutional strategic goals. Lynch [14], for example, suggests that "a mature and fully realized institutional repository will contain the intellectual works of faculty and students—both research and teaching materials..." According to JISC [1] the fact that most institutional repositories have ultimately focused on research outputs can be attributed primarily to the lack of incentives for creators to share and the lack of recognition by institutions of the value of sharing learning and teaching content.

The architecture of traditional learning management systems has not supported true integration of learning content. Access to learning materials is frequently controlled by authentication systems, which results in content not being shared across the institution and access being limited to lecturers and students in the respective course. Some LMS offer limited content management functionality as part of their system; however such functionality lacks the full potential of a repository solution.

In assessing this potential, JISC [1] asserts: "Whilst institutional VLEs [virtual learning environments] have, to some extent acted as stores for learning and teaching materials, they tend not to support the search and retrieval functions for a repository. Making this content more open, even within the institution, presents challenges for institutions with a commitment to open up their resources." This analysis reinforces the argument that research content used in learning and teaching should not actually be stored within the VLE / LMS.

With the growth of the Internet—and particularly Web 2.0—the learning lifecycle is being reconceptualised. It clearly encompasses much more than the traditional LMS, especially when so much more learning content is available through open access on the web. In this new paradigm research content should be stored in a repository environment which delivers it for multiple purposes within the institution, including the LMS. This is in fact the model proposed by Lyon nearly 10 years ago (Fig. 2 above).

This is not to say that all content should be stored in a single repository. Repositories should be considered as integrated components of larger systems and distributed infrastructure [15]. Nicholas [16] establishes a compelling case for not automatically assuming that the institutional repository should contain all content associated with all stages of the research lifecycle, particularly datasets. The core audience for this complex data tend to have requirements which are not currently supported by most institutional repositories; therefore they may store that data in a different repository federation, e.g. TARDIS for crystallography, which is better suited for the complexity of their needs. As mentioned previously, increased multimedia (film/music) content may require streaming servers, increased storage, and possibly new types of metadata.

In this new repository environment in which a wide range of research outputs are managed and in which traditional learning content is decoupled from the LMS, resource discovery becomes even more important. At Griffith University previously siloed research content across the University is being made discoverable through the aggregation of data from a range of different systems [17]. The utilisation of agreed data standards and protocols was a key enabler of the resultant service.

Interoperability and integration underpin the ways in which repositories work with other systems using common standards and protocols. It is important that each repository interoperates with other systems in order for the institution to maximize the benefits that come from the sharing of information. Critical to the discovery process is "good" metadata, i.e. metadata which not only adheres to accepted standards but also is extensive enough to allow varying degrees of granularity.

The IP and licensing issues which tend to characterise both research and traditional learning content would benefit from the proposed model. Authentication can be set within repositories to control access to individual items, e.g. limit to a defined course. At the same time institutional repositories have a history of working with publishers to allow the author's pre-print or postprint version of a publication to be made publicly accessible, i.e. no authentication.

An advantage of decoupling content from an LMS is that research content, for example, can be delivered independently from that LMS. At Griffith University several lecturers who teach in less

developed countries point their students to open access content in the institutional repository. As discussed previously in regard to the open access movement, this strategy enables the students to use scholarly research that they otherwise could not access because of the cost of journal subscriptions. The same model could be applied to learning objects.

6 CONCLUSION

This paper has shown that there needs to be a holistic approach to content management within an educational institution. However the architecture of traditional learning management systems has usually resulted in content not being shared across the institution and access being limited. To overcome these limitations, content needs to be decoupled from the LMS.

The model presented in this paper as a potential architectural solution highlights the importance of repositories in integrating different types of content for use in the learning lifecycle. Repositories provide a method of sharing content for different audiences. For example research outputs such as publications and data are not only used by other researchers but are also important resources for students. Research outputs as well as learning objects are important parts of the learning lifecycle.

Institutions need to recognise that repositories can add value to the learning lifecycle and to investigate how better to integrate them within their content management environment.

REFERENCES

[1] Joint Information Systems Committee. (2010). *Digital Repositories InfoKit*. Available from http://www.jiscinfonet.ac.uk/infokits/repositories/types

[2] IMS Global Learning Consortium. (2003). *IMS Digital Repositories v1.0 Final Specification*. Available from http://www.imsglobal.org/digitalrepositories/

[3] Crow, R. (2002). The Case for Institutional Repositories: A SPARC Position Paper. Washington, DC: SPARC. Available from http://works.bepress.com/ir_research/7/

[4] University Leadership Council. (2011). Redefining the Academic Library: Managing the Migration to Digital Information Services. Washington, DC. Available from http://www.educationadvisoryboard.com/pdf/23634-EAB-Redefining-the-Academic-Library.pdf

[5] Xia, J., & Opperman, D. B. (2010). Current Trends in Institutional Repositories for Institutions Offering Masters's and Baccalaureate Degrees, *Serials Review* (Vol. 36, pp. 10-18).

[6] Lehrmann, L., Hildebrandt, T., Rensing, C., & Steinmetz, R. (2007). Capturing, Management and Utilization of Lifecycle Information for Learning Resources. *Lecture Notes in Computer Science, 4753*, 187-201.

[7] Green, R., Awre, C., & Waddington, S. (2012). The CLIF Project: The Repository as Part of a Content Lifecycle. *Ariadne, 68*. Available from http://www.ariadne.ac.uk/issue68/green-et-al

[8] Pappalardo, K. (2008). *Understanding Open Access in the Academic Environment: A guide for authors*. Brisbane, QLD: Queensland Institute of Technology. Available from http://www.oaklist.qut.edu.au

[9] National Science Foundation. (2010). Dissemination and Sharing of Research Results. Available from http://www.nsf.gov/bfa/dias/policy/dmp.jsp

[10] Wolski, M., & Richardson, J. (2011). A Framework for University Research Data Management, *EDUCAUSE Australasia*. Melbourne. Available from http://hdl.handle.net/10072/39672

[11] Ball, A. (2012). *Review of Data Management Lifecycle Models. Research Report* (No. REDm-MED Project Document redm1rep120110ab10). Bath, UK.

[12] Wolski, M., Richardson, J., & Rebollo, R. (2011). Shared Benefits from Exposing Research Data, *IATUL 2011*. Warsaw: IATUL. Available from http://hdl.handle.net/10072/39023

[13] Lyon, Liz (2004). Realising the Scholarly Knowledge Cycle: The Experience of eBank UK. *CNI Task Force Spring Meeting*. Alexandria, Virginia.

[14] Lynch, C. A. (2003). Institutional Repositories: Essential Infrastructure for Scholarship in the Digital Age. *Portal: Libraries and the Academy, 3*(2), 327-336. Available from http://works.bepress.com/ir_research/27/

[15] Payette, S. (2010). *Repositories and Cloud Services for Data Cyberinfrastructure*. Available from http://duraspace.org/node/957

[16] Nicholas, N. (2010). *Position Paper: ADL Learning Content Registries and Repositories Summit*, from http://blog.linkaffiliates.net.au/2010/04/07/position-paper-adl-learning-content-registries-and-repositories-summit/

[17] Wolski, M., Richardson, J., & Rebollo, R. (2011). Building an Institutional Discovery Layer for Virtual Research Collections, *D-Lib* (vol. n, pp. 1-10). Available from http://hdl.handle.net/10072/39664