Readiness and responsibility for managing research data: institutional perspectives

Margaret Henty and Danny Kingsley
Australian Partnership for Sustainable Repositories

Abstract

EResearch is more than a buzzword. The creation of a special term to represent new and different approaches to research in the digital environment has implications for the provision of the infrastructure to support it. There are three players in this arena: the researchers themselves, the administrators who have overall responsibility for developing and managing infrastructure, and the government as the primary funder of Australian universities.

APSR has recently surveyed select senior university administrators (Deputy-Vice-Chancellors, Research, Pro-Vice-Chancellors, Information and University Librarians or their equivalent) seeking their views on this changing landscape. This paper looks at their responses as they contemplate what eResearch means for their universities and their institutional readiness and capability to respond. It also looks at the development of institutional repositories in this environment and the challenges which they face.

Introduction

Research increasingly involves the use, generation, manipulation, sharing and analysis of digital resources. This increase in what is generally called eResearch has created the need for improved data management strategies and sustainability practices to support research in the longer term. The data generated by eResearch ranges from high-end grid computing requirements through to smaller sets of data of varying degrees of complexity, meaning there is a spectrum of needs to be met.

The research landscape as it affects both repositories and eResearch is a rapidly changing one. New technologies can enable new forms of collaboration, data gathering and analysis, and researchers are keen to exploit this potential. The government is keen to take advantage of research findings which might benefit the economy and Australia's reputation as a forward-thinking nation. Academic administrators, with an eye on the bottom line, are anxious to support research activity which will benefit their university's reputation and income.

The government's commitment to eResearch is evidenced by the establishment of an eResearch Coordinating Committee. Access to data features prominently in the philosophy behind many current policy decisions. "As a general statement of principle, the Government regards publicly funded research as a public good, and considers that researchers and other stakeholders should be able to discover what

research is occurring, and to gain access to research data" (eResearch Coordinating Committee, 2005, p23)

The Australian Partnership for Sustainable Repositories (APSR) was established in early 2004 to focus on issues of access continuity and the sustainability of digital collections. In mid-2006, APSR conducted two surveys designed to explore how Australian universities are meeting the needs of researchers engaged in eResearch.

The AERES Project, conducted by Markus Buchhorn and Paul McNamara of the Australian National University, was based primarily on the views of those actively engaged in research. It was designed to "survey the sustainability issues for data intensive research projects, including the capabilities and demands of research groups and institutions for the storage, access, and long-term management of research data". (Buchhorn & McNamara, 2006, p2).

A second and smaller study was conducted to complement and extend the first. It sought the views of academic administrators on their university's readiness to meet the challenges of eResearch data management and on issues associated with their repository service. Those interviewed included Deputy Vice-Chancellors (Research), Pro-Vice-Chancellors (Information), University Librarians or any of their equivalents in 14 universities including Australia's eight largest research universities (the Group of Eight), partners and associates of APSR and ARROW (Australian Research Repositories Online to the World), and a small group of others. The aim of this study was primarily to examine Australian institutional repositories, their aims and history, and issues to be faced in repository management in the foreseeable future. The study did include, however, a question about the university's response to and preparedness for meeting the needs of eResearch.

The findings

Despite the differences in emphasis of the two studies, they showed a high degree of agreement on research data management issues.

Policy

The need for appropriate research data management policies at both national and institutional levels is pressing, in particular the need to address issues of responsibility. There is some support for a national approach to data management, especially for long-term sustainability. However while those interviewed recognised the need to have a body responsible for ensuring that standards are defined and put into place, overall there were few who favoured the creation of a national body to oversee data management. Within institutions, the management of research data is currently being undertaken by discrete groups, with little indication of institutional coordination. This does not mean to say that the issue is not being addressed, as nearly all of those spoken to referred to cross-campus committees set up to explore all aspects of eResearch. Discipline areas are actively managing their own data in many places, with central IT areas and libraries also taking a role.

The library is already responsible for the university repository in all but one university surveyed. For the most part, librarians interviewed expressed a willingness to take

responsibility, or at the very least, to play an active role in the management of eResearch data.

> *[…] if somebody says it's an institutional job to start working on the data curation issue then which part of the institution is going to take that on? And I think the answer should in 100 per cent of cases be the institution's library. It may end up being that in 80 per cent of cases. It depends. It depends what the state of readiness is for the library to move into it.*

If indeed libraries are to take on this role, there will have to be significant collaboration with the areas involved in the data creation and analysis, and new organisational structures will need to be designed to accommodate this.

At a national level, we need policies which provide incentives to ensure that data is appropriately sustained. At present there are more disincentives than incentives to do this. Both the Australian Research Council (ARC) and National Health and Medical Research Council (NH&MRC) have regulations about the keeping of research data and these have recently been tightened. (ARC, 2005 and NH&MRC, 2004)

Copyright is a major area of policy deficiency, and all aspects of the legal and regulatory environment such as privacy and digital rights management need addressing. Some of this can only be done at a federal level with the remainder an institutional challenge. Considerable work has been done in the area of open access for text, but data management requires attention. Privacy concerns remain paramount for personal information of any kind. Data ownership is a major issue, especially where projects are being done by teams of researchers from more than one university and where researchers move from one university to another.

A major policy consideration for any university managing research data is the need to decide what to keep, what not to keep, what to discard and when to discard it. These are all decisions that require careful judgement.

Financial and economic considerations

There are considerable financial and economic implications in having increased quantities of research inputs being created in electronic form and the consequent need to ensure that this is properly managed and preserved. At present, researchers are funded to produce results and the creation of data is a by-product of this.

Research in Australia is primarily funded by ARC and the NH&MRC, neither of which has any responsibility for infrastructure funding. This puts the responsibility for infrastructure back on the government, which itself funds both bodies, and creates a tension between the funding of research and the funding of long-term data management. If funding for data management is to be increased, some wonder, will the funds be taken away from research activity?

One infrastructure solution, university repositories, have in many instances been set up using soft money. Many of the university librarians interviewed expressed their concern that repositories are not embedded in the university's funding model and

that there is no guarantee of future funding to sustain what is becoming a core service.

Technology

The technological infrastructure to support eResearch is far from complete. The infrastructure required covers a wide spectrum from the relatively small and unsophisticated to the large and complex.

Data management for eResearch is most often carried out in the discipline area concerned, even where it might tap into national resources facilities such as the grid. This means that the systems used to support eResearch are often peculiar to the discipline and are isolated from more generalised systems, At the same time there is considerable technical development going on to support new research, as is appropriate. Problems arise primarily when data ceases to be of immediate use and becomes of historical interest.

Those interviewed expressed major concern over: issues of interoperability within and between systems, the critical importance of common standards, the quality of data, the need for metadata to ensure discoverability, the capacity to scale systems to accommodate growing needs, questions of storage management, ensuring a robust and reliable environment and the need for middleware, especially relating to security and authentication.

APSR is looking at ways to address some of these concerns. One solution is development of the Digital Scholar's Workbench. This will allow the researcher using a word-processing application to utilise a web application to convert suitably structured documents into different formats for different purposes; for example, into archival quality DocBook XML for uploading into a digital repository, into XHTML for onscreen viewing or into PDF for printing or distribution. The documents must be written using a template. (Barnes, 2006)

Skills

Australia is beset by a crisis in the availability of all kinds of skills, including those in this rapidly developing area. It is an area curiously neglected by both the higher education and vocational education sectors, with no formal qualifications or courses available in this kind of data management. Everyone interviewed identified the need for well-qualified staff for this work and the difficulty of finding them. With the added need in many cases for people with disciplinary knowledge, there is then the issue of how such staff should be classified: as general or academic staff.

In the case of the repository

Readiness for eResearch takes on a special meaning in the context of the university repository service. Most repository services in Australian universities originated within the library as a means of supporting open access, whereby researchers have been encouraged to deposit copies of their publications, including journal articles, research reports, conference papers and theses. Some repositories have been extended to include other types of format including images and sound files. There is the potential for many other developments.

In the context of eResearch, the current relatively narrow conceptualisation of what a repository service might offer will need to be revised as part of a broader university data management strategy.

> *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 particular issues here for libraries in meeting the challenge of eResearch.

Conclusion

Despite increased interest on a governmental and institutional level in the management of research data arising from eResearch, there is a large void between current practice and what needs to be established and maintained. The issues range from esoteric to extremely practical.

Conceptually, institutional repositories can offer a solution to the problem of long term preservation of data, but for this to happen they will need to be recognised by the universities as integral to the university's data management processes and supported accordingly with ongoing funding. Funding in general needs to be increased, to allow for the development and purchase of appropriate equipment, and the training and employment of staff with the requisite skills.

The challenge lies ahead of policy makers, university administrators and funding bodies to recognise the need for change and to take responsibility for the problem.

References

Australian Research Council 2005, *Funding Agreement between the Commonwealth of Australia as represented by the Australian Research Council and {Organisation} regarding funding for Discovery Projects to commence in 2006*, Available at:http://www.arc.gov.au/funded_grants/management_fund_discov.htm

Barnes, I 2006, *Preservation of Word-Processing Documents*, APSR, Canberra

Buchhorn, M & McNamara, P 2006, *Sustainability Issues for Australian Research Data: The report of the Australian eResearch Sustainability Survey Project*, APSR, Canberra

eResearch Coordinating Committee 2005, *An e-Research Strategic Framework: Interim Report of the e-Research Coordinating Committee*, Available at: http://www.dest.gov.au/NR/rdonlyres/9526F5BC-B079-4903-AE0B-87140165F37F/10519/InterimReport1.pdf

National Health and Medical Research Council 2004, *2004 NHMRC Conditions of Award for Training Fellowships,* Available at: http://www7.health.gov.au/nhmrc/funding/funded/manage/policy/deeds.htm#a2