

D-Lib Magazine
May/June 2007

Volume 13 Number 5/6

ISSN 1082-9873

Ten Major Issues in Providing a Repository Service in Australian Universities

[Margaret Henty](#)

Australian Partnership for Sustainable Repositories

<Margaret.Henty@apsr.edu.au>

Introduction

By mid 2006, all Australian universities had established, or were partway to establishing, institutional repository services. The development of institutional repository services can often be related to the open access movement, which seeks to make valued research outputs openly available by encouraging academics to place their publications into repositories, enhancing their availability and bypassing the high cost of journal subscriptions. However, many universities have extended the functionality of their repository services for other purposes, such as giving scholars the opportunity to develop their own research portfolio, providing a means of improving research reporting, establishing an electronic publishing service, or giving access to collections of images or other research outputs. The potential for development seems endless.

At the same time, university research increasingly involves the use, generation, manipulation, sharing and analysis of digital resources. The importance of what is generally called "eResearch" on the national agenda shows the need for improved data management and sustainability practices to support research over the longer term. This raises questions of the relationship between the repository and eResearch and provides challenges to repository managers to broaden their thinking still further to help meet these needs.

The Australian Partnership for Sustainable Repositories (APSR) was established in early 2004 with a focus on issues of access continuity and the sustainability of digital collections. In mid-2006, APSR began a series of interviews with senior university personnel who are responsible for the oversight of research, for the repository service or for research data management. This arose from APSR's need to better understand the higher education sector requirements for improving research data management and the information infrastructure that underpins it.

The purpose of this article is to identify the major issues that interviewees thought would be most significant for their repository services in the next five to ten years. While there are different views of the issues associated with the roles and responsibilities of repositories and research data management, this article only addresses the views of administrators. It does not claim to be exhaustive or

statistically based. Rather it aims to provide a summary of the ten major issues facing this particular group of repository services from the point of view of those who provide the policy setting and the funding for the services, or who have responsibility for them.

APSR has independently investigated the researcher viewpoint. Those results have been published as *Sustainability Issues for Australian Research Data: The report of the Australian e-Research Sustainability Survey Project*¹.

The interviews

Thirty-three people from fourteen universities were interviewed during the third quarter of 2006. Those involved were Deputy Vice-Chancellors (Research), Pro-Vice-Chancellors (Information) or University Librarians or any of their equivalents, with other personnel such as repository administrators, library staff or IT specialists on occasion. (The anonymous quotes interspersed throughout the rest of this article are just a few of those that were recorded during the interviews and later transcribed.)

The participating universities include Australia's eight largest research universities (known as the Group of Eight), partners and associates of APSR and ARROW (Australian Research Repositories Online),² and a small group of others who were geographically accessible and willing to participate. The South Australian Partnership for Advanced Computing (SAPAC) was also included, because of their expertise with research data management.

The repository environment

Any statement about institutional repositories in Australia is out of date the moment it is written. It is an area of rapid development, with occasional landmark surveys³ capturing snapshots at particular moments. The most recent assessment can be found in a survey by Joanna Richardson⁴ published in August 2006, which provides a comprehensive listing of repository software systems in use in each university. However, a repository is not, in itself, a service. The term repository is often associated with the use of a particular software and hardware platform. This article is more concerned with identifying the broader issues of the cultural, political and organisational landscape in which the service is offered. The subject of technology, however, cannot be ignored.

Before turning to the issues, it is worth providing some background as to why the various universities decided to develop their repository services in the first place. Some have been set up only to take print materials, with their development influenced largely by the open access movement. This idea has been taken up with enthusiasm by librarians dismayed by the ever-increasing cost of journals. It has not, however, found the same level of support among senior academic administrators or among the researchers who have the role of depositor.

As time has gone on, it has become apparent that attracting researchers to participate in this grand plan has been an uphill battle, leading to calls for the mandating of article deposit. This has been done successfully in one Australian university, the Queensland University of Technology, with resulting high levels of deposit. At the same time, some question whether open access is a sufficient basis for the repository, given its high cost and the fact that the materials are otherwise available.

I didn't want to go through all of that effort to put up something that was already available. Despite all the urgings of open access and all of those higher order things, to me it was just a question of resources. And also I would have to say I haven't been able really to engage more than a handful of academics around the place on open access.

Predictions that open access repositories would be taken up enthusiastically on a discipline basis do not seem to have come to anything, despite the ongoing success of the Physics "Eprint" Archive.⁵

Resonating more deeply with senior academic administrators has been the idea that the repository can serve as a showcase for their university's research output. This was mentioned by many as the prime motivating factor for setting up their repository service, which has been extended to include conference papers, theses, reports and other quality text-based research output. In some instances, the impetus for such a repository had come from an academic area but had found a home with a library willing to take on responsibility for the service.

A small number of universities have come to their repository through the need to control learning objects. Members of this group do not mix learning object repositories with research outputs, however, seeing the two as having quite different objectives and management needs. A complicating factor in the development of a repository service has been the forthcoming national Research Quality Framework (RQF) exercise.

Regardless of the reason for starting up the service, many now recognise that the repository as originally envisaged as a text-based e-print service is limiting. In such a rapidly developing area, the way ahead is not, however, completely clear. Those responsible for repository development have many questions and issues with which to deal. What are those issues? Here are the most commonly mentioned ten.

The issues

1. Roles and Responsibilities

The open access origin of many repositories has led to responsibility for the repository being held by the library in all but one of the universities surveyed. At the same time, research data is usually managed most often in the disciplinary area concerned, while the research office has its own system to manage the university's research reporting obligations.

All of the librarians spoken to seemed perfectly comfortable with taking responsibility for the repository service and for having a role in the management of research data.

...if not the library, who? There really is nobody else that I can identify in the university community who has probably the skills and necessarily the will to actually take this responsibility on. Libraries are in the business of describing and providing access to information, whether it's in a book or a journal or something like that. And at the end of the day, what we're talking about is exactly the same sort of service – but on behalf of a much more distributed community.

Governance, however, is a different matter, with some stressing the importance of the

involvement of the academic community, university administrators and those responsible for IT services. The academic administrators spoken to also seemed comfortable with the library running the repository service.

I think the university needs to address the issue of governance, and that obviously is related to management but is different in the sense that it would need to be broadly based. You would need to draw in different parts of the university who are making different contributions.

Responsibility for the long-term management of research data is ill-defined in all of the universities surveyed. This is not to suggest that data is ill-managed. However, none had clear guidelines for administrative responsibility for data, leading to a situation where obligations for data retention may not always be met and longer-term access may not be possible. The general practice is that responsibility for research data rests with the discipline area involved. The advantage of this situation is that there is a strong disciplinary understanding of the data management issues involved.

Given the distribution of responsibility for different systems and services, it is not surprising that the lack of coordination between them was commented on.

One of the issues for the University as a whole is the issue of data custodianship, whether we're talking about scholarly information and knowledge or the data that the University uses to do its business – like HR data and student data and so on.... Wouldn't it be good if in the end we had a reasonably seamless system?

2. Sustaining the service

The issue of how to sustain the service over the longer term was mentioned by everyone interviewed. For the most part, the repository service had started out as a small scale operation, often funded by a special grant of some kind. The services have grown in size and scope, leading to questions of future resourcing. Many recognised the irony of trying to provide a long-term service on short-term funding.

At the moment here, for instance, it's all funded by the library more or less. We've got to get the university centrally to agree that the equipment is part of a central infrastructure.

For libraries, particularly, the issue of resourcing comes at a time when their role is changing, needing new economic models to underpin resource allocation. The demands of print do not seem to lessen, and subscription electronic services still require a high degree of management oversight. However, libraries are keen to expand their roles to include publishing, the digitisation of print materials, support for electronic course delivery and other activities.

Is the repository going to become our new core business? I would pose that as a semi-serious question. Perhaps not in the next five years, but in the next ten years we're going to see many, many libraries wrestling with the idea that there's all of this other stuff that people are far more interested in than the sorts of traditional library services we have.

3. Engaging the community

If the repository service is to be regarded as successful, it has to demonstrate its value

to the academic community. Most of those interviewed as part of this survey recognised that this is a new enterprise, but one which will not take off unless there is something in it for everyone using it, "meeting real needs" as one person put it. It is one thing to have the support of senior administrators who can see the practical value of providing access to research outputs and research resources, but it is another to convince people to go through the necessary steps to ensure their materials are deposited.

The proof of the pudding will be its value to the university, so that's the key issue. Are we creating something that is perceived as being valuable to the university?

The need to inform and engage was seen as paramount.

...if we stopped 100 of our researchers in the street and asked them what a digital repository was, what was its purpose and how would you access it, and how would you use it and is it important to you, then you'd get a very wide range of opinions. But you get a lot of people going, "a digital what?"

One way to engage the community is to mandate the repository's use. This might be effective, but the consequences might not be wholly desirable. Could the repository service cope with the ensuing demand? Would this create ill-feeling? Would researchers actually comply? And if not, what then?

We also were surprised in science... They suggested it should be mandatory, and we thought boy, if the library had said that, they'd have had us up against the wall. But when they suggest it...

If mandating does not seem practicable, there is always persuasion, but this is not necessarily the easier route.

I certainly think if mandatory deposit doesn't kind of become generalised, that repositories aren't sustainable, because it takes a lot of work to get people self-archiving. [But] once they're self-archiving, they'll keep on doing it.

4. Guaranteeing quality of service

Not only does the repository service have to be of value, it has to be of high quality. Quality means that the service needs to be responsive to researcher needs, effective in its delivery and adaptable to change. Particular concerns mentioned were efficiency, as measured by timeliness of response; the capacity to grow and develop to match demand; and the need for better tools and technologies to improve access and delivery. Interoperability and integration of services are also key requirements.

We have to be able to deliver what we say we can. We can't offer the world and offer every solution to everyone when we actually can't do that. And if we try to do that, we're going to come unstuck. And so I think we have to deliver and be credible, as well as position ourselves to be able to extend as things grow.

Part of the problem here relates to inadequacies in the technological base of the repository service. There are barriers between the depositing researcher and the

repository itself, so every extra step to be taken involves extra effort and the likelihood that the depositor will not bother. One word describes what everybody said they want: "seamlessness". Two particular developments were singled out, although doubtless there are many more. One is the Scholar's Workbench, currently being developed at the Australian National University,⁶ a tool that will enable documents created in standard word processing applications to convert readily to other formats and provide a direct link to the repository for deposit. The other is the need to develop document management systems for email to be easily maintained as a record of a particular research project.

How do we put in the tools, the standards, the processes and approaches, and the underpinning technologies to kind of make that happen seamlessly so it is easier for the researcher? And they don't know what's happening behind the scenes. They just know it's easy and we're doing what we need to do to contribute to this.

5. Defining the collection

All those interviewed agreed that there is a need for a strong policy framework to define the role of the repository service and the scope of the collection. In practice, few universities have a written policy, but all recognise the need to have one, especially if they have leapt into providing a repository service without thinking through the collecting implications. The need for a policy framework parallels the traditional practice of university libraries and archives in having a collection development policy, agreed by the user community and in line with the aims and objectives of the university. Some now find themselves having to refuse some categories of material: difficult to do without a policy that has been accepted by the wider university community.

Aligned to the issue of what to collect is the issue of what to keep for the longer term and what then to discard. If not discarded, material may be relegated to secondary storage, available only on request and possibly after some delay.

...first of all, one of the most pressing issues, I think, is definition of the scope of content, and that picks up that whole issue of research data. And if so, how do we define research data? How do we handle it? We're already, even at this early stage, having a little bit of difficulty with the scope of content, because on the one hand we have some people who are saying we should have no scope – we should have no boundary drawn, and basically anything anybody wants to put in should be okay. Why not? Then we have another group that are saying, 'No, no.' As a university repository, we have to have some control mechanism, particularly as this currently is an open repository.

6. Research reporting and compliance

Australian universities are required to report research activity to the Government and other funding bodies on a regular basis, and have acquired systems, whether in-house or proprietary, to help them with it. The proposed introduction of a Research Quality Framework (RQF) exercise, however, suggests that the universities will not only have to report research activity, but to provide access to published research outputs, impact statements and other information. It is here that the repository has an important role.

The imminent introduction of the RQF has served to justify, and hence to hasten, the introduction of a repository in some universities, while encouraging better communication between the research office and repository managers. The RQF has raised issues of whether data can be readily exchanged between the research management system and the repository, where responsibilities lie and how the activity can best be coordinated. These issues will be resolved during 2007, as the needs of the RQF are further defined, but there is no doubt that it has served to focus minds on the value of a repository.

At the time of the survey, details of the RQF were being keenly awaited because of its likely impact of processes and workloads.

I think there's a creative tension between what the university needs to be doing in order to comply with the RQF and some of the principles of the preservation of scholarly information. I think...the need to provide access to the research content in order to measure quality and impact sometimes appears to take priority over good preservation practice.

Aligned to the RQF is a proposed Accessibility Framework, designed to ensure that the published research outputs of Australian universities are readily available to all who might want to read them. This will be even more complex to implement, mainly because of the legal and regulatory framework.

Repository services also have a role to play in meeting requirements for the retention of data as specified by funding bodies such as the Australian Research Council (ARC) and the National Health and Medical Research Council. The extent to which these requirements are currently met is debatable, with some expressing confidence that the situation is in hand and others expressing considerable scepticism.

There's a standard requirement[...] that in many disciplines the data from which publications are drawn should be kept for a specified amount of time. Typically it's five years, and at [this University] in areas other than those where it's required to be kept for longer, there's an expectation that the researchers will maintain datasets that represent the primary source of information from which their publications have been drawn. On a few occasions we've actually audited various parts of the University to find out whether that's being followed, and by and large, the auditors didn't find any problems with that although they warned us about the vulnerability of the holdings because of backup problems and things like that.

So as you would know [...] the ARC has had a longstanding requirement in its guidelines about putting data into secondary repositories. Most people don't even know what they are, let alone comply with it – let's be candid. And the tools that were developed some years ago when that requirement first came in were not very friendly. The ARC never checked up on it, and most of our people didn't even know what it was that was being referred to in the guidelines. So it was a bit of a perfunctory sort of condition really.

7. The repository and eResearch

The term "eResearch" has been the buzzword of 2007 as its importance to the university has been increasingly recognised. Everyone interviewed was asked about

institutional preparedness for eResearch and the management of research data.

Just to put my cards on the table, I'd go one step further and say that of all of the modern descriptions of the world, the fact that we're flooded by data is true, and anything that can be done to help us deal with that flood of data should be done. We will find it increasingly difficult to operate the communities that we live in if we can't get control of data and turn it into information and knowledge, and the best way I know for doing that is to thin it out. Either throw away stuff that isn't important, which requires a very careful judgement, or to aggregate the data so that it produces high levels of information and knowledge. And that's what you do see in research publications or in monographs. The notion that the primary data should be kept forever doesn't make any sense at all, although it shapes people's views about things.

A recent APSR report⁷ has considered research data management in more detail, and it is not proposed to duplicate the findings of that report here. However, it was apparent that those interviewed were quick to identify four themes relating to eResearch and its implications. Firstly, there is the need consider the financial and economic implications of increased quantities of research inputs being created in electronic form, and the need for a sustainable basis for funding. Secondly, there needs to be significant collaboration with those areas involved the data creation and analysis, and new organisational structures designed to accommodate this. Thirdly, the technological infrastructure to support these activities is currently far from complete. And lastly, the skills available to support research data management are difficult to find, as might be expected in such a rapidly growing area.

It is only recently that the repository has been seen as one of a number of contributors to the management of research data. However, it was clear from the various responses to the question about eResearch, that the current relatively narrow conceptualisation of what a repository service might offer needs to be revised within the context of a broader university data management strategy. Given that most repositories are to be found in libraries, this presents particular challenges.

Another key issue is for us to get right the continuum through the various repository layers. So if there's a large data store sitting at the bottom layer of the university that's also a repository, how will that be managed? ... And then, how does it interoperate with or speak to or be part of a landscape of repositories which includes the [...] one which is the one I think you're most interested in. But although it's the one I'm managing, I don't actually think it's more or less important than the mass data store....So it's not the size of the store that's necessarily important. It's the use that needs to be made of the data. And we've always taken an information management perspective on this. The issue for us to resolve over the next five years is how to get the landscape of repositories working together.

There are a number of possible solutions to the issue of research data management, one of which may be the creation of a centralised data service to meet national needs.

I quite like the idea of some national support, but I'm not sure it's about infrastructure in boxes. I think it's more to do with setting standards and policies, and giving people the support through that, and also that because researchers work across institutions, the idea of some national

support there is a good one. But I don't think of it in terms of a great big supercomputing facility with lots of stuff.

8. Skills and staffing

Underlying any service are the people who make it all happen, and the skills they bring to bear can be critical to the success of the enterprise. Almost without exception, those engaged in supporting a repository service are learning on the job; at present in Australia, there is no course in repository management offered within the higher education or vocational education sectors. The range of skills required in managing and operating a repository can be extensive: requiring an understanding of the underlying technology, marketing, management, classification, legal framework, scholarly communication patterns and more. Many of those interviewed identified difficulties in finding appropriately qualified staff.

I think you need two types of skill base to manage this. One is the technical knowledge and the quality assurance for the actual system that both accepts the content and checks its bona fides, and then makes it available or manages the access. The other type of skill are the staff who'll actually provide information about the system – how to use it, encourage the use of it, troubleshoot the researchers' interface. And some of those skills are transferable from existing library operations.

9. Technology

Some of those interviewed identified a particular aspect of the technological basis of their repository as an issue. Some were emphatic that technology is not an issue, pointing out that technological challenges have solutions in ways that other issues, especially organisational and political issues, do not.

A major consideration for all of those offering, or planning, a repository service is selecting an appropriate software platform. Open source or proprietary? Which open source: which proprietary? The survey did not ask about the basis for any decision making here, but it was clear from discussion that open source software is seen by some as demanding a high level of in-house technological support, which is not always available in a library setting. Proprietary solutions have therefore been preferred. This has not always been the case, with several successful open source implementations in smaller libraries. Of interest in this context was the comment from one university librarian, that there has been a muddling of open source and open access, with the expectation that a commitment to one implies a commitment to the other.

Other aspects of technology that generated comment included the importance of interoperability (both between repository systems and with other systems), the critical importance of common standards, the quality of data input, the costs and quality of metadata to ensure discoverability, the capacity to scale the system to accommodate growing needs, questions of storage management (online, offline, data replication), ensuring a robust and reliable environment, and the need for middleware, especially relating to security and authentication.

Far better, I think, if we accept that there are going to be many different flavours of repository or many different ways of storing the stuff. But if we are all talking the same language and using the same interfaces and the same management protocols and the same type of security

infrastructure, at least we've got a chance of getting interoperability.

10. The regulatory environment: copyright and digital rights management

Copyright and the other legal and regulatory aspects of maintaining a repository service were mentioned by practically everyone interviewed. Considerable work has been done on copyright in association with the use of repositories to enhance the open access for research outputs, especially published articles, and more remains to be done.⁸ Aspects of copyright that were identified as issues were the cost of checking individual items to ensure no breach of copyright and ongoing difficulties in obtaining the relevant permissions from authors and publishers. One university librarian reported that the primary motivation for putting their image repository in place was to cut down on copyright infringement.

I think intellectual property and copyright, given our experiences so far with our adviser here, is a major issue.

Data ownership was specifically seen as an issue, indicating a need for further policy development within universities.

Data that you generate as an academic at University A and you move to University B – do you leave it all behind? Do you take it all with you? Do you create it in such a way that you can do both? Who really owns it? What if it's externally funded anyway?

The issue of privacy was also seen as important, ensuring that all privacy requirements are met, especially in situations where data is being accessed by teams located in different locations. This is a further indication of the need to provide safeguards while not inhibiting the generation of knowledge.

Preservation

Preservation is mentioned here as an eleventh major issue facing repositories. It was not included in the first ten for the simple reason that few of the interviewees mentioned it. Preservation is the elephant in the room: the unmentionable that remains a huge challenge if repositories are to provide sustainable collections.

And I think, there again, there are differences of emphasis, and it hasn't always been clear in these discussions if part of the purpose is for long-term archiving, and if so, what sorts of materials is that for and to what extent it is seen as just fairly short-term open access to more recent publications. And I think that's an important question that needs university minds to think through in relation to the various sorts of things which might go into a repository. And I think it's a very difficult question, and if you're going to take that on, then I think you need to understand what it is you're taking on and think about how you might actually engage in ensuring that long-term preservation.

Conclusion

This article identifies the issues relating to repository management that are seen as important by a group of senior academic administrators. These reflect to some degree the way in which repositories have developed in Australia, where for the most part they have been introduced for the worthy purpose of giving researchers a vehicle to

enhance the availability of their publications by making them available via open access. The time is now here for a broader definition of the repository service, as some are recognising. Reconceptualising the repository is a challenge. It remains to be seen whether this resolves the issues identified here or simply presents new ones.

Notes

1. Markus Buchhorn and Paul McNamara. 2006. *Sustainability Issues for Australian Research Data: the report of the Australian e-Research Sustainability Project*. Australian Partnership for Sustainable Repositories. Online at <<http://www.apsr.edu.au/aeres/>>.
2. <<http://www.arrow.edu.au>>.
3. Gerard van Westrienen and Clifford A. Lynch. 2005. Academic Institutional Repositories: Deployment Status in 13 Nations as of Mid 2005. *D-Lib Magazine* 11 (9). Available online at <[doi:10.1045/september2005-westrienen](https://doi.org/10.1045/september2005-westrienen)>.
4. Richardson, Joanna. Integration of open access repositories with research management systems. August 2006. Available online from <<http://www.caul.edu.au/surveys/OARintegration2006.doc>>. (Accessed 6 October 2006.)
5. See <<http://www.arxiv.org>>.
6. For a description of the project, see <<http://www.apsr.edu.au/word/>>.
7. Buchhorn and McNamara, op.cit.
8. See, for example: Fitzgerald, Brian et al., *Creating a legal framework for copyright management of open access within the Australian academic and research sector*. Oak Law Project Report No 1. Report for the Department of Education Science and Training. August 2006. Available online at <http://www.oaklaw.qut.edu.au/files/LawReport/OAK_Law_Report_v1.pdf>. (Accessed January 11, 2007.)

Copyright © 2007 Margaret Henty

[Top](#) | [Contents](#)
[Search](#) | [Author Index](#) | [Title Index](#) | [Back Issues](#)
[Previous Article](#) | [Next Article](#)
[Home](#) | [E-mail the Editor](#)

[D-Lib Magazine Access Terms and Conditions](#)

doi:10.1045/may2007-henty