

Institutions, their repositories and the Web

Alma Swan^{1,2,3} and Leslie Carr¹

¹School of Electronics & Computer Science, University of Southampton, Southampton SO17 1BJ, UK

²Marketing & Strategic Management Group, Warwick Business School, University of Warwick, Coventry, CV4 7AL, UK

³Key Perspectives Ltd, 48 Old Coach Road, Truro, TR3 6ET, UK

<http://eprints.ecs.soton.ac.uk/14965/>

Abstract

It will soon be rare for research-based institutions not to have a digital repository. The main reason for a repository is to maximise the visibility of the institution's research outputs (provide Open Access), yet few contain a representative proportion of the research produced by their institutions. Repositories form one part of the institution's web platform. An explicit, mandatory policy on the use of the repository for collecting outputs is needed in every institution so that the full research record is collected. Once full, a repository is a tool that enables senior management in research institutions to collate and assess research, to market their institution, to facilitate new forms of scholarship and to enable the tools that will produce new knowledge.

Introduction

Over the last couple of years there has been considerable development and progress in the world of institutional repositories. We have seen their numbers rising at an average of one per day around the world. At national level there have been large programmes of investment in some countries as the potential of networked repositories for improving the communication of research and enhancing the productivity of the research community has been understood, for example, the UK's Joint Information Systems Committee's huge *Repositories Programme* (Joint Information Systems Committee Repositories Programme, 2005-7) and *Repositories & Preservation Programme* (Joint Information Systems Committee Repositories and Preservation Programme, 2007-8) and the *Australian Scheme for Higher Education Repositories* (Australian Scheme for Higher Education Repositories, ASHER, 2006-8). It is unlikely that any serious research-based institution will not have its own digital repository by the end of this decade. This article looks at the current value of repositories to their institution and what promise repositories hold for the future of research and learning.

Institutional repositories – what are they for?

So what are repositories for? What place do they occupy in the life of the institution? What place *should* they occupy? So far, repositories are not being used to their full potential and institutions are falling into the same thinking trap about repositories as they have about the Web itself. Institutions have so far treated the Web largely as a digital version of what they have always done, when its potential – for them and for their researchers – is so much greater. Digital repositories should be one of the institutional Web-based tools that take research into places it has not been able to reach before.

One important issue in the context of this discussion is that the primary reason for establishing a digital repository is to increase the visibility of the institution's research output by making it Open Access. Just about every institution with a repository cites this as a reason for having set it up according to our own small survey of European repositories (unpublished). Certainly, the repository is the ideal vehicle for making the work of the institution visible. Relying on pages on the institution's website is not satisfactory. Departmental and research web pages are frequently out of date or incomplete. Indeed, our own informal study on this recently revealed that even in computer science departments, where one might expect maximal and optimal use of the Web, this is the case: in our own School of Electronics & Computer Science at Southampton one third of researchers have no working home page and the figure for the world's premier computer science research department – at MIT – the situation is not much better with twenty percent of professors (of all grades) lacking a web page.

With these sorts of statistics regarding the currency and completeness of author home pages, it is absolutely clear that the place for collecting and disseminating the outputs of a research-based university is the institutional repository. At the moment, however, very few institutional repositories are being successful in amassing a sizeable proportion of the institution's current research output, though this will change as the benefits of doing this become apparent to senior research managers. They will see that the way to achieve growing repositories with more than the 'baseline' 15-20% of potential content is to put in place well thought-through, mandatory policies on deposition. The big research funders have already seen this and produced such policies; eventually, universities and research institutes will see the light and follow suit, reaping the benefits of maximal visibility for their research outputs as a result.

One tale may be salutary in this respect. It is what we call *The University of Southampton Conundrum*. Southampton University is a fairly large UK 'Russell Group' institution. The Russell group is an association of twenty research-intensive British universities that between them are awarded two thirds of the national research funding budget allocated by the UK Research Councils, thus comprising the cream of British research-based higher education establishments. Southampton can therefore be assumed to be strong in research and its teaching merits high ranking in the UK national league tables, too: a worthy institution all round. In the *Times Higher Education Supplement* (Times Higher Education Supplement World University Rankings) and the Shanghai Jiao Tong University (Shanghai Jiao Tong University Academic Ranking of World Universities) annual rankings of world universities Southampton comes somewhere between the 100th and 200th places each year. Yet it appears in 25th position in the G-Factor ranking (The G-Factor International University Ranking), calculated each year by measuring the number of hyperlinks on the sites of the world's leading universities to other universities' websites. A peer-reviewed measure of web presence, in other words. Why does Southampton do so well here? Certainly, the university has a large research programme which inevitably encourages linking from other institutions, but undoubtedly the fact that it has a digital repository housing thousands of contemporary research articles has most to do with it. Southampton simply has a huge online presence because of this (Harnad, 2006; Brody et al, 2007).

Other universities could easily boost their visibility in the same way. The trouble is that except for a small number of institutions around the world that have big, growing repositories containing current research articles (rather than just, say, theses, grey literature or legacy literature from the past) most

repositories are to all practical purposes empty. They are not only *not* enhancing their institution's online visibility, they are also actively projecting a very poor image of their institutions to the world. The shop window is empty.

In our survey of European repositories, over 60% of repository managers reported that recruiting content to their repository has been possible but not easy: one third said they have found it difficult or very difficult. The Southampton repository would not be full without the mandatory policy introduced in 2004: that much is not only borne out by the evidence about the effectiveness of policies and mandates (Sale, 2006; Sale, 2007) but is also openly acknowledged by the researchers working there. All the evidence shows that advocacy and mediated deposit by library/repository staff can be effective tools in recruiting content, but only in the presence of a mandate from senior management. Without the latter, making the case to researchers and gathering their outputs from them is extremely labour-intensive and ultimately not very successful.

Yet it is senior management who arguably have the most to gain from ensuring that their shop window to the world looks full and enticing. Around one-third of repository policies have been reviewed and approved by the vice chancellor (rector or provost) and a further 10% by his/her deputy with research responsibilities, yet so few of these senior managers have had the mettle to take it all the way to mandatory status. This is despite the fact that, having risen from the ranks themselves, they know only too well the mentality of busy researchers, who can always find priorities for their time other than administrative tasks like depositing their articles in their repository. As one researcher in our own School at Southampton put it:

"I have seen all the benefits of using our repository in the four years since the mandate was developed. I remember how we used to get constantly asked for lists of publications for administrative purposes and how tedious and time-consuming it was to create a list of all my research last year, or a list of every publication on which I had collaborated with people inside the university but outside our school, or every publication which came from such and such a source of funding, or a list of my CV or a Web page. Now it's simple - you just say 'go and look in eprints'. Even the RAE. So the repository is really, really useful and well worth it... BUT having experienced all those benefits and with all that in mind, if the mandate wasn't in place I still wouldn't use the repository because I just don't do things."

*Research Group Leader, School of Electronics & Computer Science,
University of Southampton*

The dearth of mandatory policies for filling university repositories is therefore bemusing to say the least.

An institutional investment

The bemusement grows when the investment that universities are putting into their repositories is considered. Our own small sample of European repositories revealed that the average amount spent on setting up the repository was about €10,000, with a labour investment of some one or two FTEs to keep the repository running once established. The study conducted by the Association of Research Libraries (on US repositories) reported an average start-up cost for a repository of \$183,000

(€125,000) (Association of Research Libraries, 2006); the Primary Research Group study (on repositories worldwide) found that the average cost of setting up a repository was \$79,000 (€54,000) (Primary Research Group, 2007). These are not huge sums in the context of overall university budgets, but they nevertheless represent a noticeable investment, especially when ongoing costs are taken into account.

A university's overall investment in research is a much larger scale matter, of course. Until recently, the visibility and reputation of the institution based on that research had to be contrived in one main way – by its researchers publishing their results in journals and books and hoping that the right people around the world would see these outputs, cite them and build upon them. The efficacy with which this process happened depended upon the appropriate journals being available to the appropriate people, something that was by no means a certainty. Now, by using a repository properly, a university can raise its research visibility hugely (viz. Southampton). With all the research outputs collecting day-on-day in the repository, and with Google providing the worldwide indexing service needed to search and retrieve repository content – as it does – the repository will become the competitive tool for maximising research visibility, status and reputation of an institution. Indeed, this does not just pertain to research, since the university's pedagogical reputation will also be centred around its digital resources and wealth in the same way. The locus for promotion of the institution, therefore, moves from institutional and departmental web pages and other marketing channels to a new place – its digital collections of real outputs and resources.

An institutional payoff

The promise for a vastly better presence and enhanced status is the primary payoff for universities that properly grasp the opportunities their repository can bring them. There is no doubt that a repository increases the usage of an institution's research outputs. The University of California's eScholarship repository logged 2 million downloads less than four years after it was established. Moreover, the rate of downloads is increasing. The first 0.5 million downloads took 2 years, the first million took three years and by nine months later the total had reached 2 million. The number of items in the repository at the time was 10,000. At the time of writing there are 20,000 items in the repository and these have been downloaded 5.3 million times. Downloads are currently running at around 38,000 per week (University of California eScholarship Repository). Of course, this is a large repository at one of the world's most prestigious universities but this sort of success story with respect to usage is not confined to such a repository. The repository for the School of Electronics & Computer Science at the University of Southampton in the UK is a smaller, School-based archive containing around 11,500 items, about half of which are full-text research papers. Downloads from this repository have averaged around 30,000 per month over the last year, an impressive level of outreach to researchers around the world who wish to read these articles.

There are other potential payoffs, too. A repository offers the locus for dissemination of outputs in a more formal way. In the last half of the twentieth century the proud tradition of university presses largely died out as commercial publishers came to dominate academic publishing. Only the largest university presses could compete head-on with their commercial rivals. The upshot, at the beginning of this century, is that the situation for young scholars in the arts and humanities, in particular, is very difficult. These young people need to get their 'first book' published – often it is based on their doctoral thesis work – in order to get their foot firmly on the academic career ladder. Unfortunately, the small sales volumes expected of such monographs mean that they have frequently been

considered an uneconomic prospect by most commercial publishers. Now there is something of a renaissance in thinking and acting. Presses are springing up again in universities, enabled by the move to digital information and the concomitant savings that can be made in many aspects of publishing. The logical strategy for a university is to use its digital research repository as the submission tool for works that the press intends to publish formally – that is, to provide peer review, an imprint and copy editing services to the author. Indeed, that is what is starting to happen. A number of universities now are combining press and repository to create an efficient and highly functional digital publishing platform. An example of this is the eScholarship repository at the University of Sydney (University of Sydney eScholarship). Moreover, some university presses that use the repository platform are also moving to Open Access publishing of both journal and books, such as Amsterdam University Press (Amsterdam University Press repository) and the University of Pittsburgh Press (University of Pittsburgh Library). We should expect to see other institutions follow suit.

For senior managers in a university, the repository is a tool *par excellence*. First, it provides the means of recording and assessing research in the institution. Coupled with an institutional Research Information System – which records who is working where in the institution, with whom, on what project, funded by what grant and using which equipment – the repository facilitates the measurement of research activity in the institution with great granularity. Research managers are able to retrieve exactly the information required for research assessment exercises, for institutional record-keeping and for research management and planning. Repository software has even been adapted to enable this function: for example, the EPrints software now has an add-on which enables an institution to select and present its research outputs in formats required for the periodic national Research Assessment Exercise in the UK. Second, the repository becomes part of the university's shop window for the world. Currently, no university is using the repository optimally for this function. What is needed is a holistic view of the whole institution's web offerings, with extensive links between departmental web pages, researcher web pages, project web pages, the institution's press office web pages, the institution's research office web pages and the repository where the intellectual wealth of the institution is accumulating. In these early days no institution has yet implemented the linking and inter-relationships across its entire web presence that would optimise its image, facilitate the navigation of its website by interested visitors and maximise its overall impact.

The same sort of holistic approach will see repositories play a major role in enabling the progress of scholarship within and institution and outside its confines. The trend towards interdisciplinary research is gathering pace as techniques, methodologies and approaches developed in one discipline become increasingly applicable, indeed essential, in others. This trend is not confined to the natural sciences, though it is certainly very pronounced in these disciplines now, with increasing amounts of research funding being allocated to inter- or multi-disciplinary approaches to resolving some of the big questions in biomedicine and environmental science particularly. Not all research that would benefit from a cross-discipline approach is the subject of such explicitly-organised interdisciplinarity, though. This is where repositories will come into their own over time. Collecting the research outputs of the institution in one location, tagging entries with appropriate terms and keywords, tying the repository in to other sources of information within the institution and providing a rich context-based navigation capacity will facilitate the enhanced communication, serendipitous discovery and semantics-based relating of disparate entities that could profit from partnership and collaboration. The research life – and the teaching – of institutions will be enriched by such architectures and ways

of thinking. We are seeing the beginnings of such systems in a few far-seeing US universities – Harvard, Minnesota and Cornell – and it is salient that the initiatives beginning right at the top of the institution (Harvard University Planning Committee for Science and Engineering, 2006; Rosenstone, 2006; Devare *et al*, 2007).

Finally, in a manner analogous to the facilitation of interdisciplinary research, repositories are positioned to play a major role in facilitating the workings of the Semantic Web. The true benefits of the Web on science and scholarship have yet to be realized. They will come from the intelligent (semantic) machine-to-machine linking of items, not just at item level, but at sub-item level such as by paragraph, sentence, phrase or significant concept. The benefits come not only from the more precise, meaningful linking but from the fact that machine tools will be able to work across documents and records, putting together pieces of information that have additional or new meaning and truly taking us into a realm where the whole is greater than the sum of the parts. Pieces of information from similar, dissimilar or downright different sources will be put together to create knowledge – new knowledge – and make possible more enlightened and advanced thinking. Some tools to do this already exist and more will be developed. They can only work, of course, on a fully Open Access literature and that, before long, will be provided in abundance by university repositories. They are the fundament of future scholarship and innovation.

References

Australian Scheme for Higher Education Repositories (ASHER), 2006-8.

[http://www.dest.gov.au/sectors/research_sector/policies_issues_reviews/key_issues/research_quality_framework/asher_and_iap.htm#Australian_Scheme_for_Higher_Education_Repositories_\(ASHER\)_programme](http://www.dest.gov.au/sectors/research_sector/policies_issues_reviews/key_issues/research_quality_framework/asher_and_iap.htm#Australian_Scheme_for_Higher_Education_Repositories_(ASHER)_programme)

Brody, T., Carr, L., Gingras, Y., Hajjem, C., Harnad, S. and Swan, A. (2007) Incentivizing the Open Access Research Web: Publication-Archiving, Data-Archiving and Scientometrics. *CTWatch Quarterly*, 3 (3). <http://eprints.ecs.soton.ac.uk/14418/>

Devare, Medha Jon Corson-Rikert, Brian Caruso, Brian Lowe, Kathy Chiang and Janet McCoe. 2007. VIVO: Connecting people, creating a virtual life science community. *D-Lib Magazine*, 13(2), July-August.

<http://www.dlib.org/dlib/july07/devare/07devare.html>

G-Factor International University Ranking.

<http://www.universitymetrics.com/tiki-index.php?page=G-Factor>

Harnad, Stevan. 2006. Metrics and assessment. *Open Access Archivangelism*, April 14.

<http://openaccess.eprints.org/index.php?/archives/79-guid.html>

Harvard University Planning Committee for Science and Engineering. 2006. Enhancing Science and Engineering at Harvard. Harvard University, Cambridge, MA.

http://www.provost.harvard.edu/reports/UPCSE_Interim_Report.pdf

Joint Information Systems Committee. Digital Repositories Programme, 2005-7.

http://www.jisc.ac.uk/whatwedo/programmes/programme_digital_repositories.aspx

Joint Information Systems Committee. Repositories and Preservation Programme, 2007-8

http://www.jisc.ac.uk/whatwedo/programmes/programme_rep_pres.aspx

Primary Research Group. 2007. The International Survey of Institutional Digital Repositories.

<http://www.primaryresearch.com/200711071-Information-Science-excerpt.html>

Sale, Arthur. 2006. Comparison of IR content policies in Australia. *First Monday*, 11 (4).

http://www.firstmonday.org/issues/issue11_4/sale/

Sale, Arthur. (2007. The Patchwork Mandate.

D-Lib Magazine, 13 (1/2).

<http://dx.doi.org/10.1045/january2007-sale>

Shanghai Jiao Tong University Academic Ranking of World Universities.

<http://ed.sjtu.edu.cn/ranking.htm>

Times Higher Education Supplement World University Rankings.

<http://www.thes.co.uk/worldrankings/>

University of California eScholarship Repository.

<http://repositories.cdlib.org/escholarship/>

Rosenstone, Steven J. 2004. University of Minnesota President's Interdisciplinary Initiative on Arts and Humanities. College of Liberal Arts, Twin Cities Campus, Minneapolis, MN.

http://www1.umn.edu/pres/00_images/pdf/Abstract_7-15-04.pdf

University of Sydney eScholarship.

<http://escholarship.usyd.edu.au/>

Amsterdam University Press repository

http://www.aup.nl/do.php?a=show_visitor_repository&m=29&l=1

University of Pittsburgh Library. 2007. Pitt's Libraries and University Press Collaborate on Open Access to Press Titles. Press release, 29 November.

<http://www.library.pitt.edu/uls/news/pittpress.html>