Building E-Research Infrastructures for Collaboration in Humanities Research Networks

Toby Burrows and Elzbieta Majocha
ARC Network for Early European Research, School of Humanities (M209), University of Western Australia, 35 Stirling Highway, Crawley WA 6009.

Abstract

The ARC Network for Early European Research (NEER) is funded under the Australian Research Council’s Research Networks programme. Its goal is to enhance the scale and focus of research in this multidisciplinary field, and to build collaborative and innovative approaches to the way research is planned and managed. One of NEER’s key strategies involves the use of digital technologies to promote communication among its 300 participants and to develop shared research resources.

This paper will look at two major digital services implemented by NEER: the Confluence collaborative software environment, and the PioNEER digital repository for research outputs and data. Confluence is Wiki-type software with a high degree of access control. PioNEER is unique in being a “virtual organization” repository, rather than an institutional repository or a subject-based archive. These services form the basis for a new framework for managing and communicating research in the humanities. The paper discusses the effects of these new technologies on a large national humanities research group.

Keywords: e-research, humanities, collaboration, networks

1 Introduction

Academic research is being fundamentally changed 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 organizations.

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 2003). Among the numerous investigations which followed on from this report was that sponsored by the American Council of Learned Societies into cyberinfrastructure for the humanities and social sciences (ACLS 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. The Australian Research Council (ARC), for example, recently funded several projects under its Special Research Initiative in e-Research Support.

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 AS$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 e-research infrastructure which would be needed to underpin collaborative research in a national setting.

2 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 (Trigg 2006). 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 39 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 like Brepols, ProQuest and the University of Western Australia Press; public collecting institutions like the State Libraries of New South Wales and Victoria, and the Western Australian Maritime Museum; and community groups like the Perth Medieval and Renaissance Group, Australians Studying Abroad, and the Woodside Valley Foundation.

The Network is organized 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 organizing 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 fourteen 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.

3 E-Research Services and the Research Network

In response to the ARC’s emphasis on the use of information technologies within its Research Networks, NEER has developed its own digital services agenda as an integral part of its activities (Burrows 2005). 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.

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 (EEBO) while Brepols NV is providing access to a selected group of its full-text services. The main beneficiaries of this approach are researchers in smaller universities and regional institutions. 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, 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 institutions 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. An important model for this service will be the MuseumFinland project (Hyvönen 2005).

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 (ANZAMEMS), is issued in electronic form through Project Muse. NEER provides publishing and subscription management services for this electronic version of Parergon.

To address its other goals – more effective communication between participants, and promotion of the Network’s research and achievements – the Network is relying on two major digital services: NEER Confluence (a collaborative software environment) and PioNEER (a digital repository of research outputs). The remainder of this paper is concerned with these services.

4 NEER Confluence: collaborative workspaces on the Web

In February 2007, NEER launched its collaborative workspace on the Web. NEER Confluence (http://confluence.arts.uwa.edu.au) contains various different sections, known as “spaces”. There are separate spaces for each of the fourteen Research Clusters being funded by NEER, as well as general spaces for postgraduates, the digital projects, and the Management Committee. Within each space there is a mixture of Web pages, comments, and attached files (including documents and images).

Each NEER participant also has their own personal space, where they can promote their research, record their work-in-progress, and manage their access to other Web sites and blogs. All of the spaces have a “News” function which is equivalent to a blog. An individual user can keep
Confluence is an open source Wiki software that supports version control, document management, collaboration, and tracking. It is designed to help teams work together to create, publish, and manage content. Confluence is particularly useful for projects that require coordination and documentation, such as software development, research collaboration, and academic publishing.

Confluence enables teams to organize their work into spaces and pages, allowing for permissions to be set at the space or page level. It supports version control, so users can track changes and revert to previous versions if needed. Confluence also allows users to add comments to pages, facilitating discussions and tracking changes.

One of the key features of Confluence is its support for pluggable features, which can be used to extend its functionality. This allows teams to customize Confluence to meet their specific needs. Confluence also integrates well with other tools, such as email and RSS feeds, making it easy to stay up to date with changes and additions to the site.

Confluence is important to NEER (the National Endowment for the Humanities) for a number of reasons. It provides a central repository for NEER’s research clusters to share plans and documents. It also provides a way for NEER participants to work together, regardless of their location, and to keep each other informed about ongoing projects.

Confluence is designed to be user-friendly and intuitive, making it easy for users to navigate and find the information they need. It is also scalable, so it can be used for small projects or large organizations.

In conclusion, Confluence is a powerful tool for managing collaboration and documentation in a variety of contexts. It is particularly well-suited for academic and research projects, where version control, versioning, and collaboration tools are essential.
5 PioNEER: Building a Research Repository

The other major goal for NEER’s digital services programme is to promote and disseminate the research and achievements of the Network’s participants. The main priority in this area is to develop NEER’s own digital repository of research outputs, known as PioNEER. 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 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 NEER providing advice and input on configuration and design issues.

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 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 eventually 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 AIATSIS. UWA’s DigiTool has been installed, and is currently being configured and tested.

Digital research repositories have been the focus of a great deal of activity over the last five years or so, but this has mostly concentrated on institutional repositories. Before this, a small number of successful discipline-based e-prints archives had been developed, notably the ArXiv service. These served mainly as an informal method for rapid dissemination of new work and were not necessarily linked to a specific disciplinary body or organization. Initially, at least, institutional repositories emerged from the e-prints movement as a means for individual academic institutions to make their research output available freely to the scholarly community. One of their key attributes was considered to be 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). 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). The only notable exceptions are institutions like 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 like Research Councils UK and the Wellcome Trust which require the open archiving of papers resulting from research funded by them. It is also difficult to predict the effect of the Australian Research Quality Framework (RQF) on the development of institutional repositories. The RQF only requires universities to make available to the assessors digital copies of the four ‘best’ research outputs for each eligible researcher. There is no requirement for these digital objects to be placed on open access, and permissions from publishers only provide for access by the assessors. Only outputs published between 2001 and 2006 will be acceptable. At the same time, the researchers covered by the RQF will only be those who are Level B and above, who have a full-time equivalent status of 0.4 or above, and who are included in one of the nominated research groups. While the RQF process will undoubtedly ensure that all Australian universities have an institutional repository, it remains to be seen whether this will make a significant difference to the proportion of research outputs which are made publicly available in this way.

NEER’s aim is to build a national research repository which 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, 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. NEER does not intend to place any limitations on eligibility similar to those of the RQF. Nor will it be limiting its coverage to the period after 2001.

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-organizing, discipline-based archive either. It will contain outputs from an Australia-wide “virtual organization” which 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 which 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).

Because PioNEER is a national, discipline-linked repository, its relationship to existing institutional repositories is an important issue for the project. NEER’s preference would be to link to relevant digital objects
which are already held in institutional repositories, thus providing a “virtual view” of its research output. But a scoping study earlier in 2007 revealed that very little material from NEER participants (or of interest to NEER) is currently available in Australian repositories. The only significant exception to this seems to be the Flinders Academic Commons at the Flinders University of South Australia, which already contains articles and book reviews by several staff members who are participants in NEER (Phiddian 2006).

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 2005). There are various reasons for this, including lack of awareness, concerns about copyright and quality, and lack of time. Informal discussions with NEER participants have revealed some of the same concerns. While senior researchers can see the merits of having a single location where they can get access to the output of their colleagues, they see less incentive to promote their own work. The possible benefits are generally outweighed by the time and effort required to collect and deposit material. Postgraduate students and early career researchers, on the other hand, have expressed enthusiasm for using PioNEER to promote their own work. This group is often overlooked in the process of building institutional repositories, but is likely to play an important role in contributing to PioNEER.

NEER will be taking a two-pronged approach to the business of acquiring content for the repository. Participants will be encouraged to give preference to their institutional repository if they are interested in depositing their own digital objects. Participants who are not eligible to deposit material in an institutional repository – or whose institution does not have a generally available repository – will be given access to the PioNEER repository for depositing their digital objects.

NEER’s own staff will give priority to identifying and linking to objects already available in institutional repositories. They will also be identifying, collecting and depositing material for the PioNEER repository. Retrospective material will be the main priority. 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. The aim will be to provide a representative selection from the total work of Network participants, which adds up to thousands of publications produced over a period of at least forty years.

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. NEER will be working closely with authors and publishers to ensure that items are deposited in PioNEER with appropriate permission from the copyright owners.

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 a repository (either PioNEER or an institutional repository) as part of their continuing communication with fellow researchers in their discipline. By embedding PioNEER 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 which is required for such approaches to be successful.

Metadata schemas and standards are an important issue for the repository. There are various metadata schemas available for use, ranging from the complexities of MARCXML to the simplicities of Dublin Core. MARCXML is based on the MARC (MAchine-Readable Cataloging) standard for library catalogue records. In Australia, the ARROW (Australian Research Repositories Online to the World) 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, NEER is adopting MODS as its metadata schema. MODS (Metadata Object Description Schema) was developed by the Library of Congress as a cut-down version of MARC, designed for the exchange of bibliographic information (Guenther 2004). MODS will meet 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.

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.

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 – or linked to by – PioNEER).

6 Conclusion

Appropriate digital services are vital for building successful research networks across Australia and beyond. Without “transparent, easy to use and affordable
ICT structures” (Camarinha-Matos 2005: 83), it will be impossible to implement effectively the collaborative networks of the type envisaged by the ARC’s Research Networks programme. In the sciences, the research process itself has been extensively modeled and increasingly replicated in a software-driven Web environment (Coles 2006; Ludäscher 2006; Oinn 2006). Research processes in the humanities, on the other hand, do not readily lend themselves to such approaches.

NEER’s strategy is to focus on the use of information technology to improve collaboration and communication between humanities researchers at a national, discipline-based level. This kind of approach 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 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.

NEER is aiming to contribute to the design of “a next generation system for scholarly communication” (Van de Sompel 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.

While NEER itself is currently only funded until 2009, negotiations are already underway to ensure the continuity of its digital assets after that date. Grants from other ARC programmes will be actively sought, as well as from other public and private sources. Continued support from the University of Western Australia, as the host of the Network, is also likely, and the UWA Library is already committed to maintaining the availability of the PioNEER repository beyond 2009.

NEER has a unique opportunity to test technologies for scholarly communication and the dissemination of research in the humanities, 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 which will be a necessary part of this transformation. Confluence and PioNEER are just the beginning.

7 References


Van de Sompel, H. et al. (2004): Rethinking scholarly communication: building the system that scholars deserve. DLib Magazine 10 (9):