Repositories: What's the target? An ARROW Perspective

Derek Whitehead
Swinburne University of Technology

Repositories: What's the Target? An ARROW Perspective

Derek Whitehead

Swinburne University of Technology

dwhitehead@swin.edu.au

Abstract

The key observation about institutional research repositories is that they are under-used. This paper looks at why take-up has been slow, and what might encourage researchers to use these facilities.

This paper surveys the evolution of practice and rationale for the institutional research repository, looks at existing use studies and opinion surveys of users, and examines ways in which university libraries, the main proponents, have changed their approaches in response to their experience. The experience of Swinburne University of Technology, a partner in the Australian ARROW project, is drawn upon.

Keywords: ARROW, institutional, repository, research

Introduction

At the 2005 Online conference, Greg Tananbaum [1] summarised the key messages to be conveyed regarding an institutional research repository – it will save you time, and make you famous.

However, the clearest observation about institutional research repositories at present is that they are seriously under-used – not saving anyone time, not making anyone famous. Table 1 shows the take-up of Australia's institutional repositories of research

Table 1: Australian Repositories Accessible Through ARROW

Source	Number of items (27 April)
Australian National University	2,453
Curtin University of Technology	246
Monash University	122
Queensland University of Technology	832
University of Melbourne	459
University of Queensland	1,757
University of Tasmania	122

materials. [2]

This paper looks at why take-up has been slow, and what might encourage researchers to use these facilities. That this is the wrong order of things is obvious – the logical order is to determine the need, and then develop and implement means of meeting that need. This paper argues that that approach was taken, but that the need as originally perceived was insufficient to drive the level of adoption required for critical mass to be achieved.

This is no longer new territory – there has been a wide range of studies now, and most of them look at the needs of and benefits to the user. This is unsurprising, although it is surprising that it took so long to get down to seriously analysing the user's needs.

Definition

The term repository, as well as being inelegant and opaque to user groups, is a broad one. Any repository is a database. This paper is about repositories or databases of research outputs. A common approach to these has emerged over the past decade; the primary influence has been the software developed and provided by eprints.org and the OAI-PMH metadata standard [3]. An institutional research repository is a database which has developed to have certain additional features, which include

1   Institutional location and focus, in contrast to subject-based repositories not limited to one institution, or individual web sites. An institutional focus has some weaknesses, since scholarship transcends institutional boundaries and scholars often barely recognise them.

2   Focus on research outputs. There are fuzzy borders here, too, since one person's scholarly refereed article is another person's teaching material. Research outputs include theses, refereed journal articles, books and book chapters, and unpublished research reports.

3   Web visibility: a key feature of institutional repositories is that, unlike many databases, they provide access via the web to their content.

4   Full text availability. Repositories are means by which the outcomes of research are made widely available. They are also bibliographies, but for research repositories, the goal is for the full content of the item to be openly available. However, it is not always possible to make the full text widely available.

5   Structured information (metadata) about the documents, following the standard established by eprints.org; this enables the content of the database to be located using simple resource discovery metadata by eprint harvesters.

6   The repository is sustained and effectively managed over the longer term, including permanent locations.

There is debate over most of these characteristics. For example, it is sometimes – say for copyright reasons – not possible to make a full copy of an article available within a repository, or it may be possible to make a copy available only on a more limited basis, or merely provide a bibliographic listing. And, the fact that the content of repositories

is effectively published, using most but not all definitions of published, creates some issues for categories of research usually regarded as unpublished such as theses, in the sense that their availability does not mean prior publication in a way which would make them ineligible for first publication elsewhere.

Use and Users

Institutional repositories have three main groups of users: personal, institutional and global. For users within the institution – personal and institutional – the repository is a means of organising content, publishing it or making it available, managing it over the longer term, providing a fixed copy for subsequent re-uses, standardising the means and form of storage with similar objects. For external users the main concern is to location of items, access to full texts, and discovery of the existence of content of repository contents, as well as information about them.

The ARROW Project

The Australian ARROW project was funded by the Australian Government as one of four projects, and runs from 2004-2006. The project is to develop and implement institutional repository software and processes. The partners include three universities – Monash University, Swinburne University of Technology and the University of New South Wales – and the National Library of Australia. Software development work is being undertaken by VTLS, and the ARROW software is being built on the Fedora repository software, and VTLS's own VITAL product.

While there is a strong focus on software development, the ARROW partners are each exploring ways in which an institutionally-based research repository can provide value to the research communities of their universities. They are discovering that this value is more varied than anticipated, and varies significantly from one institution to another. The lessons of the project to date have informed this paper. [4]

Discussion of Institutional Research Repositories

Early institutional research repositories mostly followed the clear logic of Stevan Harnad's long-standing advocacy of open access – the green and gold routes to open access. [5]. This approach was backed up by the development of simple, easy to use software and a range of ancillary services and resources, such as the data about publisher policies developed as part of the ROMEO project and incorporated into SHERPA [6]. Harnad's initial approach was focused mainly on external users – the focus was on putting the body of scholarly literature online and making it openly accessible. But while the logic of the Harnad approach is obvious, over the past decade it has persuaded relatively few people in the user target group, researchers. A large part of the agenda was driven by providers, particularly librarians – but there was not automatic or even widespread support amongst researchers.

A SPARC paper by Raym Crow, published in 2002 [7] had a limited focus on the rationale for repositories, which was seen to be obvious: repositories would create a

new publishing paradigm, providing better access to scholarly literature. Crow made a second point – that repositories would also enhance the profile and status of the institution, and thereby contribute to institutional visibility and prestige.

This was an approach focused on external users. The influential PALS study in 2004 described the main "documented uses for IRs" as being [8]

1 Scholarly communication
2 Education
3 E-Publishing
4 Collection management
5 Long-term preservation
6 Institutional prestige
7 Knowledge management
8 Research assessment exercises

Again, a largely institutional approach, but more broadly focused, with a wider range of institutional goals referred to – for example, research assessment has been added to the core external institutional goals of prestige and visibility. However, in later lists of purposes, such as that of the Foster & Gibbons [9], some of these do not appear at all.

The EPIC study, by Alma Swan and others, undertaken by Loughborough University and Cranfield University, with Key Perspectives Ltd, was another survey, akin to the PALS study, and completed in July 2004 [10]. Like that study, it took a twofold approach, with a primarily outward focus. The study summarised the "advantages" of institutional e-print archives in a conventional externally-focused style, also incorporating research assessment:

1 Increased access to published research
2 Increased impact of published research
3 Provision of enhanced citation analysis (new measures of impact)
4 Provision of a tool for the compilation of 'institutional CVs' and institutional impact (a marketing tool for institutions)
5 Provision of a tool for the compilation of individual researchers' CVs and personal impact

Although this section of the study received little space, it was complemented by a cost benefit analysis, which set out just three benefits to the institution

1 The impact of British research would be maximised
2 Visibility and accessibility would be improved
3 The RAE would be cheaper and easier to administer, at both institutional and funding-council level.

Not all repositories arose from universities. OCLC, which conducts research "on behalf of the library community", among other things, established a research repository in September 2003. Three purposes were specified [11] – the principal reason was to maximise the visibility and impact of OCLC research; make the OCLC research web site "more nimble" (a process re-engineering goal), and demonstrate OCLC research technologies (a commercial goal). Interestingly, only half of the 900 or so items in the repository exist as full text.

Broad studies of institutional repositories continue to appear. Heery and Anderson reviewed current repository activity for UKOLN (UK Office of Library Networking) in early 2005 [12], the scope extending beyond research repositories. The SPARC ACRL

IR Workshop held in April highlighted the range of software options – Digital Commons, Digitool, DSpace and eprint.org – and also examples from large and small universities. [13].

JISC has supported work on defining an e-print and open access model for the UK in recent work by Alma Swan and others [14]

The recent Berlin 3, a follow-up to the 2003 Berlin Declaration [15] featured a presentation from Bill Hubbard of the SHERPA consortium which demonstrated a broad-ranging approach to the issue of purpose and a focus on the individual researcher. [16]

The most recent discussion has widened the range of possible purposes of institutional research repositories, and much new discussion has focused on their value in internal processes. Harnad himself now takes the view that a precondition for success of institutional repositories is that their use be mandated [17]. This will rapidly ensure that they reach a critical mass which will enable them to fulfil their primary roles, he suggests. But he and others are also now exploring the variety of internal reasons for institutional repositories – 'The point, Harnad explained to delegates, is that the best way to encourage researchers to self-archive is to request that they do it "not for the sake of Open Access, but for record-keeping and performance evaluation purposes." [18] Then, a single keystroke can make it open access. A very interesting recent article by Carr and Harnad looks at the issue of balancing effort and benefit, examining user workflows in self-archiving to demonstrate the very limited time requirements of archiving. [19] This has brought the discussion a long way from the basically external user focus of earlier work.

Focus on Reasons for Slow Adoption of Repositories

There have been some interesting contributions from institutions seeking to explain the slow take-up of institutional research repositories by the ostensible beneficiaries – a good idea, but too much trouble, they seemed to be saying.

For example, the Daedelus Project at the University of Glasgow [20] produced an excellent analysis of reasons for relative failure. In Australia, Paula Callan at Queensland University of Technology also provided a good analysis – from an institution which had actually mandated deposit. [21]

At the same time, there has been a proliferation of software, some of which incorporated features which aimed to make the process more appealing in a variety of ways. The University of California eScholarship Repository provided a simple interface which permitted the user to establish an account, be notified of new papers in areas of interest, obtain statistical information, and take an individual view of the database [22] Proquest, using the same software base, developed a commercial version (Digital Commons) which aims to make use of an IR as easy as possible [23].

Research on Repositories

There have also been several recent pieces of systematic research about research repositories. Two of these are of particular interest in the context of this paper. The

highlight the repetitive internal processes which exist in universities in their handling of information about research, and shift the focus not only to the institution, but to the individual too.

First, Hey has reported on surveys undertaken at the University of Southampton. The study analysed the extent to which faculty had their publications listed on the Web, and the proportion which also included full text. She reports "Academics are asked to provide information about their publications for a large variety of demands and for which the format required varies each time." The study of academic staff threw useful light on what the nominal beneficiaries really wanted. These included [24]

- o The ability to "input data once, and use this for multiple outputs"; for example, to providing information for a school's research report and the University Research Report.

- o Provide outputs to meet a range of other needs, each of which might require a different format, such as information and documents for CVs, job and fellowship applications, research proposals, promotional material for a school's web site or in print, export into EndNote and other bibliographic software, the School's website including lists on its Web site by person, group, project or topic, or an individual's personal Web site

- o "Providing a secure store for publications in electronic form at an early stage for sharing with the research community or enabling visibility of one's work as a professional".

- o "Providing information and documents for the Research Assessment Exercise (however it is presented in its new form)"

Second, at the University of Rochester, Foster and Gibbons [25] describe the results of a systematic study of academic staff using a methodology derived from anthropology. They found that faculty members wanted to be able to do a number of things, and these were often very mundane.

A large category related to management of their documents during the research and writing stages – working with co-authors, managing version control, sharing versions of an article, working from different locations and different hardware. At present, however, few institutional repositories aim to help manage workflow.

A second category related to management of completed or published material. Researchers wanted to be able to organise their own material the way they wanted it, make sure that it was always available at all times and locations, eliminate concern about hardware and software and keep everything related to computers "easy and flawless", and "reduce chaos or at least not add to it." They also wanted assistance with copyright issues.

Finally, they wanted to keep up in their fields, but not be any busier. This picture of embattled researchers struggling amongst competing demands to keep their heads above water perhaps reflects the real context for institutional research repositories. Their impatience with the complexity of using a networked computer is real.

The University of Rochester study offers a number of conclusions which just involve saving time – not maintaining a server, not doing complicated things, not having to file and manage copies of their articles on computers, not having to find and email copies

of their articles to colleagues. In the end, the research showed that researchers wanted help in managing their research and information about it as easily as possible. Individual researchers, not surprisingly, had a focus on their personal needs, and there was little institutional dimension in the outcomes of the study.

Repositories and Research Evaluation and Impact

In the UK context, it is been suggested for some time that research repositories may be useful in assisting research evaluation.

For example, Michael Day of UKOLN develops the argument that institutional repositories, in time, can be used to support the UK Research Assessment Exercise (RAE) [26]. He suggests that repositories could support the UK RAE (Research Assessment Exercise) in three ways, and develops this case in detail. The three approaches are: provide accurate metadata and the full text of documents; more ambitiously, use citation linking and analysis across repositories; and apply other Web-based metrics, such as accesses or downloads. As simple metrics are succeeded by greater interest in impact and access, repositories clearly have a role.

At the same time, a growing preoccupation with measuring the impact of research, whether indicated by citation or some other method, has given rise to a number of studies of the effect of open access (in an institutional repository or an open access journal) on impact [27]. Antelman's study of four disciplines finds that open access does lead to a higher level of research impact. [28]

In Australia, the Commonwealth Department of Education, Science and Training (DEST) has recently (March 2005) published an issues paper on assessment of research quality and impact [29]. The paper discusses means of measuring research quality and impact, as well as knowledge diffusion.

Governments are very interested in the potential of institutional repositories to meet their policy objectives relating to research, and particularly research which they have funded. In Australia, $12 million has been allocated to projects which are principally focused on institutional repositories. The expression "public funding, public knowledge, public access" was coined in a government presentation to scholars. [30]

User-Centred Approaches

The presentation by Greg Tanenbaum mentioned above [31] had a focus on the repository as a way to save time and get famous at the same time.

Saving time focuses on the approach of the University of California's BE Press, which provides a hosted service with a variety of time-saving features. He suggests that posting objects takes only 3 minutes, files are automatically converted to a standard format (PDF), version histories can be created, and there is "no need to rely on Larry, the local IT guy." The service is sold as an easy alternative to home-grown sites and commercial publishers – as an outsourced means of managing your published output in a stable, long-term environment.

How will an institutional repository make you famous? Tanenbaum suggests that services such as push email, configurable saved searches, customised email alerts, full-

text searching, the capacity to browse by author, and the addition to OAIster and new search engines like the new Google Scholar service will combine to make the content of IRs immensely more available to interested people.

Tanenbaum emphasises providing value to the individual user. He suggests that sites need customized branding (logos and links, photographs for those who like them), that we must recognise the wide variety of content and publication types and community structures: don't shoehorn, keep in touch. His approach combines the original approach of Harnad's subversive proposition – externally focused, enhancing visibility – and the recent discovery of extensive value in re-engineering institutional and personal processes

Building a Composite Picture of Value

During 2004 and 2005 there has been rapid development in thinking about repositories and their users. Much of this is based on more systematic examination of user requirements and the role of a repository in meeting them, and the experience that we now have. It has been assisted by

- o Emergence of more "commercial" looking (i.e. customer focused) software, and a better understanding of the needs and interests of different kinds of users and beneficiaries – personal, institutional and global.

- o Alignment of repositories with other kinds of functionality, like data collection (DEST data in Australia), or journal publishing (BEPress in the US), or production of research reports, or justification of promotion – and the possibilities of re-engineering processes to make them more efficient.

- o Awareness that management of research outputs, for particular purposes, fits into the wider context of management by universities of their rapidly growing online information resources.

- o Perception of clear value in an IR which still lacks critical mass. Early arguments focused on the importance of putting all research online, on open access; but critical mass can occur for an individual, a department, a cluster of researchers, or an institution.

The creation of institutional repositories over the past few years has been extremely rapid, and is now likely to be followed by a corresponding development of content.

This will increasingly recognise the diversity of interests of the users of these repositories – personal, institutional and global. Within institutions, use is likely to focus first on the needs of individuals, and second on the needs of the institution itself, and of component elements – departments, faculties and research centres, research administrators, and marketing staff. Beyond the institution, the focus will be on resource discovery and access, but also on a wide range of government and other administrative uses.

A composite picture of value is complex and still evolving. Here is a suggested outline.

1 Resource discovery and access: the basis for this has been around since the development of OAI-PMH provided the means of providing simple access to

distributed repositories. This can only improve, as new tools are added to those that exist, and the capacity to search full text develops further. The sheer efficiency of locating an item and locating it again when needed through simple search or browse functionality is a major benefit.

2   New modes of publication and dissemination: the other side of the first point, but harder to pin down as a benefit to the user. There may be elements of the existing mode of scholarly publication that we don't want to lose (quality control) and other elements we are happy to lose (costly overheads) but the future is uncertain. There is, however, a steady advance of open access to research outputs. In some contexts (such as the BE Press software, and planned for ARROW) there is a repository base for production of open access journals – green and gold in one!

3   Research evaluation and assessment is seen as an area of growing importance for institutional repositories, and for open access more widely. Within the UK, Australia and elsewhere institutional repositories are being actively incorporated into new thinking.

4   Institutional and personal impact is a major driver. The addition of new tools which measure the extent of use of individual items had made this more significant. Increase in interest in impact and access measures for published research has also potentially enhanced the role of the repository as a tool which both enhances access and counts it.

5   Information, knowledge or asset management, down to the individual level, and up to the sub-institutional (e.g. faculty or centre) and institutional levels. Asset management covers a wide range of elements – the creation of outputs in a wide range of formats (published journals, research reports, promotion applications), format management, version management, long-term preservation, and the management of intellectual property rights.

6   Efficiency, process improvement, and the saving of time for the researcher are all achievable goals of institutional repositories. At present, universities have a number of processes which essentially involve repetition of the same kind of thing – compiling lists of references for various purposes – which can be replaced by a repository with effective output processes. Foster & Gibbons emphasise just making things easy – most of their conclusions involve saving time.

The future of institutional research repositories rests on an understanding of their usefulness, both by those developing and marketing, and those for whose use they are intended. Over the past year there has been rapid development in this understanding. The target is value to the user, and we are only just beginning to understand it adequately.

References

[1] Tananbaum, G, 2005, Institutional Repositories: crossing the Chasms: Paper presented at Information Online 2005: 12th Exhibition and Conference, Sydney, 1-3 February, 2005. http://conferences.alia.org.au/ online2005/papersandpresentations.php
[2] See http://www.arrow.edu.au   The table provided here excludes two sources which are not institutional research repositories: Agrigate and Australian Policy Online.
[3] http://www.eprints.org
[4] See the ARROW web site at http://www.arrow.edu.au for more information.

[5] Poynder, R, 2004, "Ten years after", Information today, 21(9) (Oct 2004) gives a brief history of the progress of Stevan Harnad's original "subversive proposal".

[6] See http://www.sherpa.ac.uk/romeo.php for the list of publishers' copyright policies. The SHERPA site is at http://www.sherpa.ac.uk

[7] Crow, R, 2002, The Case for Institutional Repositories – a SPARC Position Paper. Release 1.0 SPARC, 2002. p.6 http://www.lib.duke.edu/its/diglib/digarchive/repositories.html

[8] Mark Ware Consulting, 2004, Publisher and Library/Learning Solutions (PALS): Pathfinder Research on Web-based Repositories: Final Report. Bristol, PALS. Pp.20-21.
http://www.palsgroup.org.uk/

[9] Foster, N and Gibbons, S, 2005, "Understanding Faculty to Improve Content Recruitment for Institutional Repositories", D-Lib magazine, 11 (1) (Jan 2005).
http://www.dlib.org/dlib/january05/foster/01foster.html

[10] Swan, A, et al, 2004, Delivery, Management and Access Model for E-prints and Open Access Journals within Further and Higher Education: a joint report by The Electronic Publishing Innovation Centre (EPIC) in partnership with Key Perspectives Limited. Loughborough: EPIC and Key Perspectives Ltd. Pp.7-8 and 72-75.
http://www.keyperspectives.co.uk/OpenAccessArchive/E-prints_delivery_model.pdf

[11] Hyatt, S and Young, J, 2005, "OCLC research publications repository", in D-Lib magazine, 11 (3) (Mar 2005) http://www.dlib.org/dlib/march05/hyatt/03hyatt.html   For the repository, see http://www.oclc.org/research/publications

[12] Heery, R and Anderson, S, 2005, Digital Repositories Review. Final draft, 19-02-2005. UKOLN, 2005. http://www.jisc.ac.uk/uploaded_documents/rep-review-final-20050220.pdf

[13] ACRL, 2005, SPARC ACRL IR Workshop: Building a Successful Institutional Repository: an Introduction for Smaller Libraries, ACRL 12th Annual Conference, April 8, 2005, Minneapolis. http://www.arl.org/sparc/meetings/acrl05/index.html

[14] Swan, A, et al, 2005, "Developing a model for e-prints and open access journal content in UK further and higher education", Learned publishing, 18(1) (Jan 2005) pp.25-40.
www.keyperspectives.co.uk/ OpenAccessArchive/Eprints_LP_paper.pdf

[15] Harnad, S, 2005, "The Implementation of the Berlin Declaration on Open Access: Report on the Berlin 3 Meeting Held 28 February – 1 March 2005, Southampton, UK", D-Lib Magazine, v.11, no.3 (Mar 2005). http://www.dlib.org/dlib.march05/harnad/03harnad.html

[16] Hubbard, B, 2005, Open Access Institutional Repositories in UK Universities: Presentation to the Berlin 3 Meeting, 28 February – 1 March 2005, Southampton, UK.
www.eprints.org/berlin3/ppts/03-BillHubbard.ppt

[17] Harnad, S, et al, 2004, "The access/impact problem and the green and gold roads to open access", in Serials review 30(4). http://dx.doi.org/10.1016/j.serrev.2004.09.013

[18] Poynder, R, 2005, "Time to Walk the Talk?", in Open and Shut, March 17, 2005. p.4
http://poynder.blogspot.com/2005/03/time-to-walk-talk.html

[19] Carr, L and Harnad, S (2005), Keystroke Economy: a Study of the Time and Effort Involved in Self-Archiving. Public Draft 1, 15/3/05. eprints.ecs.soton.ac.uk/10688/
01/KeystrokeCosting-publicdraft1.pdf

[20] Mackie, M, 2004, "Filling institutional repositories: practical strategies from the DAEDALUS Project", Ariadne Issue 39 (Apr 2004). http://www.ariadne.ac.uk/issue39/mackie

[21] Callan, P, 2004, The development and implementation of a university-wide self-archiving policy at Queensland University of Technology: Insights from the frontline, Paper presented at Institutional Repositories: the Next Stage. Workshop presented by SPARC and SPARC EUROPE November 18-19, 2004, Washington DC. Washington: SPARC.
http://www.arl.org/sparc/meetings/ir04/presentations/callan_files/callan.pdf

[22] See http://repositories.cdlib.org/escholarship

[23] See http://www.umi.com/proquest/digitalcommons/

[24] Hey, J, 2004, "Targeting academic research with Southampton's institutional repository", Ariadne Issue 40 (July 2004). pp.8-9. http://www.ariadne.ac.uk/issue40/hey

[25] Foster, N and Gibbons, S, 2005, "Understanding Faculty to Improve Content Recruitment for Institutional Repositories", D-Lib magazine, v.11, no.1 (Jan 2005).

http://www.dlib.org/dlib/january05/foster/01foster.html

[26] Day, M, 2004, Institutional repositories and research assessment. A supporting study for the ePrints UK project, v. 0.1 (draft), 2 December 2004. pp.13ff

http://www.rdn.ac.uk/projects/eprints-uk/docs/studies/rae/rae-study.pdf

[27] Harnad et al, 2004, "The access/impact problem and the green and gold roads to open access", in Serials review 30(4). http://dx.doi.org/10.1016/j.serrev.2004.09.013, cited above.

[28] Antelman, K, 2004, "Do open access articles have a greater research impact?', College & Research libraries 65(5), pp.372-382.

http://www.lib.ncsu.edu/staff/kantelman/do_open_access_CRL.pdf

[29] Department of Education, Science and Training (DEST), 2005. Research Quality Framework: assessing the quality and impact of research in Australia. Issues paper, March 2005. Canberra, DEST. http://www.dest.gov.au/resqual/issues_paper.htm

 [30] Arthur, E, 2004, Presentation to Changing Research Practices in the Digital Information and Communication Environment, Canberra, June 1, 2004.

http://www.humanities.org.au/NSCF/current.htm See also related documents located at http://www.ecs.soton.ac.uk/~harnad/Hypermail/Amsci3098.html

[31] Tanenbaum. See note 1.