

Content in Institutional Repositories: A Collection Management Issue

Dr Paul Genoni
Dept of Media and Information
Curtin University of Technology
p.genoni@curtin.edu.au

Abstract

Many libraries are facing the challenges to develop and manage an institutional repository. This article addresses the issue of content in repositories, and suggests that librarians need to approach the task of content development by applying some of the procedures and skills associated with collection management within more traditional environments. It also considers the types of content that might be suitable for institutional repositories, and notes that several recent Australian reports have recommended the need for a more standardised and regulated approach to the content of institutional repositories. It is argued that this is inappropriate.

Introduction

Recently a significant interest has developed in the concept of institutional repositories as a means by which libraries can support the communication of the research output of universities and other research organisations. This interest has been fed by a realisation that institutional repositories can not only improve access to research related information, but at the same time they can serve to extend the influence of a library within a university (Carver, 2003). The enthusiasm felt for institutional repositories is expressed in article titles such as 'Institutional repositories: the library's new role' (Chang, 2003), and 'Libraries in the lead: the institutional repository phenomenon' (Smith, 2004).

The focus to date has frequently been on the technical and structural matters related to building institutional repositories. This includes issues relating to the selection of software (Day, 2003); the application of metadata to items placed in repositories (Medeiros, 2003); copyright in open access environments (Bide, 2002), and the role that repositories might play in furthering reform to scholarly publishing by supporting the move towards open access archives (Prosser, 2003). As the Online Computer Library Center (OCLC) has recently pointed out, however, the technical issues involved in creating institutional repositories are not necessarily difficult, and the developers of a repository will more likely face challenges related to the politics and culture of an institution.

The most significant challenge facing academic libraries undertaking these institutional repository projects is not technical... The major challenge is cultural. Too few initiatives include all the stakeholders—faculty, library staff, IT staff and instructional designers—and there is no common view of what an institutional repository is, what it contains and what its governance structure should be. (OCLC, 2003)

OCLC are also right to point to the largely unresolved matter of content in institutional repositories. Content related issues will be crucial to the success of these repositories, be they in universities or some other institutional setting. To date, however, there has been remarkably little consideration given to the issue of exactly what type of material might be suitable for inclusion in repositories; who should be responsible for selecting this material; or how the task of content development for repositories might relate to other content selection responsibilities managed by a library.

Content in institutional repositories

The precedents for institutional repositories are found in the subject-based repositories that became popular – particularly in the sciences – during the 1990's. These repositories were based, at least loosely, on the model endowed by the creators of ArXiv.org (<http://arXiv.org/>), a repository based at Cornell University and offering content derived from physics, mathematics, computer science and quantitative biology. There are now numerous other important e-print services supporting research and scholarly communication in a number of disciplines, including the Chemistry Preprint Server (<http://www.chemweb.com/preprint?url=/CPS>), the Philosophy of Science Archive (<http://philsci-archive.pitt.edu/>), and Research Papers in Economics (<http://repec.org/>).

These subject repositories were created as an attempt to use the power of the Internet to provide an alternative and cheaper form of access to research literature. For the most part they require that for inclusion articles should have either been published in a peer reviewed

outlet, or to have successfully concluded the peer review process and be awaiting publication (Garner et al, 2001). Subject repositories have therefore evolved as primarily as archives of published material united by discipline, although at least some of them also encourage the archiving of pre-refereed drafts

Institutional repositories have been described as having four key attributes:

- Institutionally defined
- Scholarly
- Cumulative and perpetual
- Open and interoperable. (Johnson, 2002)

It is the first of these attributes, that content will be limited to the output of one institution, which distinguishes institutional repositories from their subject based predecessors. For many libraries in the process of developing institutional repositories, it is apparent that the criteria set for inclusion in subject repositories will not comfortably 'scale-down' to the institutional level. That is, there is an understanding that documents relevant to the output of a single research-active institution are much more diverse than those intended for formal peer review.

The Coalition for Networked Information has recently warned of the dangers inherent in simply transforming 'a review process that is traditionally disciplinary in scope to one that is institutional. We need to be very careful about the boundaries between scholarly publishing and repositories' (CNI, 2003). There is a growing appreciation that content in institutional repositories needs to be more diverse than is appropriate for subject-based repositories, and that they should unite both formal and informal scholarly communication in a single archive (Robins, 2002; Genoni, Merrick and Wilson, 2004).

The Scholarly Publishing & Academic Research Coalition (SPARC), one of the major proponents of open access archives and institutional repositories, has also endorsed the need to take a broad view of the material that might possibly be included in repositories. SPARC has stressed that institutional repositories should not serve to simply duplicate the existing channels of formal scholarly publishing.

While surveys of faculty attitudes and perceptions reflect faculty-author concerns for the perpetuation of traditional scholarly journal publishing, they also indicate that faculty consider institutional repositories to be particularly well-suited for various types of gray literature and other fugitive and unpublished material. This material includes:

- preprints;
- working papers;
- theses and dissertations;
- research and technical reports;
- conference proceedings;
- departmental and research center newsletters and bulletins;
- papers in support of grant applications;
- status reports to funding agencies;
- committee reports and memoranda;
- statistical reports;
- technical documentation; and
- surveys.

Such gray literature forms a part of the informal scholarly communication process ... In some instances, an item may be followed by a formal publication. Often, however, that it is not the case and the material becomes difficult to identify and access, let alone preserve. Further, even when gray documents are

subsequently published, significant detail—for example, on research methods and experiment techniques—is frequently omitted. Thus, while peer-reviewed journals provide the principal venues for formal communication within scholarly communities, informal gray literature serves a valuable supplementary role. (SPARC, 2002)

Clifford Lynch is another who has argued for an open mind on the content to be included in institutional repositories. Lynch suggests that repositories need to be understood and developed as a new form of scholarly publishing, and he has called for libraries to use them to ‘digitally capture and preserve many of the events of campus life – symposia, performances, lectures’ (Lynch, 2003).

These calls to broaden the content of institutional repositories, to the point where they include many types of texts that would fall outside those categories of material previously collected by libraries, make considerable sense. One of the major changes to scholarly communication engendered by the universal uptake of ICTs to support research and publishing, has been a steady erosion of the distinction between formal and informal scholarly communications. As this distinction means increasingly less to researchers, libraries have an opportunity to enhance their role in the research and scholarly communication processes by expanding the range of content under their management.

Many libraries have already begun to recognise the value of informal scholarly communication. This is apparent in the numerous subject ‘portals’ that now form part of libraries’ information services, and which frequently include guides or links to non-formal as well as formal sources of information. Similarly, institutional repositories should not necessarily be configured as a surrogate form of distribution for formal or peer reviewed scholarship. Rather their full potential may only be realised if they are allowed to develop into a hybrid form, to be used for the dissemination of both formal and informal scholarly communication.

Recent Australian developments

Despite the calls for a broad approach to the selection of content for institutional repositories, there have also been suggestions that content should be controlled and regulated in some way. This approach has been evident in a series of recent Australian reports investigating aspects of the infrastructure required for research support. Each of these reports has emphasised the potential benefits of institutional repositories, but they are worth considering briefly for the manner in which they deal with issues related to content. In general they have been hesitant to address the matter of content, but when they have done so they have supported standardised approaches to content assessment and selection.

The Commonwealth Department of Education, Science and Training (DEST) Information Advisory Committee report, *Research Information Infrastructure for Australian Higher Education*, issued in November 2002, called for financial support for the development of institutional repositories. On the matter of content it recommended ‘identifying international standards for use with the contents of E-print repositories’ (Department of Education, Science and Training, 2002, 54).

Not only are such standards not available at present for institutional repositories, but they are almost certainly undesirable. These repositories are, after all, designed to serve the needs and interests of the institutions that support them, and their content should be developed with local requirements foremost. Just as academic libraries would not look to international

standards to determine the content of print collections, nor should they look to such standards to dictate the content of an institution-based repository. It may be that some form of acknowledged 'best practice' will emerge with regard to content selection, but each institution or library must be responsible for selecting material that suits its own needs.

A 2003 report by the DEST Higher Education Information Infrastructure Implementation Steering Committee identified institutional repositories as being essential for the development of the Australian research information infrastructure. The report recommended, *inter alia*, that 'a common repository infrastructure approach for the management, discovery and dissemination of Australian research information be supported'. While the report did not address issues of content in detail, it did recommend that an important early step in the development of repositories will be to 'identify the various categories of research materials that would be deposited in institutional repositories, and their characteristics' (Department of Education, Science and Training, 2003, 10). The precise meaning of this recommendation is unclear, but it does appear to be suggesting the need for a standardised approach to not only the general technical matters related to repositories, but also to issues of content.

The subsequent 2003 Houghton Report into the impact of the changing information and communication technology environment on Australian researchers addressed briefly the issue of content in repositories, and came down in favour of replicating the 'quality' standards which are applied to traditional forms of scholarly publishing. The report concluded that:

Perhaps the most important feature, given the strong adherence to peer review and widespread concerns about quality control, will be quality control of the material populating repositories and clear standardised means for identifying the level of quality control that each item meets. (143)

This conclusion seems, however, to be tenuously supported by the evidence in the Report itself. The evidence presented regarding researchers' opinions of the value of peer review is far less emphatic than implied in this recommendation, and is anyway gathered in quite a different context. That is, respondents were asked to address the issue of peer review within the context of formal scholarly publishing, and the Report's recommendations have transposed these results to a discussion about the content of institutional repositories. The mention of 'widespread concerns about quality control' is also surprising, as the concerns recorded in the Report's data were expressed in the context of the Internet generally, and not directed at the issue of material placed in repositories by, or with the agreement of, researchers.

Where the Houghton Report is on firm ground is in the call for 'clear standardised means for identifying the level of quality control that each item meets'. If repositories choose to include material of variable levels of authority, then this should be recorded in standardised metadata and made transparent to users. The content included in institutional repositories should be categorised and able to be searched accordingly. That is, if desired a user can restrict a search to peer-reviewed material or some other category, but this is a matter of regulating metadata rather than restricting content.

The recommendations contained in these reports leave open the possibility that libraries or some other body will be responsible for imposing standards related to the 'quality' of material placed in repositories. It is not clear if the standard suggested is as rigorous as peer review, or if some lesser standard is being implied. The need to adjudicate on or enforce such matters of quality is outside the traditional role played by libraries and is almost certainly unnecessary. On the contrary, the policing of quality of material loaded to institutional repositories should come from those who are most vulnerable – that is, the academic staff who elect to add material to those repositories. To insist on peer review of this material, or

some other form of regulated quality control, will risk limiting repositories to replicating content that finds an outlet in formal scholarly publications. Rather they should be encouraged to develop in a manner reflecting the new scholarly communication environment which increasingly blurs the distinctions between formal and informal communication.

It is of course open to those charged with developing an institutional repository to confine content to peer-reviewed material. But that is a collection management decision to be made with the needs of a particular institution in mind. It is important that institutions and their libraries resist attempts to impose external standards relating to content which may have the effect of unnecessarily limiting the categories of content that can be added to a repository.

Institutional repositories and collection management

Libraries and archives of course have long experience with developing and managing content, and many of the skills applied to print and other forms of digital collections will be transferable to the institutional repository environment. By building upon their existing collection management experience and skills and adapting them to the selection of content for institutional repositories, libraries will be able to reinforce their role in the institutions they serve.

The realm of professional practice referred to as ‘collection development’ – or more broadly ‘collection management’ – has already been through a series of transformations as a result of the revolution in information and communication technologies. Whereas it was once a point of contention as to whether the key principles of collection management could be applied to digital formats, it is now taken for granted that the realm of collection management incorporates content in whatever form it is acquired or provided by a library.

Libraries should now move to extend the ‘normalisation’ of digital content within standard collection management practices to include the content of institutional repositories. It is libraries, more than any other of the competing institutional interests, which possess the necessary experience in the selection, description, storage and management of information content.

There are a number of standard collection management practices that can easily be adapted to incorporate the content of institutional repositories. These include:

- Negotiation of content priorities with users.

Libraries have extensive experience in talking to user groups about their information needs as a part of the selection of material for inclusion in print and digital collections. This will be a crucial step in the selection of content for institutional repositories. Even given a general acceptance that repositories should include peer reviewed and non-peer reviewed material, there will still remain a number of issues to be agreed. Will the repository contain research material only, or will it be extended to teaching material? How will successive drafts of a document be accommodated? What will be the relationship between the repository and other university based archival services? Will content be duplicated in an institutional repository when it may well be available from another repository source (i.e. theses which might be available from an archive of digital theses)?

The categories of ‘users’ to be included in these negotiations may well be more extensive – at least initially - than was previously the case. It will include not only the academic

staff, but also the management of the university who will almost certainly have a view as to how the institution can best be represented by the content of the repository.

- Description and definition of priorities

Libraries have for a number of years been using collection development policies to describe their collection priorities and practices to their users and other libraries. Although originally used to describe print based collections they are now widely used to also describe access to digital information and collections. As such, they have become an essential part of the accountability process for many libraries. These collection development policy documents should be adapted to describe the decisions made regarding content and priorities for institutional repositories. This should include descriptions of the different categories of content, the priority that each of these categories is afforded, and the relationship between the material in the repository and other collections managed by the library.

- Evaluation of the performance of the collection

In most cases, once a collection has been established libraries have undertaken the measurement and evaluation in order to determine if the original collecting and service goals have been met. This evaluation can either be based on the collection itself (i.e. how many documents it contains and how this compares with other similar collections), or on how successfully the collection is meeting the needs of its users (i.e. how much use the collection is receiving, and what levels of satisfaction can be measured amongst the user group). Issues of use and performance will remain as important with institutional repositories in order that their continued development may be properly planned.

This may entail traditional methods of evaluation (for example identifying which groups of users are accessing the collection, and to what degree different categories of content are being used), although it will entail a very different understanding of both 'user' and 'use'. One of the key features of a repository is that it should be 'open and interoperable'. This means that whereas library collections have traditionally been developed by collecting externally published material for the benefit of a user group internal to the collecting institution, in an institutional repository the production of content is internal to the institution, and it is likely that most of the use will be from external sources. This may make some forms of user-based qualitative evaluation more difficult, but libraries and their institutions will remain interested in knowing who is using the content of a repository and what categories of content are being used.

- Decisions relating to access

The management of collections has required libraries to make decisions about who has access and under what circumstances. Despite the requirement for repositories to support open access, there may still need to be decisions made as to some categories of content to which unrestricted access cannot be given. This may include, for example; draft versions of documents that are available in a later, more complete version; content that might have temporary copyright restrictions; or items which may have restricted access due to reasons relating to likelihood of causing offence or cultural sensitivity. Again, libraries have long been called upon to manage these issues as part of their collection management practices, although the institutional repository environment will create a new range of issues with regard to management of access.

- Management of conservation and preservation

The collection development practices of libraries have long assumed responsibility for proper conservation and preservation practices, and relevant priorities and programs have been recorded in collection development policies. As described previously, another of the key attributes of a repository is that the content should be 'cumulative and perpetual'. This therefore requires careful management, as whereas previously some content would be discarded at the end of its useful life, libraries will now assume responsibility for a greater proportion of their collection to be managed indefinitely.

While most of the developments associated with content in institutional repositories can be comfortably incorporated within the collection management program of a library, there is one area of practice that will be altered substantially, that of acquisitions. Whether dealing with print or digitally formatted material, libraries have previously identified wanted items, and then dealt with a publisher or supplier in order to acquire the item by paying for permanent retention or licensed access. With institutional repositories, the producers (authors) of material are requested to provide digital copies without financial compensation. Most repositories will be built upon a system of 'self-archiving', with the producer being responsible for adding an item to a repository and creating the associated metadata. It will probably be unwise (and almost certainly unfeasible) for institutions to attempt to mandate self-archiving of relevant material. Therefore the onus will be on libraries to convince producers of the benefit of archiving their research output, sometimes in successive versions. The process of 'acquisitions' in the context of institutional repositories is therefore transformed into the twin roles of promoting the benefits to be achieved by contributing to a repository, and educating contributors as to when and how this should be undertaken.

The negotiations regarding content suitable for inclusion in institutional repositories will require librarians to become more familiar with the full life cycle of scholarly research, communication and publishing. Hitherto the emphasis of library collections has been almost entirely in favour of formally published artefacts, principally the scholarly journal and monograph, and librarians have developed a good understanding of the role these publications play in the process of scholarship. This has included being able to make sophisticated selection decisions based on judgements about the presence or absence of peer-review; the extent to which a title is indexed; and measures of a journal's 'impact factor'.

Librarians have dealt far less frequently, however, with the categories of informal communication, the 'gray (sic.) literature' as referred to above by SPARC, that will populate institutional repositories. Although it precedes formal (peer-reviewed) publication, this grey literature is nonetheless an essential part of scholarly communication, often presenting research data in a more timely and detailed manner than is possible in formal publications. Indeed well organised and freely accessible institutional repositories will potentially transform the concept of grey literature. As Leah Halliday has argued:

The electronic environment makes these [informal communication formats] more accessible, thus moving them further along a scholarly publication continuum. Formats that are 'grey' in a print environment may, in an electronic environment, fulfil some of the functions of a scholarly publication more effectively than the archetypal formats. (Halliday, 2001)

And it is institutional repositories that are best placed to bring about this overdue reassessment of the status of grey literature. They are both more receptive to its inclusion than are the subject based repositories, and far better equipped to provide the necessary levels of indexing access and long term preservation than can be achieved by simply making the material available on the Internet from some other source.

If libraries are to build repositories that incorporate informal scholarly communication, then they will need to develop an improved understanding of how this material contributes to the processes of research and scholarly communication. To put it simply, you can't collect what you don't understand. Indeed it is a present lack of understanding of the role of these primary research documents and their relationship to formal scholarly publishing which may have led some librarians to be tentative in their approach to including this material as content for institutional repositories.

A particular challenge libraries face in assuming the lead role in the development of institutional repositories and their normalisation within collection management programs will be that of providing adequate staffing support. Many collection management programs are already struggling under the dual burden of managing print and electronic collections. Only the largest and most well funded libraries may be able to absorb an additional workload within their current resources. It may therefore be necessary for library managers to persuade their funding bodies that institutional repositories are a crucial means by which an institution can raise its profile. Although it may be difficult, librarians should be aggressive in lobbying for project funding that should deliver important benefits to a university or other institution. Universities are obviously well aware of the prestige associated with large library collections, an active research profile and a vibrant teaching culture. They should therefore be open to persuasion that an institutional repository provides an irresistible opportunity to showcase their achievements in each of these areas.

Conclusion

A recent report on the establishment of an institutional repository at the University of Glasgow noted that within the university 'the library is the standard bearer for the advent and implementation of e-print archives and open archives services' (Nixon, 2002). The report also stated that the 'Library has a number of distinct roles beyond its technical provision and maintenance'. These roles were described as;

- Encouraging members of the University to deposit material into the ePrints archives...
- Providing advice to members of the University about copyright and journal embargo policies for material which they would like to deposit in the archive...
- Converting material to a suitable format such as HTML or PDF for import into the archive
- Depositing material directly on behalf of members of the University who do not, or cannot self-archive their material
- Reviewing the metadata of content which has been self-archived to maintain the quality of the record. (Nixon, 2002)

These are all roles one would expect the library to play. There is, however, no indication that the Library is actively engaged in negotiating, selecting or documenting the priority to be given to content in the repository.

For libraries developing institutional repositories the issue of content should be foremost. They are, after all, simply another form of collection, which should be subject to established levels of decision making with regard to collecting priorities, and the same level of management with regard to associated matters, such as access and preservation. So familiar are these tasks that they should be able to be smoothly incorporated into the current collection

management programs of libraries, and the associated policy decisions can be recorded within the framework of existing collection development policy documents

References

Bide, M. (2002), *Open archives and intellectual property: incompatible world views?* Bath: Open Access forum, available at: http://www.oaforum.org/otherfiles/oaf_d42_cser1_bide.pdf (accessed 19 February 2004)

Carver, B. (2003), 'Creating institutional repositories: a role for libraries', *Ex Libris*, No 181, available at: <http://marylaine.com/exlibris/xlib181.html> (accessed 23 February 2004).

Chang, S. (2003), 'Institutional repositories: the library's new role', *OCLC Systems and Services*, Vol. 19 No. 3, pp. 77-79.

Coalition for Networked Information (2003), *Summary Report of the December 8, 2003 CNI Executive Roundtable on Institutional Repositories*, available at: <http://www.cni.org/projects/execroundtable/fall2003summary.html> (accessed 22 February 2004).

Day, M. (2003). *Prospects for institutional e-print repositories in the United Kingdom*, available at: <http://www.rdn.ac.uk/projects/eprints-uk/docs/studies/impact/> (accessed 22 February 2004)

Department of Education, Science and Training (2002). *Research Information Infrastructure Framework for Australian Higher Education*. Department of Education Science and training, Canberra, available at: <http://www.dest.gov.au/highered/otherpub/heiiaac/report.pdf> (accessed 19 February 2004)

Department of Education, Science and Training (2003). *Report of the Higher Education Information Infrastructure Implementation Steering Committee*. Department of Education Science and training, Canberra, available at: <http://www.dest.gov.au/highered/otherpub/heiisc03/report.pdf> (accessed 19 February 2004).

Garner, J., Horwood, L. and Sullivan, S. (2001), 'The place of eprints in scholarly information delivery', *Library Management*, Vol. 25 No. 4, pp. 250-256.

Halliday, L. (2001), 'Scholarly communication, scholarly publication and the status of emerging formats', *Information Research*, Vol. 6 No. 4, available at: <http://informationr.net/ir/6-4/paper111.html> (accessed Feb 10 2004).

Genoni, P., Merrick, H. and Willson, M. (2004), 'Virtual symposia: an investigation into scholarly communities online', paper presented at *Breaking Boundaries: Integration and Interoperability: VALA 2004, 12th Biennial Conference and Exhibition*, available at: <http://www.vala.org.au/vala2004/2004pprs/prgm2004.htm> (accessed 19 February 2004).

Halliday, L. (2001), 'Scholarly communication, scholarly publication and the status of emerging formats', *Information Research*, Vol. 6 No. 4, available at: <http://informationr.net/ir/6-4/paper111.html> (accessed 10 February 2004).

Houghton, J. W., Steele, C., and Henty, M. (2003), *Changing Research Practices in the Digital Information and Communication Environment*, Department of Education, Science and Training, Canberra.

Johnson, R. (2002), 'Institutional repositories: partnering with faculty to enhance scholarly communication', *D-Lib Magazine*, Vol. 8 No. 11, available at: <http://www.dlib.org/dlib/november02/johnson/11johnson.html> (accessed 23 February 2004).

Lynch, C. A. (2003), 'Institutional repositories: essential infrastructure for scholarship in the digital age', *ARL Bimonthly Report*, 226, available at: <http://www.arl.org/newsltr/226/ir.html> (accessed 23 February 2003).

Medeiros, N. (2003), 'E-prints, institutional archives, and metadata: disseminating scholarly literature to the masses', *OCLC Systems & Services*, Vol. 19 No. 2, pp. 51-53.

Nixon, W. (2002), 'The evolution of an institutional e-prints archive at the University of Glasgow', *Ariadne*, No. 32, available at: <http://www.ariadne.ac.uk/issue32/eprint-archives/> (accessed 19 February 2004).

OCLC (2003), 'OCLC Environmental Scan 2003: Research and Learning Landscape', available at: <http://www.oclc.org/membership/escan/research/repositories.htm> (accessed 24 February 2004)

Prosser, D. (2003), 'The next information revolution: can institutional repositories and self-archiving transform scholarly communications', *IATUL Proceedings*, Vol 13, available at: http://www.iatul.org/conference/proceedings/vol13/papers/PROSSER_fulltext.pdf (accessed 19 February 2004).

Robins, D. (2002), 'From virtual libraries to digital libraries: the role of digital libraries in information communities' in *Libraries, the Internet and Scholarship: Tolls and Trends Converging*, ed C. F. Thomas, Marcel Dekker, New York, pp. 45-76.

Smith, M. 2004, 'Libraries in the lead: the institutional repository phenomenon', paper presented at *Breaking Boundaries: Integration and Interoperability: VALA 2004, 12th Biennial Conference and Exhibition*, available at: <http://www.vala.org.au/vala2004/2004pprs/prgm2004.htm> (accessed 19 February 2004)

SPARC, 2002. *SPARC Institutional Repository Checklist & Resource Guide*, http://www.arl.org/sparc/IR/IR_Guide.html#management (accessed 23 February 2004).