

# DEVELOPING A DIGITAL REPOSITORY FOR A HUMANITIES RESEARCH NETWORK: THE PIONEER PROJECT

**Toby Burrows**

*The Australian Research Council's Network for Early European Research (NEER) is a recent initiative aimed at broadening and deepening research in the field of medieval and early modern European studies. An integral part of the Network is its Digital Services Programme, which integrates a range of digital strategies designed to advance the Network's goals. Among these strategies is a project to develop a national discipline-based digital repository known as PioNEER. This paper looks at the issues that have arisen in the course of this project.*

## **Research Networks, Cyberinfrastructure and the Humanities**

Academic research is being changed fundamentally by two major imperatives. In the first place, it operates increasingly within an interdisciplinary and international framework. Research teams are tending more and more to be composed of temporary groupings of researchers from a range of disciplines, brought together to address a specific problem. The complexity and scale of these research problems require the assembling of expertise from different disciplinary perspectives as well as from different institutions and organisations.

Closely allied to this is a second trend: the growing centrality of information technologies to the ways in which research is done. The importance of information technologies in academic research has been highlighted by a number of recent initiatives and investigations. The Atkins Report to the US National Science Foundation in 2003 used the term 'cyberinfrastructure' to describe the way in which software platforms can be interlinked to manage the deluge of scientific data, enabling researchers around the world to tap into an international grid of digital research (Atkins *et al.* 2003). Among the numerous investigations that followed on from this report was that sponsored by the American Council of Learned Societies into cyberinfrastructure for the humanities and social sciences (American Council of Learned Societies 2006). In the United Kingdom, e-Science and e-Social Science programmes have been established to channel funding into the development of such infrastructures (Hey and Trefethen 2005). In Australia, the term e-Research has been used for similar initiatives.

In Australia, the main government funding body for research in higher education institutions—the Australian Research Council (ARC)—responded to these imperatives by establishing its Research Networks programme in 2004. The purpose of this programme is to build large-scale groups of researchers and encourage them to collaborate across institutional and disciplinary boundaries. It

goes beyond the ARC's existing support for smaller-scale collaboration, and aims to develop these linkages at a national and international level. In all, 24 research networks were funded under this programme, with a total of A\$42 million being allocated over a five-year period (2004–2009).

One of the key areas in which the ARC envisaged that the networks would be active was the development of shared information technologies and knowledge management tools, new databases, and new technologies for communication and interaction. These activities were seen as part of the essential infrastructure that would be needed to underpin collaborative research in a national setting.

### **The Network for Early European Research**

The Network for Early European Research (NEER) was one of only two ARC networks to be funded in the humanities. It is based at the University of Western Australia, where its executive and secretariat are located, but most of its academic activities (conferences, seminars and workshops) take place 3,000 kilometres away on the eastern side of Australia. The Network's structure is a mixture of individual researchers and institutional members. Its individual participants include researchers in most of Australia's 37 universities, ranging from eminent academics through to postgraduate students and early-career researchers. More than 300 individuals are currently listed as Network participants. Their research covers all aspects of the culture and history of Europe in the medieval and early modern period, extending up to the initial European contacts with Australia in the late eighteenth and early nineteenth centuries.

The Network has a range of institutional members including several of the larger Australian universities, such as Melbourne and Sydney, which are making a financial contribution to the Network. There are also a number of industry partners, including commercial publishers such as Brepols, ProQuest and the University of Western Australia Press; public collecting institutions such as the State Libraries of New South Wales and Victoria, and the Western Australian Maritime Museum; and community groups such as the Perth Medieval and Renaissance Group, Australians Studying Abroad, and the Woodside Valley Foundation.

The Network is organised around five main themes or broad research areas: Cultural Memory; Social Fabric; Intellectual Formations: Science, Medicine and the Environment; Religion and Spirituality; and Early European/Australasian Connections. Each of these has a leader whose role is to coordinate Network activities and communication between researchers with an interest in the specific research area. Many of the academic activities of the Network are focused around these five themes, with each area organising and supporting conferences, seminars, postgraduate advanced training workshops, and meetings to develop a research agenda and develop collaborative grant applications. The Network has also funded a total of 13 Research Clusters—smaller groupings of researchers

assembled to address specific research topics. The Network also has a range of specific strategies for encouraging participation by postgraduate students and early career researchers.

### Digital Services and the Network

In response to the ARC's emphasis on the use of information technologies within its Research Networks, the NEER has developed its own digital services agenda as an integral part of its activities. This programme brings together various activities in the digital arena, with three main goals: to provide resources for the Network's participants, to enable them to communicate more effectively with each other, and to promote the Network's research and achievements.

The Network's initiatives in promoting more effective communication among its participants are focused mainly on its collaborative Web service: NEER Confluence (<http://confluence.arts.uwa.edu.au/dashboard.action>). This service uses the Confluence software, developed by the Australian company Atlassian, to provide Network participants with a number of workspaces for discussion, collaborative writing, storage of documents and files, and dissemination of news and comments. Confluence is marketed as 'enterprise Wiki' software, which simply means Wiki-type software with various additional features, including the ability to control access to specific spaces and pages. Each NEER Research Cluster has a space on Confluence, for example, and can restrict viewing and editing of specific pages within that space as required. Among the other spaces in Confluence are one for postgraduates and early career researchers, and one for the Management Committee of the Network.

Providing resources involves two main strategies. The Network is working in partnership with two major commercial publishers of specialist databases to provide access for those Network participants whose institution does not have a subscription. CSA ProQuest is providing access to *Early English Books Online* while Brepols NV is providing access to a selected group of its full-text products. The main beneficiaries of this approach are researchers in smaller universities and regional institutions. The NEER has also established an internship programme with Brepols, under which selected Australian postgraduates will be able to work on the development of the *Europa Sacra* database.

At the same time, the NEER is developing its own resource discovery service for identifying Early European artefacts, artworks and manuscripts in Australian collections. This will involve federated searching and browsing across the records of different types of institution with relevant collections: libraries, museums, archives, and galleries. Where digital versions of these objects are already available, the Network will enable researchers to find and view them.

As far as possible, NEER will avoid duplicating work already done in the cultural heritage institutions themselves, and will reuse metadata from their sites. The same principle applies to the digital objects identified through the resource discovery service. The Network prefers to point to a file on the server of the

appropriate institution, and avoids creating or storing its own digital files. The Network will, however, be aiming to add value to these resources through such techniques as collaborative annotation and semantic ontology-based frameworks. Early European research is a difficult area for metadata because of the many European languages used in the original sources and in contemporary scholarship, and the lack of consistent terminology in some fields. Mapping variant forms of names is a particular challenge. As far as possible, the Network intends that its work on metadata and subject vocabularies should link into the broader framework of the Semantic Web.

The Network is also involved in electronic journal publication. The refereed journal *Parergon*, which is published by the Australian and New Zealand Association of Medieval and Early Modern Studies, is issued in electronic form through Project Muse. The NEER provides publishing and subscription management services for this electronic version of *Parergon*.

### **Building a Research Repository**

The other major goal for NEER's digital services programme is to promote and disseminate its research and achievements. Its main priority in this area is to develop its own digital repository of research outputs. Digital research repositories have been the focus of a great deal of activity over the past five years or so, but this has mostly concentrated on institutional repositories. Initially, at least, institutional repositories emerged from the e-prints movement as a way for individual academic institutions to make their research output available freely to the scholarly community. One of their key attributes was considered that they should be open and interoperable. The assumption was that researchers would be prepared to deposit copies of their publications in this kind of institutional repository, and that self-archiving would become a widespread practice.

This has proved not to be the case, and most institutional repositories have found it hard to attract content voluntarily. A recent British survey found that fewer than 2,000 of more than 58,000 papers published by British researchers in 2004 had been deposited in institutional archives, leading the author to the conclusion that 'self-archiving . . . appears to be having problems' (Wilson 2006). This is borne out by figures from other countries, including Australia (van Westrienen and Lynch 2005). The only notable exceptions are institutions such as the Queensland University of Technology where depositing copies of papers has been made mandatory. It remains to be seen whether this situation will be significantly affected by the recent declarations of bodies such as Research Councils UK and the Wellcome Trust, which require the open archiving of papers resulting from research funded by them. Before the recent focus on institutional repositories, a small number of successful discipline-based e-print archives were developed, including the well-known ArXiv service for physics. These served mainly as an informal method for rapid dissemination of new work and were not normally linked to a specific disciplinary body or organisation.

The NEER's aim is to build a national research repository that is linked specifically to the Research Network as the sponsoring body, and which reflects the disciplinary areas covered by NEER participants. The primary aim of this repository is not so much rapid dissemination or communication within the Network; those requirements are being addressed by the Confluence service. Instead, the NEER repository—known as PioNEER—will provide a record of the research output of NEER participants, as well as a body of retrospective material previously produced by this group of researchers. PioNEER will serve to promote and record the work of NEER, both as a formal body and a series of research groups, and as a collection of individual researchers.

The PioNEER repository is being developed in partnership with the University of Western Australia Library. The Library is responsible for installing and managing the hardware and software, while the NEER is responsible for acquiring and loading content, determining standards, monitoring quality, handling copyright issues, and setting access permissions. The implementation project is being managed by the Library, with the NEER providing advice and input on configuration and design issues. The project plan envisages the repository going live in the third quarter of 2007. The initial Memorandum of Understanding between the Library and NEER will run until the end of 2009.

### **Issues Arising**

The PioNEER project is something of a hybrid as far as repositories go. It is not an institutional repository, but it is not a self-organising, discipline-based archive either. It will contain outputs from an Australia-wide 'virtual organisation' that has been called into being by a government grant and has a financial and administrative structure, but no independent legal existence outside that grant. The issues that have arisen in the planning and development stages of the PioNEER project are, as a result, partly those derived from the unique circumstances of this project and partly those already identified as common to institutional repositories (Henty 2007).

### *Choice of Software*

The University of Western Australia did not already have an institutional repository at the start of the PioNEER project, which meant that selection, acquisition, installation and configuration of suitable software was one of the project's first major tasks. Various software packages were evaluated jointly by the Library and the NEER against a detailed list of functional requirements. One of the critical considerations in this evaluation process was the relative merits of Open Source software and commercial products, and their respective fit with the Library's management and support structures for information technology services.

The software selected for PioNEER was a commercial product. DigiTool is a digital asset management system from Ex Libris, which is also being used at various Australian and overseas institutions, including the University of Melbourne, Curtin University of Technology, and the Australian Institute of Aboriginal and Torres Strait Islander Studies.

### *Relationship to Existing Repositories*

Because PioNEER is a national, discipline-linked repository, its relationship to existing institutional repositories is an important issue for the project. If Australian universities had already made extensive content available through their institutional repositories, PioNEER would be mainly focused on developing a collection of links to relevant digital objects already openly available. About 65,000 records from Australian university repositories are included in the ARROW Discovery Service, including 13,500 digital theses. But relatively little of this material appears to fall within the NEER's area of interest.

To put this another way, very little of the output of the NEER's participating researchers is known to be freely available in digital form. The only significant exception to this seems to be the Flinders Academic Commons at the Flinders University of South Australia, which contains articles and book reviews by several staff members who are participants in the NEER (Phiddian 2006). PioNEER will eventually contain metadata for, and pointers to, items in the Flinders Academic Commons, for the sake of completeness. A similar approach will be applied for any relevant items that are openly available in other repositories, including theses available through the Australian Digital Thesis programme.

The major priority for Australian institutional repositories over the next 12–18 months is likely to be the Research Quality Framework (RQF). This is a national research assessment process that will take place in 2008 and will include an evaluation of selected publications. Research groups nominated for assessment will need to submit a full list of research output from the six-year assessment period, and each researcher will need to nominate their 'best four' research outputs for the same period. Institutional repositories will be used to make outputs available for the RQF, and the Federal Government has recently allocated funds for establishing or updating repositories under the Australian Scheme for Higher Education Repositories. But it is not yet known whether the publisher's version of an output will be prescribed, and whether repositories can simply link to the appropriate publisher's Web service (Cooke 2007).

From the point of the view of the PioNEER project, the RQF may make some relevant material available—but this is unlikely to be substantial, and may not result in open access. PioNEER itself is unlikely to contribute directly to the RQF process, although it may make some digital objects freely available that will be included in the RQF evaluation.

### *Sustainability*

The NEER itself has a finite lifespan, with the current funding extending only until mid-2009. The Memorandum of Understanding between the NEER and the University of Western Australia Library will expire at the end of 2009. No commitment has yet been made to ensure the preservation and availability of the PioNEER repository beyond that date. The Library will also be developing and maintaining the DigiTool installation as a repository for the University of Western Australia, however, so the software itself is likely to remain available.

Avenues and options for continuing and maintaining the services and activities of the NEER beyond the current end-date are currently under discussion. The expectation is that, by the beginning of 2009, decisions will have been made about the continuation of these services, including the PioNEER repository. Ownership of the digital objects in the repository after 2009 will be one of the elements to be addressed as part of this process.

### *Roles and Responsibilities*

It is abundantly clear from a range of studies and practical experiences that most academic researchers are reluctant to deposit their research outputs in institutional repositories (van Westrienen and Lynch 2005). A recent analysis of the situation in Australia found that two-thirds of the 21 Australian university repositories contained fewer than 1,000 items (Kingsley 2007). There are various reasons for this, including lack of awareness, concerns about copyright and quality, and lack of time.

It is also clear from informal discussions with NEER participants that this reluctance extends to the discipline-based repository planned for the Network. While researchers can see the merits of having a single location where they can get access to the output of their colleagues—as well as promoting their own work—the possible benefits are generally outweighed by the time and effort required to collect and deposit material.

NEER's initial approach is for its own staff to identify, collect and deposit as much material as possible. NEER staff will liaise with researchers to obtain publication lists from them or from their web sites, and will work through the processes of obtaining electronic copies, checking copyright, creating metadata, and submitting the items to the repository. Access to the repository for depositing their own digital objects will also be made available to NEER researchers, but the primary responsibility for building the PioNEER collection will rest with the NEER staff, at least initially.

The NEER's expectation is that, once a sufficient corpus of material is available in the repository, researchers will see the value of contributing their research output to the repository as part of their continuing communication with fellow researchers in their discipline. By embedding the repository into the existing pattern of disciplinary communication, the Network aims to provide a

sufficient incentive for researchers to participate, and to encourage the cultural change required for such approaches to be successful.

### *Scope and Content*

The NEER's initial goal in populating the PioNEER repository is to provide a substantial proportion of the research publications of a representative selection of Network participants. The total output of the NEER participants runs to thousands of publications, going back to the 1960s at least. Because items are initially being added to the repository by NEER staff, rather than being deposited by the authors, an order of priority has had to be developed. To do this, the list of participants was analysed against various criteria, including the extent of their involvement in the Network, their inclusion in the initial list of 50 key participants submitted with the original funding application, and their track record in obtaining competitive grants. The initial selection of material is also designed to provide a balance between established researchers and early career researchers, as well as a suitable gender balance and a cross-section of different Australian universities. Three lists of 20 researchers each have resulted from this process, and are being used for the initial selection and loading of material to the repository.

Publication lists for many participants had already been collected as part of the Network registration process and have also been recorded in their 'personal spaces' in Confluence. These lists are initially being checked for readily available digital copies. Reuse of the publisher's digital copies is the preferred option (where publishers' policies permit), followed by use of the author's personal digital copies. Some digitisation from print copies is planned, but only as a last resort.

The aim is to give a rounded picture of the careers of these researchers, within the limits imposed by time and resources. While the initial focus is on published articles, the coverage of the repository is not intended to be limited to articles and papers. Monographs and theses will be included if they are available in digital form, and researchers will be encouraged to deposit the underlying research data from these studies, whether in the form of databases, spreadsheets, correspondence, images, sound files, maps, or other formats.

### **Copyright and Permissions**

Initially most publishers were suspicious of institutional repositories and tried to prevent authors from depositing articles in them. But this situation quickly changed, to such an extent that many publishers now permit this kind of open archiving. Of the 155 publishers in the SHERPA/RoMEO database, 78% formally allow some kind of self-archiving. An increasing number of publishers, including the proponents of Open Choice approaches, also make their own copies of articles freely available at a defined interval after publication—usually 6 or 12 months.



The NEER is working closely with authors and publishers to ensure that items are deposited in PioNEER with appropriate permission from the copyright owners. Immediacy is not the over-riding priority for PioNEER, however, so embargos imposed by publishers are not a major concern.

## **Metadata and Access**

Within the PioNEER repository, the Network is focusing particularly on metadata content and standards. There are various metadata schemas available for use in repositories, ranging from the complexities of MARCXML to the simplicities of Dublin Core. Among Australian repositories, the ARROW project has developed schemas based on MARCXML for six major types of research output, while others have stuck with Enhanced Dublin Core. Following the example of the Fez implementation at the University of Queensland, however, the NEER is adopting MODS as its metadata schema. MODS will meet the NEER's need for a considerable level of granularity without requiring the full complexity of MARC. It will also enable all types of material to be covered with a single schema.

There are several pathways for browsing the PioNEER repository, reflecting the nature of the NEER and the way it is organised. Browsing by author or institution will be supplemented with browsing across the NEER's five Research Themes and its 13 Research Clusters. The NEER is also testing different subject approaches. These include tagging by NEER participants (not just by authors) and ontology-based approaches.

## **Interoperability**

The NEER is very interested in exploring how the existing repository framework might be used to help in transforming the ways in which research is communicated and disseminated. At present, repositories are a mainly static service. They collect and archive research papers and articles and make them available to other researchers. The immediacy and availability of research results are improved, but the effect of repositories on the research process itself is not necessarily transformative.

The NEER is aiming to test the integration of repositories with other emerging technologies in an effort to design new structures for communicating research in the humanities. In particular, it aims to encourage younger researchers to experiment with new approaches. A key element in this process will be interoperability between Confluence and PioNEER. The aim is to encourage links between work in progress (recorded in Confluence) and the finished products (stored in PioNEER).

Among the components of this structure will be blog-like narratives and discussions of research activities. These will be linked to research data and source materials, housed in databases, and similar structures assembled either by the research group or externally by other groups, including publishers. They will also

link to more formal publication of results in such structures as repositories, journals and monographs. This framework will be a record of both individual and collective activities within the national network of Early European researchers.

This kind of structure will open the workings of humanities research to a more continuing scrutiny than is possible in the traditional system of publication. It will also promote this research to a much wider audience across the Internet, through exposure to Google Scholar and through interlinking to external subject gateways and similar sites. A crucial element will be the incorporation of methods for enabling and recording peer review of the research. Providing avenues for other researchers to evaluate, comment on and respond to research will be critically important in a truly transformative use of the digital environment.

In these ways, the NEER is aiming to contribute to the design of 'a next generation system for scholarly communication' (Van de Sompel *et al.* 2004). The Network is able to work closely with a vigorous national community of researchers to design and test new ways of distributing and evaluating their communications.

## Conclusion

The main goal of the Network's digital strategy is to promote collaboration and communication between humanities researchers at a national, discipline-based level. This is being done by harnessing and integrating a wide range of information technologies. Some of these are well established and widely used, while others are newer approaches that are intended to transform the way in which research is communicated in the humanities.

The NEER offers a unique opportunity to develop and test technologies for scholarly communication and the dissemination of research, within the framework of a new government approach to funding academic research communities. The main emphasis is on transforming the way in which researchers communicate within a well-established humanities discipline, and on fostering the kind of cultural change that will be a necessary part of this transformation.

The PioNEER repository is a crucial element in this digital strategy. By building a national repository of research outputs for a research network in the humanities, PioNEER is testing the applicability of the repository concept in a setting other than a single institution or a self-archiving discipline. By linking the repository to other collaborative activities, the NEER is aiming to demonstrate the value of this approach to humanities researchers and to encourage their participation.

## REFERENCES

- AMERICAN COUNCIL OF LEARNED SOCIETIES, 2006. *Our cultural commonwealth: the report of the American Council of Learned Societies Commission on Cyberinfrastructure for the Humanities and Social Sciences*. New York: ACLS.

- ATKINS, D. E., *et al.*, 2003. *Revolutionizing science and engineering through cyberinfrastructure: Report of the National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure*. Arlington, VA: National Science Foundation.
- COOKE, A., 2007. Update on the implementation of the Research Quality Framework. *The RQF Explained: Information Management and Repository Needs for the RQF*. Melbourne. 15 February.
- HENTY, M., 2007. Ten major issues in providing a repository service in Australian universities. *D-Lib Magazine* 13(5/6). <<http://www.dlib.org/dlib/may07/henty/05henty.html>> (accessed 3 October 2007).
- HEY, T. and A. TREFETHEN, 2005. Cyberinfrastructure for e-Science. *Science* 308(5723): 817–21.
- KINGSLEY, D., 2007. *The one that got away?: Institutional reporting changes and open access in Australia*. Pre-print available. <<http://dspace.anu.edu.au/handle/1885/45158>> (accessed October 2007).
- PHIDDIAN, R., 2006. 'Who is the third who walks always beside you?' or how to engage humanities researchers in building digital repositories. *Digital Resources for the Humanities and the Arts 2006 Conference*. Dartington, UK. 3–6 September.
- VAN DE SOMPEL, H. *et al.*, 2004. Rethinking scholarly communication: building the system that scholars deserve. *DLib Magazine* 10(9). <<http://www.dlib.org/dlib/september04/vandesompel/09vandesompel.html>> (accessed October 2007).
- VAN WESTRIENEN, G. and C. A. LYNCH, 2005. Academic institutional repositories: deployment status in 13 nations as of mid 2005. *D-Lib Magazine* 11(9). <<http://dlib.anu.edu.au/dlib/september05/westrienen/09westrienen.html>> (accessed October 2007).
- WILSON, T., 2006. Institutional open archives: where are we now? *Library & Information Update* 5(4): 31–3.

**Dr Toby Burrows**, Digital Services Director, ARC Network for Early European Research, School of Humanities (M208), University of Western Australia, 35 Stirling Highway, Crawley WA 6009, Australia. Fax: +61 8 6488 1128. E-mail: [tburrows@library.uwa.edu.au](mailto:tburrows@library.uwa.edu.au)

Copyright of *New Review of Academic Librarianship* is the property of Routledge and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.