# TOWARDS AN AUSTRALIAN HUMANITIES DIGITAL ARCHIVE



GRAEME TURNER

Towards an Australian Humanities Digital Archive Graeme Turner

© 2008 Graeme Turner and the Australian Academy of the Humanities

Published in 2008 by: The Australian Academy of the Humanities 3 Liversidge St Canberra ACT 2601 Australia enquiries@humanities.org.au +61 2 6125 9860

Funding for the scoping study and this publication was provided to the Academy under a special grant from the Australian Government through the then Department of Education, Science and Training. The views expressed in this report do not necessarily reflect the views of the Government, its Ministers or DEST's successor Departments.

Research assistant: Lesley Pruitt

Project manager and editor: John Byron

Thanks are due for assistance with the administration of the study to:

- John Shipp, University Librarian, University of Sydney
- Anne-Marie Lansdown, General Manager, Research Infrastructure Branch, Department of Innovation, Industry, Science and Research
- Sarah Howard, Phoebe Garrett and Christina Parolin of the Academy Secretariat

Cover image: Gateway Arch, St. Louis, Missouri, USA; courtesy of John Byron.



## TOWARDS AN AUSTRALIAN HUMANITIES DIGITAL ARCHIVE

A REPORT ON A SCOPING STUDY FOR THE ESTABLISHMENT OF A NATIONAL DIGITAL RESEARCH RESOURCE FOR THE HUMANITIES

#### PREPARED BY

## PROFESSOR GRAEME TURNER FAHA

WITH THE ASSISTANCE OF LESLEY PRUITT

SEPTEMBER 2008

PROFESSOR GRAEME TURNER FAHA is an ARC Federation Fellow, Professor of Cultural Studies, and Director of the Centre for Critical and Cultural Studies at the University of Queensland. He is one of the key figures in the development of cultural and media studies in Australia and has an outstanding international reputation in the field. He was elected to the Fellowship of the Australian Academy of the Humanities in 1997 and was its President from 2004-2007. He is the Convenor of the Australian Research Council Cultural Research Network and has served on the ARC's College of Experts in the Humanities and Creative Arts panel. He convened the Humanities, Arts and Social Sciences Working Group for the 2008 Review of the National Collaborative Research Infrastructure Strategy Roadmap. In 2008 he was appointed to the Prime Minister's Science, Engineering and Innovation Council, as only the second humanities scholar to serve since the Council's inception in 1989.

### **CONTENTS**

Overview & Recommendations	4
Introduction	6
Method	6
The need for an Australian Digital Archive for the Humanities	7
Preservation and Access	8
Governance	9
Principles for Selection	11
Collaboration	12
Technology: Possibilities for Innovation & Challenges	13
Researcher engagement, education and training	14
References	16
Appendix A: Digital Humanities outside Australia: A selection	20
Appendix B: List of those interviewed for this scoping study	24
Appendix C: Suggested non-electronic resources suitable for high priority	25

#### **OVERVIEW**

The Australian Academy of the Humanities is pleased to present this report on its scoping study on the development of a national scholarly digital archive in the humanities. The primary aim of the scoping study was to establish exactly what materials would need to be included in such an archive. The resulting report provides rational, evidence-based advice leading towards an answer to that question.

This report is based on a broad, interdisciplinary enquiry that has consulted experts in the humanities in across Australia. Consultations were conducted with a view to establishing an agreed, rational basis for a set of Principles for Selection, and a Governance model for the management of the prioritisation process. Recommendations are presented that address these fundamental questions.

The report also offers recommendations in response to questions of preservation and access, collaboration, technological change and training issues. Finally, the report incorporates an indicative list of possible collections of data or textual material that should be included in a programme of digitisation to set up such an archive.

#### RECOMMENDATIONS

#### The need for an Australian Digital Archive for the Humanities

1. That a national research archive dedicated to digital research resources for the humanities be established, to enable the transformation of humanities and interdisciplinary research practices and outcomes.

#### **Preservation and Access**

- 2. That an audit of existing digital humanities resources and a consideration of their interconnection constitute the first step towards the development of the digital archive.
- 3. That digital resources for humanities research be preserved in accessible, interoperable and long-term sustainable institutional networks.
- 4. That the resources digitised through this programme be freely accessible to researchers and to the public.

#### Governance

5. That the governance of the archive and its procedures for selection, access and preservation be managed by a board comprising representatives of the humanities research community, the e-humanities research community, and the collecting institutions.

#### **Principles for Selection**

- 6. That the principles guiding the selection and prioritisation of materials for digitisation include:
  - demonstrable demand;
  - · academic or research significance; and
  - public interest.
- 7. That the archive manifest a special obligation to prioritise Australian research and cultural materials.
- 8. That, in the first instance, materials which are not subject to copyright clearance be prioritised.

#### Collaboration

9. That the selection of materials support the potential to facilitate interdisciplinary and collaborative research.

#### **Technology: Possibilities for Innovation & Challenges**

10. That the establishment of the archive acknowledge the need to develop appropriate analytical tools and services to accompany the development of digital resources.

#### Researcher engagement, education and training

11. That funding for this initiative also address the requirements for stakeholder education and training.

#### INTRODUCTION

This study emerged from an approach to government by the Australian Academy of the Humanities which proposed the establishment of a national digital scholarly archive for humanities researchers in Australia. The argument for such a proposal included the recognition of the transformation of humanities research that was occurring internationally, and which would be facilitated by developments in technology and resources here; and the recognition that such a resource would have a national role, by providing public as well as academic access to key components of the nation's social and cultural history.

In response to this approach, the Academy was asked to provide considered advice to government on the best way to approach the potentially enormous task of digitising, preserving and making accessible and operable the existing analog resources for humanities researchers in Australia. These resources are widely dispersed, and highly varied in format and textual form: they range from newspapers to literary texts, from census data to explorer's journals.

Specifically, the Academy was asked to provide advice on how such a project might be conceptualised, and what kinds of priorities should drive its selection of materials. The Academy was also asked to research the current developments in humanities digital research internationally, and to provide an account of the best practice in the field. Finally, the Academy agreed to consult researchers, e-research experts, libraries and other representatives of the collecting institutions for their advice on the development and management of a national digital archive for humanities researchers, as well as for their suggestions on particular collections of materials that were especially important to the objectives of such a project.

#### **METHOD**

In the first phase of this study, the Academy consulted with its members by distributing a questionnaire which sought their views on priorities and the selection of material for digitisation. However, this dimension of the study demonstrated the difficulty in generating a prioritised list of resources that acknowledged the interests of the sector as a whole: most of the Fellows' responses took a relatively narrow view of the sorts of resources that would be valuable. While that advice was taken on board, the aggregate of these views was clearly going to produce a very substantial 'wish-list.' A more distanced, less discipline-specific approach was needed.

As a result, the subsequent aspects of the research were directed more towards setting criteria and principles for selection, rather than generating a list of specific resources. There is, at Appendix C of this report, an indicative list of appropriate resources: however, it is included in order to provide useful examples to which agreed selection criteria might apply.

The Academy has undertaken an ongoing international review of the existing literature on current developments in the field to frame our research. Resources included documents on file at the Academy, conference reports, previous in-house reports, and a wide range of applicable scholarly literature. Resources documenting experiences around the world have been useful in framing our plans for an Australian model. For a complete list of the international digitisation programmes considered, see Appendix A.

Following our initial literature review, we interviewed key figures and research leaders in the digital humanities field (see Appendix B). In the interviews we sought participants' views on:

- prioritising digital preservation and achieving desired access;
- governance of the archive;
- community agreements around preservation and access;
- the lessons from previous practice outside Australia; and,
- any specific resources they wished to highlight for digitisation.

The outcomes of these interviews have fed into the recommendations and discussion in the following sections. It is important to note that there was a remarkably high degree of agreement on all the major issues, and so we are entitled to regard the advice from the sector as representing a clear consensus on the best ways to undertake this project.

#### THE NEED FOR AN AUSTRALIAN DIGITAL ARCHIVE FOR THE HUMANITIES

It is important to acknowledge that the original impetus for a proposal to create an archive was derived from the fact that the NCRIS program had not addressed the specific needs of the humanities and some of the social sciences. The inclusion of the humanities, arts and social sciences (HASS) in the 2008 review of the NCRIS Roadmap has to some extent addressed this issue. Given the detail of the HASS capability outlined in the review, which does incorporate an analogous facility, the advice emerging from this study will be very useful in the eventuality that the HASS capability is funded. Given that there is no guarantee that the NCRIS proposal will indeed receive funding, on the other hand, the advice herein will be even more useful if the HASS capability is not funded. In the latter eventuality, there would be a case for continuing to proposing this archive as a nation-building resource in its own right.

Our consultation with leading figures in the sector and with our international colleagues reinforced our view that the technological context for humanities research was in the process of transforming, and transforming dramatically. There is no doubt that Australia must embrace such innovation by creating the necessary infrastructure to access and use the enormous range of information contained in existing and future data (ESFRI 2006: 20). There is unequivocal support for an Australian e-research project that recognises the particular contributions and needs of the humanities (ACLS 2006: 8). The term 'e-research' here in Australia is equivalent to 'cyberinfrastructure' in the US and 'e-science' in the UK. The terms include the levels of shared "expertise and best practices, standards, tools, collections, and collaborative environments" as well as software and applications for specific disciplines and project data collections (Unsworth cited in Anderson 2007: 5). In order to take advantage of developments in e-research, it is important that humanities researchers take a role in constructing and designing such efforts (ACLS 2008).

The international literature supports the view that this work will not only be of use to humanities researchers; it will be a national asset that could benefit many Australians from diverse walks of life. As the American Council of Learned Societies argues, "[t]he value of building an infrastructure that gives all citizens access to the human record and the opportunity to participate in its creation and use is enormous, exceeding even the significant investment that will be required to build that infrastructure." The ACLS goes on to suggest that if their recommendations were adopted, in only five or six years the general public would

take much more interest in research from humanities scholars, and those same scholars would be "answering questions that today they might not even consider asking" (2006: 40).

The Australian scholars we consulted had particular familiarity with projects in the U.S. and the U.K and had observed, over the last 10-15 years, the various experiments, failures, and successes. There is now a considerable amount of information that will assist Australia in avoiding the mistakes of others in the past in relation to the management of preservation and access, governance, principles of selection and a range of related issues.

#### **Recommendation 1:**

That a national research archive dedicated to digital research resources for the humanities be established, to enable the transformation of humanities and interdisciplinary research practices and outcomes.

#### PRESERVATION AND ACCESS

As one European project notes, resources of scientific and cultural heritage hold fundamental value for our present and future, "both as a unique knowledge basis and in terms of their commercial utilization." The development of research capacities in the humanities should therefore ensure that institutions can best utilize the new opportunities presented by the emergence of digital technologies to offer universal access to citizens while preserving resources for future use (DigiCult 2008). We have seen through the experience of others that "[p]utting the historical record online opens it to people who rarely had such access before;" including people, with visual or other disabilities who can gain access through enhanced technology (ACLS 2006: 14). Researchers on an American initiative were surprised to learn that their findings clearly showed "widespread support for a national initiative in long-term preservation of digital content across a broad range of stakeholder groups outside the traditional scholarly community" (Friedlander 2002).

According to our Australian interviewees, the first step in the process of preservation and access should be enhancing discovery: working out what has already been digitised and gathering the meta-data into a distributed network. While countless institutions and individuals have completed small stand-alone digitisation projects, this work has been mostly *ad hoc* and unconnected. Remedying this may involve the development of a federation of collections similar to Picture Australia. A central discovery service with information held on repositories around the network is needed. Ideally, this would have the capacity to include images and oral material as well as text, and would be more sophisticated than basic Google. Once the existing resources have been rationalised and connected, we can better focus on the decisions about what analogue material should be digitised.

Preservation raises the issue of continued funding. It was regarded as crucial that there be a sustainable plan for the continued preservation of materials, as well as the continued management of access to those materials. Digitisation plans should incorporate a long-term plan for their preservation; this means that plans must be institution-based rather than project-based, and that clear lines of responsibility will need to be articulated.

Experience elsewhere has underlined the importance of freedom of access. A local example would be the Australian Dictionary of Biography Online, a globally celebrated and well-used

free database. It is much cheaper than the corresponding project in the U.K., and those we consulted suggested that it has been more widely used precisely because it does not require subscription.

One concern of scholars that relates to access is ensuring that this project is a benefit to the Australian public. The preference for items of public interest should be understood broadly to privilege Australian cultural materials. That raises secondary problems of regional representativeness, for instance, that necessitates a strategic plan for the management of the programme of digitisation, preservation and access. Among the possible priorities our informants suggested was a focus on various themes of Australian history such as discovery and exploration. The public records around key social issues were also mentioned, such as the extension of the franchise to women or to Aboriginal people. There was also a concern for maximising the potential of the digital environment by ensuring that we do not simply (for instance) digitise text records of parliamentary debates, but instead integrate Hansard text with images, sounds, and other forms of data. The need to use the digital technology so as to enable multidisciplinary research was a high priority for the academic researchers we consulted.

#### **Recommendation 2:**

That an audit of existing digital humanities resources and a consideration of their interconnection constitute the first step towards the development of the digital archive.

#### **Recommendation 3:**

That digital resources for humanities research be preserved in accessible, interoperable and long-term sustainable institutional networks.

#### **Recommendation 4:**

That the resources digitised through this programme be freely accessible to researchers and to the public.

#### **GOVERNANCE**

Implementing a project of this size requires the development of a detailed and sustainable governance plan. Only a few reports from abroad give even a brief explanation of how digitisation projects have been governed. The LC project, for example, began with an Advisory board that made scoping decisions and decided that the first phase would focus on 'born digital' information, and the 4 key aspects of framing the project would include meetings with stakeholders, collaborative research, constructing a conceptual framework, and scenario planning (Friedlander 2002).

Pointing to the necessity of establishing standards, policies, and procedures, Allen notes that development work should commence with the generation of a policy framework that explains that digitisation will be included in institutional goals and objectives; once that framework is approved by management staff, it can be the basis for developing particular guidelines and policies to reflect the framework's vision. Standards must then be addressed and explained to

stakeholders. And, in the end, procedures should be planned and documented for all processes connected to managing and preserving digital data in a simple but flexible way (Allen 2006).

Some scholars note that, while a widespread collaborative infrastructure for humanists could pave the way for innovations in research questions and practices, it remains challenging to sell the idea to humanities scholars, since "the universal benefits are not immediately recognizable across the community, and while individuals may accept the overall value of such a foundation, they do not necessarily see a direct benefit to themselves" (Goldenberg-Hart 2004: 3). For this reason, ACLS commissioners have considered asking research funding bodies "to require recipients to follow certain protocols for the dissemination of their research that would promote advancement in making that research accessible in digital form and to do so in ways that support inter-institutional collaboration and collaboration among scholars generally" (Goldenberg-Hart 2004: 3).

Reports on the US Project Bamboo also discuss how these projects can be governed. Broughton and Jackson's proposal for the project outlines their 1½-year community design and planning agenda, which is to take place through a progression of workshops and conversations. These forums will "map out the scholarly practices and common technology challenges across and among disciplines, and discover where a coordinated, cross-disciplinary development effort can best foster academic innovation" (2008: 3). Workshops will establish the process with "principal investigators, academic leaders, key institutional stakeholders and critical program staff" (Broughton and Jackson 2008: 27). Other responsibilities will include exploring scholarly practices to best meet needs in the area, and forming a leadership council that will be tasked with helping to direct actions to ensure the programme remains attentive to its objectives (Broughton and Jackson 2008:29; 34).

During our expert consultations, we asked scholars to share their perspectives on governance issues, particularly as they concern key interests identified in the literature reviews. Informants generally agreed that sustainability could best be achieved by linking the project to an existing institution, either a national institution or group of institutions. While we do see merit in having a central location for the service, the chosen agent needs to have a strong acceptability to key players. This should be supported by the presence of an advisory board with decision-making capabilities that is separate from the contributing institutions. This advisory board of 5-12 people should include representatives from the research field as well as from the collecting institutions. A possible model is the Bamboo project in the U.S.: Bamboo commenced with the foundation of a leadership council, and then moved toward an open and evolving board.

On the whole, it is clear that the governance model is vital for successful implementation of this project. The board will need to have national reach and supra-institutional authority, and its members must enjoy the confidence and respect of the various communities: the humanities scholars, the e-research community and the collecting institutions.

#### **Recommendation 5:**

That the governance of the archive and its procedures for selection, access and preservation be managed by a board comprising representatives of the humanities research community, the e-humanities research community, and the collecting institutions.

#### PRINCIPLES FOR SELECTION

The most difficult component of designing a project of this of this sort is developing the principles for selection. Hazen et al. (1998) note that selecting items for digitisation is a complex practice that is very similar to other strategic decision-making processes that libraries conduct – in areas such as purchase, withdrawal, and microfilming – and that these judgments involve many complex factors. However, it is complicated by the fact that any current project will be working across a vast domain of possible materials from which to make its selection. There is an urgent need to arrive at an agreed set of criteria to organise this process. For example, the Library of Congress (LC) Preservation Division uses the following criteria when selecting items for digitisation: value, condition, use, characteristics of originals, acceptability of digital product, and access aids.

Some authors urge the weighting of some criteria over others: particularly that of use. For example, Paula De Stefano argues that "use holds significant promise as a guiding factor in selecting materials for digital conversion," not least because "[t]he idea of use, especially high use, is fundamental to collection development" (2001: 66). In addition to these published discussions, our discussion with the Arts and Humanities Research Council revealed that the AHDS had come to end after 11 years in operation, largely because the service was not being widely utilised.

Our research and our subsequent interviews resulted in a high degree of consensus within the field on three criteria that should be used to prioritise the selection of items for digital preservation. The material selected should be:

- (1) demonstrably in high demand already;
- (2) of high strong academic or research importance; and
- (3) assessed as likely to generate strong public interest.

The experts consulted broadly agreed that these are reasonable criteria, and added several further comments and suggestions regarding implementation and challenges. At times, for instance, they warned that these criteria will compete with one another if there is a limited budget. In terms of how one might address this, one interviewee suggested that the key driver should be research usage, or capacity for immediate academic use. Another argued that born-digital material should be a priority where there is no other version of it.

Importantly, our interviewees stressed that public lawful accessibility is key. Since copyright is such a central issue in this area, we were advised to first focus on material that is copyright cleared and thus available for free and open access.

We also asked our experts to offer their top three suggestions for non-electronic resources for digitisation. We have compiled and condensed these recommendations (see below), but the exercise also flushed out some interesting observations on prioritisation. One respondent rightly pointed out that it would be a considerable shame to only digitise obvious things, since no one is going to revolutionise their research that way. It was also pointed out that we must pursue collaborative research that will enable us to deal with a substantial quantity of information in new ways. The real transformation of humanities research that such a resource could enable is the capacity for people to work together in large-scale projects on questions they were previously unable to address or even conceptualise.

Our interviewees agreed that we should focus on material that is particularly Australian, originating in Australia or at least directly relating to Australia. Some of the more obvious

candidates for digitisation are items of historical and news significance. Newspapers and political documents would be useful, and our panel could probably determine which ones to do by using national interest as a criterion. Other priorities would be items of Australian culture, such as Australian music, film, or text, particularly any music or film that pre-dates the 1950s.

It was also agreed that the governing board should establish procedures for dealing with these issues and handling the management of priorities as its primary task. It might be necessary for the board to take specialist advice from time to time, from researchers and from those with expertise in the relevant technologies.

A list of the key items proposed by our interviewees is included in Appendix C. Each item is accompanied by a brief explanation of its provenance and why it should be digitised. This list is offered as a set of examples rather than as a wish list of items to be proposed. The overwhelming view of those we interviewed was that the construction of a list of prioritised items – initially an objective of this study – was inappropriate at this stage, and that it would be more useful to establish a set of principles to guide this process, and a mode of governance through which it might be managed.

#### **Recommendation 6:**

That the principles guiding the selection and prioritisation of materials for digitisation include:

- demonstrable demand;
- · academic or research significance; and
- public interest.

#### **Recommendation 7:**

That the archive manifest a special obligation to prioritise Australian research and cultural materials.

#### **Recommendation 8:**

That, in the first instance, materials which are not subject to copyright clearance be prioritised.

#### **COLLABORATION**

Another key factor is the potential this proposal holds for enabling collaboration between researchers and across institutions. Deegan notes that in any country it is critical to ensure collaboration with a diverse array of actors, including libraries, educational institutions, community organisations, technology providers, public sector bodies, and end users, among others (2002). Likewise, Allen discusses practitioners' emphasis on collaboration and partnerships as key to conducting digital management and preservation with adequate resources through shared resources, responsibilities, and results. To this end, she says that "[d]igital preservation programmes are most effective when they are built with partnerships in mind" (Allen 2006). Collaboration can also be important in addressing the need to "facilitate sharing and cooperative ownership with other libraries" (Payne 2007: 29).

The experience of earlier projects with commercial collaboration is also worth considering. Some prospects discussed include agreements made between universities and commercial entities that give the libraries rights to non-commercial use of their holdings, and prevent the commercial supporter from levying fees for searching digitised books through the libraries (CLIR 2008: 14). Oya Rieger, in her recent CLIR white paper addressing digital preservation, looks at library and commercial partnerships, but explains that there are complex questions being posed about whether commercial companies can or should be trusted with preserving cultural heritage for the public good (2008: 9-12). In practical terms, we have been advised to examine how this project relates to existing work in the international context, such as Google Books. There is a certain amount of Australian content in Google Books, for instance, because it was digitised as part of the holdings of North American collections.

Copyright and intellectual property rights are also implicated here. It is imperative that the archive project obtains the benefit of expert advice in order to establish mechanisms which can effectively deal with such issues. This work may include the development of contracts setting out network terms and conditions or end-user agreements. Several scholars (e.g. Hazen et al. 1998 & Chen 2005) also note the importance of resolving copyright issues surrounding access and potential income.

Overall, the general consensus is that this project should aim to bring together many existing institutions and individuals on a network basis, rather than simply confining itself to seeking representation from individual universities. Copyright issues must also be addressed through deliberate and informed planning, including discussions with government about the role it may play in making scholarly information more open to public accessibility. All of these tasks are important and should be directed by the governing body recommended in the previous section.

#### **Recommendation 9:**

That the selection of materials support the potential to facilitate interdisciplinary and collaborative research.

#### TECHNOLOGY: POSSIBILITIES FOR INNOVATION & CHALLENGES

Technology can bring with it both prospects for innovation and new challenges. Advances in technology have paved the way for significant opportunities for research innovation. CLIR reports on how mass digitisation projects can open doors to investigating new research topics that are cross-disciplinary or require large-scale data. Some examples include tracking linguistic patterns or intellectual history over culture, space, and time, as well as social and economic history and cultural and cross-linguistic studies (CLIR 2008: 6-7). At the same time, they note the need for document comprehension, multi-lingual services, and conversion of basic text to machine actionable data (CLIR 2008: 9-12).

Likewise, Anderson notes "the need for existing highly dispersed content to be joined up through appropriate tools – ontology connectors for example, and in the longer term, through the development of a Semantic Grid for the arts and humanities," but acknowledges that "this must extend beyond search and retrieve mechanisms, and encompass curation, annotation, de-duplication and processing data for analysis," which are substantial goals that require serious participation by experts to guarantee consistency and quality (2007: 13). One major

priority is to address the deficiency in search capability for materials that are not text based, since rich new resources are becoming available, while researchers' abilities to meaningfully engage with them remains limited (Anderson 2007: 14). Despite these challenges, Allen (2006) suggests that practitioners recognise that they must implement 'good enough' practices for now, even as they acknowledge that technologies will continue to evolve and require updates, creating a need for flexibility.

Other technological aspects must also be addressed by the experts involved in the implementation of this plan. Key questions include: What tools do you need at the front end? How do we enable different sets of data and users to talk to each other? Among the particular difficulties for humanities e-research is that we need to build tools simple enough for people to use without first acquiring IT degrees, while ensuring that the data is usable to them in the present and also robust enough for changing and unforeseeable future use.

Moreover, some early, rare, fragile or ephemeral hard-copy resources have been successfully format shifted, but to analogue formats such as microfilm, and these need to be made digital. Standards for completing these shifts need to ensure that the digitised items are made susceptible to maximum future data mining and techniques to maximise research outcomes. This may make initial capitalisation greater, but it is important for the long-term payoff on our cyberinfrastructure investment. Given the willingness of many scholars in this area to take part in our original planning, we are confident this should not be a great hindrance to the project overall.

#### **Recommendation 10:**

That the establishment of the archive acknowledge the need to develop appropriate analytical tools and services to accompany the development of digital resources.

#### RESEARCHER ENGAGEMENT, EDUCATION AND TRAINING

While noting the positive prospects here, it is also necessary to look at the challenges and problems encountered by others so that we can address or avoid them in our scheme. At the same time, Chen (2005) has identified the need for considering specific national perspective and needs rather than adopting methods uncritically from other countries such as the US. The ACLS also notes the particular challenges resultant upon the lack of sufficient "resources, will, and leadership to build cyberinfrastructure for the humanities and social sciences" (2006: 18).

A key challenge which emerges from the international comparisons is what has been called 'stakeholder buy-in'. Christy Allen (2006) argues that stakeholder support and buy-in is crucial for the success of this kind of initiative. It requires both internal and external education efforts, as well as an awareness among those charged with framing digitisation priorities that we need to privilege those projects that will provoke both interest and innovation. One expert we consulted recommended the value of using a prototype to demonstrate the benefits of e-research: this should be sufficiently broad to interest different disciplines and might include something like a multidisciplinary approach to a particular historical era (say, the 1920s).

A further key challenge is education and training: informing educators and leaders of educational institutions on the use and development of e-learning environments; and training our students in the technologies and research practices they enable. Internationally, this has been a prominent concern (Anderson 2007: 21). The task is significant but imperative, as CLIR suggests: "we should examine the future education and training required to create a new information professional" (2008: 25-7).

#### **Recommendation 11**:

That funding for this initiative also address the requirements for stakeholder education and training.

#### REFERENCES

- Abate, Tom. 2008. 'Librarians discuss how to store world's data.' 29 May 2008. San Francisco Chronicle. Page C-1. Online. Accessed 30 May 2008. Available at: http://www.sfgate.com/cgi-bin/article.cgi?f=/c/a/2008/05/28/BU3710U908.DTL.
- Allen, Christy E. 2006. 'Foundations for a Successful Digital Preservation Program: Discussions from Digital Preservation in State Government: Best Practices Exchange 2006. RLG DigiNews. 15 June 2006. Volume 10, No. 3. Accessed 16 July 2008. Available at:
  - http://digitalarchive.oclc.org/da/ViewObjectMain.jsp;jsessionid=84ae0c5f82400f32ffe0d 8f0469480bfff9f59028dae?fileid=0000070519:000006287177&reqid=766.
- American Council of Learned Societies (ACLS). 2006. Our Cultural Commonwealth: The Report of the American Council of Learned Societies on Cyberinfrastructure for the Humanities and Social Sciences. Accessed 12 June 2008. Available at: http://www.acls.org/uploadedFiles/Publications/Programs/Our\_Cultural\_Commonwealth.pdf.
- American Council of Learned Societies (ACLS). 2008. 'ACLS Commission on Cyberinfrastructure.' Webpage. Accessed 6 June 2008. Available at: http://www.acls.org/programs/Default.aspx?id=644.
- Anderson, Sheila. 2007. 'The Arts and Humanities and e-Science: Scoping Survey Report and Findings.' King's College London. Centre for e-Research. December 2007. Available online at: http://www.kcl.ac.uk/content/1/c6/03/60/27/e-sciencereportfinal.pdf. Accessed 2 June 2008.
- Arts and Humanities Data Service (AHDS). 2008. Website. Accessed 2 June 2008. Available at: http://ahds.ac.uk/.
- The British Academy. 2005. *E-resources for research in the humanities and social sciences: A British Academy Policy Review*, April 2005.
- Broughton, Janet and Gregory A. Jackson. 2008. 'Bamboo Planning Project: An Arts and Humanities Community Planning Project to Develop Shared Technology Services for Research.' A Proposal to the Andrew W. Mellon Foundation. 16 January 2008. Public Document Revision, January 28, 2008.
- Chen, Chao-Chen. 20005. 'Going Digital: Taiwan Experiences.' *Going Digital Experiences* from East Asia and the Pacific. Collected Works from the June 2005 American Library Association Conference. Available online at: http://csuspress.lib.csus.edu/goingdigital/chenpaper.html. Accessed 12 June 2008.
- China National Knowledge Infrastructure (CNKI). 2006a. 'History.' Website. Accessed 12 June 2008. Available at: http://www.global.cnki.net/grid20/Aboutcnki/History.htm.
- China National Knowledge Infrastructure (CNKI). 2006b. 'Introduction.' Website. Accessed 12 June 2008. Available at: http://www.global.cnki.net/grid20/Aboutcnki/Introduction.htm.

- Corpus Automatum Multiplex Electorum Neolatinitatis Auctorum (Camena). 2008. Homepage. Website. Accessed 12 June 2008. Available at: http://www.unimannheim.de/mateo/camenahtdocs/camena e.html.
- Crane, Gregory and Amy Friedlander. 2008. 'Many More than a Million: Building the Digital environment for the age of abundance.' Report of a One-Day Seminar on Promoting Digital Scholarship Sponsored by the Council on Library and Information Resources (CLIR).
- DARIAH-Digital Research Infrastructure for the Arts and Humanities. 2008. Homepage. Website. Accessed 12 June 2008. Available at: http://www.dariah.eu/.
- De Stefano, Paula. 2001. 'Selection for Digital Conversion in Academic Libraries.' *College and Research Libraries*. 62 (January): Pp. 58-69.
- Deegan, Marilyn. 2002. 'Position Paper on the Digitization of Historic Newspapers.' Digitization & Preservation Online Resource Center. Available at: http://digitalarchive.oclc.org/da/ViewObjectMain.jsp;jsessionid=84ae0c5f824072cff2afec 6e458cb0c6e8bcf1687c6a?fileid=0000016179:000000676918&reqid=58. Accessed 12 June 2008.
- Digital Curation Centre. 2008. 'About the DCC.' Website. Available at http://www.dcc.ac.uk/about/. Accessed 12 June 2008.
- Digital Heritage and Cultural Content (DigiCult). 2008. 'DigiCult Home.' Website. Available at: http://cordis.europa.eu/ist/ka3/digicult/. Accessed 12 June 2008.
- Digital Library Federation. 2008. 'DLF Aquifer: Bringing collections to light through the Digital Library Federation.' Webpage. Accessed 10 June 2008. Available at: http://www.digilib.org/aquifer/.
- Doorn, Peter. 2007. 'Preparing for the construction of the Digital Research Infrastructure for the Arts and Humanities: Preparing DARIAH.' 2 May 2007. Combination of Collaborative Project and Coordination and Support Action. Construction of new infrastructures—preparatory phase. FP7—INFRASTRUCTURES—2007—1. http://www.dariah.eu/files/DARIAH\_preparatoryproject\_proposal\_final.pdf.
- e-Taiwan. 2008. 'e-Taiwan Program.' Accessed 14 July 2008. Available at: http://www.etaiwan.nat.gov.tw/content/application/etaiwan/english/index.php.
- European Strategy Forum on Research Infrastructures (ESFRI). 2006. *European Roadmap for Research Infrastructures*. Report 2006. John Wood, ESFRI Chair.
- Friedlander, Amy. 2002. 'The National Digital Information Infrastructure Preservation Program: Expectations, Realities, Choices and Progress to Date.' *D-Lib Magazine*. April 2002. Vol. 8, No. 4.
- Goldenberg-Hart, Diane. 2004. 'Libraries and Changing Research Practices: A Report of the ARL/CNI Forum on E-Research and Cyberinfrastructure.' *ARL: A Bimonthly Report on*

- Research Library Issues and Actions from ARL, CNI, and SPARC. Washington: Association of Research Libraries.
- Hazen, Dan, Horrell, Jeffrey, and Jan Merrill-Oldham. 1998. 'Selecting Research Collections for Digitization.' August 1998. Council on Library and Information Resources (CLIR). Commission on Preservation and Access Digital Libraries. Online. Accessed 6 June 2008. Available at: http://www.clir.org/pubs/reports/hazen/pub74.html.
- Joint Information Systems Committee (JISC). 2008. 'Supporting Education and Research.' Website. Available at http://www.jisc.ac.uk/. Accessed 12 June 2008.
- Kirkham, Ruth and John Pybus. 2006. 'Building a Virtual Research Environment for the Humanities: Report of the User Requirements Survey.' Oxford University.
- The Library of Congress (LC). 2006. 'Selection Criteria for Preservation Digital Reformatting.' The Library of Congress Preservation Homepage. Online. Accessed 6 June 2008. Available at: http://www.loc.gov/preserv/prd/presdig/presselection.html.
- LOCKSS. 2008. Homepage. Accessed 10 June 2008. Available at: http://www.lockss.org/lockss/Related\_Projects.
- Mateo: Mannheimer Texte Online. Website. Accessed 10 June 2008. Available at: http://www.uni-mannheim.de/mateo/epo.html.
- National Endowment for the Humanities (NEH). 1998. Division of Preservation and Access. Online. Accessed 12 June 2008. Available at: http://www.neh.gov/news/report98/preserve.htm.
- National Endowment for the Humanities (NEH). 2008. Office of Digital Humanities: About Us. Webpage. Accessed 12 June 2008. Available at: http://www.neh.gov/ODH/About/tabid/56/Default.aspx.
- Payne, Lizanne. 2007. *Library Storage Facilities and the Future of Print Collections in North America*. Dublin, Ohio: OCLC Programs and Research.
- Project Bamboo. 2008. 'Project Bamboo.' Homepage. Accessed 18 June 2008. Available at: http://projectbamboo.uchicago.edu/
- Rieger, Oya Y. 2008. *Preservation in the Age of Large-Scale Digitization: A White Paper*. February 2008. Council on Library and Information Resources: Washington, D.C.
- Rimmer, Jon, Warwick, Claire, Blandford, Ann, Gow, Jeremy, and Buchanan, George. 2008. 'An Examination of the physical and digital qualities of humanities research,' *Information Processing and Management: an international journal*, 44(3): 1374-1392.
- Schmitz, Dawn. 2008. The Seamless Cyberinfrastructure: The Challenges of Studying Users of Mass Digitization and Institutional Repositories. April 2008. Council on Library and Information Resources.

Universitat Bielefeld (UB) Bibliothek. 2008. Website. Accessed 4 June 2008. Available at: http://www.ub.uni-bielefeld.de/diglib/aufklaerung/zeitschriften.htm.

Virtual Knowledge Studio (VKS). 2008. Website. Accessed 4 June 2008. Available at: http://www.virtualknowledgestudio.nl.

# APPENDIX A DIGITAL HUMANITIES OUTSIDE AUSTRALIA: A SELECTION

#### United Kingdom (UK)

- The Arts and Humanities Data Service (AHDS) has recently been closed down after 11 years in operation. A national service, the AHDS supported "the creation, curation, preservation, and on-line dissemination of digitized research and teaching materials and collections in the arts and humanities" (AHDS 2008). The AHDS was financially supported by the Arts and Humanities Research Council (AHRC) and the Joint Information Systems Committee (JISC), which is a national funding body that seeks "to provide world-class leadership in the innovative use of ICT to support education and research" (JISC 2008). The project was a collaboration of six partner institutions with a Managing Executive based at King's College London" (AHDS 2008). Unfortunately, due to the lack of the use of its materials, and perhaps due to a strong focus on project-based development, its funding has not been continued after 2008. In many ways, it provides a good example of the mistakes we wish to avoid.
- The Digital Curation Centre. Noting that born digital documentary heritage is at risk of becoming technologically obsolete, the DCC supports "UK institutions who store, manage and preserve these data to help ensure their enhancement and their continuing long-term use." (DCC 2008). The centre focuses on digital curation and preservation, providing advocacy and expert advice to the community, facilitating "an informed research community with established collaborative networks of digital curators," and providing "a wide range of resources, software, tools and support services" (DCC 2008). Funding is provided by JISC and the e-Science core programme.

#### Continental Europe

- The Virtual Knowledge Studio (VKS) for the Humanities and Social Sciences, run by the Royal Netherlands Academy of Arts and Sciences (KNAW), is working on e-research. A core element "of the Virtual Knowledge Studio is the integration of design and analysis in a close cooperation between social scientists, humanities researchers, information technology experts and information scientists. This integrated approach provides insight in the way e-research can contribute to new research questions and methods" (VKS 2008).
- Germany's Corpus Automatum Multiplex Electorum Neolatinitatis Auctorum (Camena) project, while not large, offers an excellent selection of materials, particularly in the fields of history and politics. Our understanding is that they requested that a couple of leading experts in the field select a core group of crucial texts, which is the method many of our interviewees suggested should be adopted in Australia for the preliminary stages (Camena 2008).
- The University of Mannheim's Mateo project is a case of significant digitisation carried out by a library (Mateo 2008).
- Bielefeld University has also embarked on digitisation work that is of particular benefit to the humanities community. While the site is available in German only, it appears that they

- have focused their efforts largely on the digitisation of Enlightenment journals and magazines, which has produced a really useful resource (UB Bibliothek 2008).
- The Herzog August library at Wolfenbüttel also has a digitisation programme, and it is likely that several others must as well. In this work it appears that the Deutsche Forschungsgemeinschaft works in partnerships and plays a key coordinating role.

#### Joint European Initiatives

- DigiCult stands for Digital Heritage and Cultural Content. DigiCult is part of the research activity for "the Information Society Technologies (IST) Programme, a European Commission programme addressing the pervasion of Information and Communication Technologies (ICT) into all aspects of the European citizen's life (DigiCult 2008). Likewise, DigiCult research promotes innovation and development of technical systems and tools for using "both traditional and digital Cultural Heritage resources" (DigiCult 2008).
- DARIAH is slated to be a massive advance in the European digital landscape. The Digital Research Infrastructure for the Arts and Humanties (DARIAH) aims to enable long-term access and preservation for research data relevant in the arts and humanities realm. Additionally, DARIAH can be used to "find and use a wide range of digital content from across Europe, find and use tools that aid data interpretation, participate in experimentation and innovation with other scholars across multiple domains, seek advice, and exchange ideas and knowledge on all aspects of digital scholarship," and "ensure that they work to accepted standards and best practices." The project's website is available at http://www.dariah.eu/?q=node/37.
- SHARE, which commenced in 2002, makes data infrastructure available for fact-based analyses of the continual changes occurring in Europe based on the ageing of the population. The survey began with 8 countries and has since then expanded to 10 member states; project developers eventually hope to expand to cover all EU member states (ESFRI 2006: 36).
- MEDLIB, "a pan-European network of digitised catalogues and texts based on surviving collections of medieval monastic libraries," seeks to create a central, coordinated database that includes the holdings of European medieval monastic libraries (ESFRI 2006: 68). The project's three phases include producing a register of existing catalogues, compiling and digitising unpublished catalogues, and combining the information gather from the two previous tasks in order to form a centralised network (ESFRI 2006: 68).
- Since 1984, the International Social Survey Programme has functioned as a yearly series based on international collaboration, bringing existing social science projects together to match up research objectives and in doing so provide a "cross-national perspective to the individual, national studies" (ESFRI 2006: 69). The project is active across Europe and includes 15 non-European countries as well, making it "by far the most widely utilized international social research infrastructure with respect to scientific publications" (ESFRI 2006: 69).
- The Hanseatic Historical Archives Network, in association with the Wittgenstein Archives (WAB) and MENOTA, "supports development of electronic archival tools and

gives access to archival material. It will help historians and librarians and archivists digitise Hanseatic archives within a maximum availability network...WAB is a humanities research infrastructure specialising in philosophy, Wittgenstein research, and editorial philology and text technology. Menota is a co-operative effort of the leading Nordic editorial institutions (archives, libraries and societies). It aims at implementing standards for the encoding and display of medieval primary sources, in the Nordic vernaculars as well as in Latin, and for building up a pan-Nordic lexicographical database" (ESFRI 2006: 69)

#### United States (US)

- The National Endowment for the Humanities (NEH), unsurprisingly, has been a key player in the field of digital humanities for many years. In 1998 the NEH's Division of Preservation and Access finished a decade "of a multiyear program for the preservation of brittle books and serials," and reported on its "national initiative to catalog and preserve on microfilm the country's newspapers on a state-by-state basis" (NEH 1998). Since then the NEH has modernised further, with the Office of Digital Humanities (ODH) taking a lead in funding and facilitation in the digital scholarship field (NEH 2008).
- The Digital Library Federation's Aquifer initiative aims "to promote effective use of distributed digital library content for teaching, learning and research in the area of American culture and life," by developments that aid access to digital content for students and scholars in the places where they work (DLF 2008). The project we are proposing would pursue similar goals in the Australian context.
- The LOCKSS initiative (Lots of Copies Keep Stuff Safe) is based at Stanford University. It "is a community based, open source, persistent access digital preservation system" that "enables publishers to guarantee long term perpetual access to their content, by allowing libraries to store, preserve, and provide back-up access to the content they have purchased" (LOCKSS 2008). Journal publishers have allowed alliance member libraries of LOCKSS to use the "software to collect, preserve, and provide back-up access to its content" (LOCKSS 2008).
- Project Bamboo is also a significant American effort related to humanities digitisation. "Bamboo is a multi-institutional, interdisciplinary, and inter-organizational effort that brings together researchers in arts and humanities, computer scientists, information scientists, librarians, and campus information technologists to tackle the question: How can we advance arts and humanities research through the development of shared technologies?" (Project Bamboo 2008). While it is still in the preliminary stages, Australian scholars have been present in framing sessions and can draw something from what has been learned so far, as discussed in the issues section. One of the directors of Bamboo, Chad Kainz, came to Australia in August this year to address a workshop run by the Expert Working Group on the NCRIS HASS Capability.

#### Asia

• Since its inception in 1996, the CNKI has become a "key national e-publishing project," that includes, among other things, e-journals, "newspapers, dissertations, proceedings, yearbooks and reference works" (CNKI 2006b). The initiative began with the 1996 release of CD-ROM versions of China Academic journals; this was followed in 1999 by

the release of "the online version of China Academic Journals," and in 2003 CNKI announced its goal of becoming China's gateway for all academic resources (CNKI 2006a). To date, CNKI's academic databases have served over "5,500 universities, public and corporate libraries, hospitals and other institutions" in China and abroad (CNKI 2006b).

• The e-Taiwan Project's National Science and Technology e-Learning Program (ELNP), also requires consideration. The ELNP is "projected to lead Taiwan to the cutting edge in digital content creation and applications" by building "a high quality e-Learning environment," and upgrading "Taiwan's overall competitiveness in the era of knowledge economy," stimulating "the development of industries related to e-Learning," and bringing "about new waves in academic research" (Chen 2005). A major part of this is the National Digital Archives Program (NDAP), which seeks to "promote and coordinate content digitisation and preservation at leading museums, archives, libraries, universities, research institutes, and other content holders in Taiwan" (Chen 2005). Through this, the NDAP seeks to preserve and maintain Chinese culture and civilisation (Chen 2005).

# APPENDIX B LIST OF THOSE INTERVIEWED FOR THIS SCOPING STUDY

- Professor Paul Turnbull: Professor of History, Griffith University
- Professor Brian Fitzgerald: Director, Intellectual Property: Knowledge, Culture and Economy, Queensland University of Technology
- Professor Tom Cochrane: Deputy Vice-Chancellor, Queensland University of Technology
- Mr. Gavan McCarthy: Director, eScholarship Research Centre, University of Melbourne
- Professor Richard Maltby: Head of School of Humanities, Flinders University
- Dr. Toby Burrows: Principal Librarian, Scholars' Centre, University of Western Australia
- Mr. Ross Coleman: Director, Sydney eScholarship, University of Sydney
- Mr. John Shipp: University Librarian, University of Sydney
- Dr. Warwick Cathro: Assistant Director-General, National Library of Australia

In addition, the questionnaire was sent to Fellows of the Australian Academy of the Humanities. 37 responses were taken into consideration, although the replies were not identified by name.

# APPENDIX C SUGGESTED NON-ELECTRONIC RESOURCES SUITABLE FOR HIGH PRIORITY

- Historical Legal Books and Materials of Significance: This would include cases, legislation, and commentaries. Most will pre-date 1900, as much material after that date has been digitised already. Through this, lawyers could get a better understanding of the history of Australian law, and digitising these resources would mean they could refer to it and access it consistently. Anything in the public domain (out of copyright, at least 100-120 years, pre-1900 maybe pre-1920) should definitely be digitised and openly accessible.
- Tasmanian Convict Records: These are world heritage listed records that have already been subject to some development. They could be put to multiple uses and could go to the public domain quickly. Researchers working with them currently have the technology to put them online now but there is no more funding. About 70,000 images, around half of the total, had been completed before the funding ran out.
- Australian Literature: The highly regarded AustLit provides a work-ready infrastructure for a large programme to digitise literary texts. There is a strong cultural heritage argument for working on these texts.
- Documents relating to the foundation of Australia: 18<sup>th</sup> Century British documentation, First Fleet journals, from discovery to federation, the end of convict transportation in 1848, early census data, NSW blue books, and early explorer diaries. Other significant epochs of historical interest that would meet the three criteria articulated in the report include:
  - o Documents relating to mass migration in the 40s, mostly from Ireland, and generally this period of early free migration;
  - o Materials relating to the discovery of mineral resources in Australia—gold, uranium, coal, natural gas, etc.;
  - o Documents, images, and other materials relating to World War I; and
  - o The Depression and its social effects.
- Australian Newspaper Records: There are projects under way at the moment, but they
  constitute a good example of the need to link up independent programmes of digitisation
  to develop an integrated and networked national archive. Newspapers are of profound
  significance, and we should aim for digital public access to the entire corpus of
  newspaper records in Australia. There is no question that Australian newspapers are
  extremely important to research on social history and events, and how they were thought
  of at the time.
- Other Australian Periodicals: The argument was made that these are also very important, and it would be worth aiming at a systematic digitisation of journals of opinion in the Australian context. It would be a significant challenge for the management board, however, dealing with small magazines, literary magazines, science magazines, and other ephemera of periodical literature outside newspapers.
- Australian Census Data: Having access to key demographic data sources would be extremely useful and could be used by researchers in numerous fields. However, this is

just the tip of the iceberg and needs to be undertaken in collaboration with colleagues in the social sciences, and in consultation with other archives and databases.

- Records of Notable Australian Researchers: the example given was Sir Macfarlane Burnet, who was (with Sir Peter Medawar) awarded the Nobel Prize in Physiology or Medicine in 1960. Some of Burnet's records are imaged, but none are in the public domain. They are, however, used frequently by local and overseas scholars.
- Historical Scientific Data from and/or about Australia or our region: This is a large field of course, but there may be particular topics relating to issues of national priority such as environmental change and sustainability (historic flora and fauna journals for instance), or cultures of the South Pacific (explorers and missionaries notes and memoirs).
- Australian Database Projects: One example is the AusStage project, which is compiling a comprehensive online database of all performance events in Australia since 1788. There are a number of record sources scattered around the country: none is large, but the whole will be once they are compiled, digitised and put into a database. The scholar who put forth this idea suggested that we need demonstrator projects to show what can be done and illustrate the potential coming from the proposed archive. Thus, he suggested that we identify a particular collection, such as AusStage, and showcase it as an exemplar to demonstrate methodologies and possibilities. We should then focus more on the process of building on work that has already been done through project grants (e.g. ARC-LIEF grants). At the moment, project apparatus are seldom transformed into enduring infrastructure, and we need to turn established projects into infrastructure for future researchers.

THE AUSTRALIAN ACADEMY OF THE HUMANITIES is the college of leading scholars of the human sciences in Australia. It comprises almost 500 outstanding scholars, teachers, researchers and artists, all internationally recognised leaders in their spheres. The Academy serves the Australian community and the disciplines of the humanities by working to ensure that the contribution of the humanities in Australia is valued, supported and expanded.

The Academy is the primary source of advice on matters relating to the humanities in Australia, for the benefit of the Government, the media, industry and the general public. The Academy is uniquely positioned to draw on the extensive goodwill of the humanities community in Australia, including the Fellows of the Academy, their peers in humanities scholarship, and the many people working in academia, government, the cultural sector and beyond who value the contribution of the humanities and the work of the Academy. The nation benefits substantially from the voluntary expert contributions made through the Academy by its Fellows and friends.



THE AUSTRALIAN ACADEMY
OF THE HUMANITIES