

Humanities, Computers and Cultural Heritage

Proceedings

of the XVI international conference  
of the Association for History and Computing


14-17 September 2005

Royal Netherlands Academy of Arts and Sciences  
Amsterdam, 2005

© 2005 Royal Netherlands Academy of Arts and Sciences

No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photo-copying, recording or otherwise, without the prior written permission of the publisher.

ISBN 90-6984-456-7

The paper in this publication meets the requirements of  iso-norm 9706 (1994) for permanence.

### **Design**

Edita-KNAW, Amsterdam  
[www.knaw.nl/edita](http://www.knaw.nl/edita)

### **Communication**

[www.ahc2005.org](http://www.ahc2005.org)

### **Low Countries Organisation Committee**

Onno Boonstra, Humanities computing, University of Nijmegen  
Leen Breure, Computer and Information Science, University of Utrecht  
Peter Doorn, Data Archiving and Networked Services (DANS), The Hague  
Jaap van den Herik, Computer Science, Universities of Leiden and Limburg  
Bart de Nil, Institute for Social History (Amsab), Gent, Belgium  
Paula Witkamp, European Commission on Preservation and Access, Amsterdam

### **Local organisation**

Michelle van den Berk, Berry Feith, Yamit Gutman and Annelies van Nispen  
(Netherlands Institute for Scientific Information Services – NIWI-KNAW)

### **Organizing institutions**

AHC: international Association for History and Computing  
KNAW: Royal Netherlands Academy of Arts and Sciences  
NIWI: Netherlands Institute for Scientific Information Sciences  
VGI: Low Countries branch of the Association for History and Computing  
SIKS: Dutch research school for Information and Knowledge Systems  
DANS: Data Archiving and Networked Services

# Preface

'History & Computing' and 'Humanities Computing' are in a crucial phase of development. The cultural heritage sector is turning digital, and ever more archival and other historically relevant sources are becoming available online. As a result there is a need for innovative methods and techniques to process the flood of digital resources. Currently, computer scientists show a keen interest in the information problems of cultural heritage and the humanities. Advances in grid computing and the Semantic Web are stimulating a new kind of e-Science, e-Culture, e-Humanities and even e-History.

This volume contains about fifty contributions presented at the XVIth international conference of the Association for History and Computing, that took place in Amsterdam from September 14th till 17th 2005.

The conference papers are intended for an audience of specialists from three broad fields:

- Scholars using computers in historical and related studies (history of art, archaeology, literary studies, etc.)
- Information and computing scientists working in the domain of cultural heritage and the humanities
- Professionals working in cultural heritage institutes (archives, libraries, museums) who use ICT to preserve and give access to their collections

The subject matter of these proceedings is primarily oriented at methodological issues. It is not restricted to one particular domain within history and the humanities. The papers included in this volume were selected by the conference committee and the session convenors. Due to time pressure, the papers could only superficially be refereed and marginally edited. Some papers that arrived too late to be processed could regrettably not be included in the volume. Such papers and the abstracts of poster presentations can be found on the conference website ([www.ahc2005.org](http://www.ahc2005.org)). A selection of the proceedings is being considered for publication in international journals.

The papers in this volume of proceedings can be characterized on the following characteristics:

- A first group of papers deals with *portals* and *gateways* to heritage information. More particularly, several papers on *virtual libraries* and *digital archives* are included.
- *Data enrichment* is the overall theme of the papers on *electronic text editing* and digital source editions. *Text analysis and retrieval* is a subject that has always received some, but not much attention from computing historians. We are happy that in this volume, a number of papers is dedicated to text analytical and ontological problems, partly inspired by the discussions

on the Semantic Web, but also influenced by participants from the field of literary and linguistic computing.

- *Images & multimedia* is one of the subject that attracts special attention from computing scientists. Papers on visual object detection and content-based artist identification show advances that are made in this area.
- *Geographical Information Systems* is a topic that has become remarkably 'hot' in historical studies over the past few years. The conference includes five sessions and 14 papers on historical GIS applications, ranging from building a historical GIS to time-space analysis and applications in urban history.
- *Quantitative data analysis* was one of the most important subjects in the early years of historical computing, but now it is attracting a relatively modest attention. Computer applications of statistics have become mainstream in social and economic history and apparently require less specific attention at AHC conferences.
- A number of papers deals with *digitization strategies* in heritage institutions and on the digitization of historical sources. However, the session convenors and other referees were fairly strict on referring 'me and my database' kind of papers to poster sessions, unless they clearly presented new methods of database design. The *XML markup language* offers a strong tool for the encoding of irregular source structures. A score of papers is dedicated to the role of XML in the structuring of heritage information.
- Large cross-sectional, nominative *databases in historical research* is a subject that might be called 'traditional' at AHC conferences. Two sessions and a handful of papers were dedicated to this subject.
- Finally, there is a number of papers with a *theoretical and methodological* component, in which *virtual networks* and collaboratories play a role. Moreover, several papers claiming *new approaches* to history and computing are included in these proceedings.

Peter Doorn  
*President of the AHC*

# Contents

- Alkhoven – *Digitizing cultural heritage collections: The importance of training* 7
- Alves – *Using a GIS to reconstruct the nineteenth century Lisbon parishes* 12
- Anderson & Healey – *Broadening the scope of electronic book publishing* 18
- Andersen & Erikstad – *Making a national census coding system internationally comparable* 23
- Berezhnoy, Postma & Van den Herik – *Computerized visual analysis of paintings* 28
- Bergboer, Postma & Van den Herik – *Visual object detection for the cultural heritage* 33
- Berger – *Microhistory and quantitative data analysis* 39
- Boot – *Advancing digital scholarship using EDITOR* 43
- Boughida – *CDWA lite for Cataloguing Cultural Objects (CCO): A new XML schema for the cultural heritage community* 49
- Breure – *PROGENETOR: An editorial framework for reuse of XML content* 57
- Broadway – *The early letters of The Royal Society 1657-1741: Managing diversity and complexity* 64
- Van den Broek, Kok, Hoenkamp, Schouten, Petieta & Vuurpijl – *Content-Based Art Retrieval (C-BAR)* 70
- Van den Broek, Wiering & Van Zwol – *Backing the Right Horse: Benchmarking XML editors for text-encoding* 78
- Brunnhöfer & Kropač – *Digital archives in a virtual world* 83
- Burkard – *Collaboration on medieval charters – Wikipedia in the humanities?* 91
- Burrows – *Reinventing the humanities in a networked environment: the Australian Network for Early European Research* 95
- Clausen – *Digitising parish registers – principles and methods* 100
- Delve & Healey – *Is there a role for data warehousing technology in historical research?* 106
- Doorenbosch – *Computer science and the Dutch cultural heritage* 112
- Fogelvik – *Large longitudinal, nominative databases in historical research* 119
- Garskova – *Towards a standard for MA programs in historical computing: (the experience of Russian and CIS universities)* 123
- Glavatskaya – *Indigenous peoples of the North-western Siberia: Ethnohistorical mapping* 126
- Gregory – *Creating analytic results from historical GIS* 131
- Gruber – *Occupational migration in Albania in the beginning of the 20th century* 136
- Heller & Vogeler – *Modern information retrieval technology for historical documents* 143
- Hoekstra – *Integrating structured and unstructured searching in historical sources* 149
- Ivanovs & Varfolomeyev – *Editing and exploratory analysis of medieval documents by means of XML technologies* 155
- De Jong, Rode & Hiemstra – *Temporal language models for the disclosure of historical text* 161
- Juola – *Language change and historical inquiry* 169
- Kröll – *Not ready for the Semantic Web: A field study of subject gateways on Contemporary History* 176
- Laloli – *Moving through the city: residential mobility and social segregation in Amsterdam 1890-1940* 182
- Lopes – *Historical geographic data dissemination through the web: the site Atlas and future developments towards its interoperability* 190

- Melms – *Reconstructing lost spaces. Affordably, that is* 194
- Mirzaee, Iverson, Hamidzadeh – *Computational representation of semantics in historical documents* 199
- Nagypál – *History ontology building: The technical view* 207
- Ordelman, De Jong, Huijbregts & Van Leeuwen – *Robust audio indexing for Dutch spoken-word collections* 215
- Pasqualis Dell Antonio – *From the roman eagle to E.A.G.L.E.: harvesting the web for ancient epigraphy* 224
- Perstling – *Layers and dimensions. The representation of complex structured sources* 229
- Petty – *Transnational histories in Roshini Kempadoo's ghosting. Cyber)Race identities.* 237
- Pieken – *Jewish life in Germany from 1914 to 2004 – The story of the Chotzen family* 243
- Robichaud – *The old Montréal heritage inventory database: Toward a renewed collective memory* 246
- Tschauner & Siveroni Salinas – *On the ground and '6 feet under'. Mobile GIS and photogrammetric approaches to building 3D archaeological spatial databases in the field* 250
- Valetov – *World museums on the Internet: A brief overview* 255
- Verheusen – *National digital repository for cultural heritage institutions* 263
- Voegler – *Virtual libraries and thematic gateways in German history: Strategies and perspectives* 267
- Weller – *A new approach: The arrival of informational history* 273
- Wiering, Crawford & Lewis – *Creating an XML vocabulary for encoding lute music* 279
- Wouters – *Writing history in the virtual knowledge studio for the humanities and social sciences* 288
- Zandhuis – *Towards a genealogical ontology for the Semantic Web* 296
- Zeldenrust – *DIMITO: Digitization of rural microtoponyms at the Meertens Instituut* 301

# Reinventing the humanities in a networked environment: The Australian Network for Early European Research

*Toby Burrows\**

The Network for Early European Research (NEER) is a new Australian initiative to broaden and deepen research in the field of medieval and early modern European studies. An integral part of the Network is a Digital Services Programme, which integrates various digital strategies designed to advance the Network's ambitious goals. This paper looks at the ways in which the Network is using information technologies to reinvent a traditional humanities discipline.

## **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 increasingly international, and 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. Research into the problem of soil salinity, for example, requires the involvement of soil scientists, environmental scientists, plant biologists, animal biologists, water engineers, and even historians.

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 vast deluge of scientific data, enabling researchers around the world to tap into an

international grid of digital research (Atkins 2003). In the United Kingdom, e-Science and e-Social Science programmes have been established to channel funding into the development of such infrastructures (Hey 2003). In Australia, the term e-Research has been used for the same type of initiative.

These trends are already evident in disciplines such as astronomy, where national and international groupings of researchers are being formed through the agency of 'virtual observatories' (Djorgovski 2004). By linking together shared data from observatories around the world and providing appropriate software tools to analyse and manipulate the data, astronomers are increasingly able to work together across national boundaries as part of a single global knowledge community.

This kind of approach is well-suited to scientific research, but its appropriateness for the humanities is yet to be tested. Of particular relevance and interest, then, is the Commission on Cyberinfrastructure for the Humanities and Social Sciences, established by the American Council of Learned Societies in 2004. The Commission is intended to go beyond the use of digital technologies as 'tools enhancing research methodologies', and to examine ways in which they can become 'a force creating environments that enable the creation of new knowledge' (ACLS 2004). A range of experts in the humanities and social sciences have been presenting the Commission with their ideas for the creative and innovative use of technology to transform their disciplines.

In Australia, the main government funding body for research in higher education institutions – the Australian Research Council (ARC) – has 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

---

\*Digital Services Director, Network for Early European Research, University of Western Australia

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. The ARC has also initiated an e-Research programme, but this is as yet on a fairly small scale. The funding for each Research Network is of the order of A\$1.5 million over a period of five years. Twenty-four networks were funded in 2004, mostly in the areas of science, technology and medicine.

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 crucial infrastructure which 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 officially 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 150 individuals are currently listed as Network participants. Their research covers all aspects of the culture and history of Europe in the Middle Ages and the Early Modern period, extending up to the initial European connections with Australia in the late eighteenth and early nineteenth centuries.

The Network has a range of institutional members including most of the larger Australian universities, such as Melbourne, Queensland and Sydney, which are all making a financial contribution to the Network. There are also a number of institutional 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 four main themes or research problems: Cultural Memory; Social Fabric; Science, Medicine and the Environment; and Early European/Australasian Connections. Each of these has a team leader, whose role is to coordinate Network activities and communication between researchers with an interest in the specific research area. Most of the academic activities of the Network are focused around these four 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.

While most of these are the type of activity in which academic researchers have been regularly involved, the difference in the Network's case is that they take place at a national level and within a comprehensive national framework. They are not simply reliant on university departments, learned societies or even individual researchers. The Network contributes financially, and also has an overall strategy for encouraging participation by postgraduate students and early career researchers. The other important difference is that these are only one strand in the Network's programmes. Its activities in the digital arena are where it aims to be most innovative and forward-looking.

### **Digital Early European Resources: Stage 1**

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. Known as Digital Early European Resources (DEER), this programme brings together various activities in the digital arena which have two main goals: to provide resources for the Network's participants, and to enable them to communicate more effectively with each other.

The Network's initiatives in promoting more effective communication among its participants are focused partly on such well-established mechanisms as mailing lists, discussion groups, and a database of participants. But other activities are more innovative: the Network has set up a scheme known as e-Consult, which enables postgraduate students to identify and contact senior researchers who are willing to provide advice. While the initial contact is made through the Network's Web site, and includes an agreement to

observe the protocols set down by the Network, subsequent communication is entirely confidential to the parties concerned.

Providing resources will be done in two main ways. The Network is working in partnership with two major commercial publishers of specialist databases to provide access for Network participants whose institution does not have a subscription. ProQuest is providing access to Early English Books Online (EEBO) while Brepols is providing access to a selected number of its full-text products. The main beneficiaries of this approach are researchers in smaller universities and regional universities.

At the same time, NEER is developing its own resources, beginning with a discovery service for identifying Early European artefacts, artworks and manuscripts in Australian collections. This involves a federated search service 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, the Network aims to avoid duplicating work already done in the cultural heritage institutions themselves, and prefers to harvest existing 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. It is, however, sponsoring a digitization programme on a small scale, by identifying significant objects which have not been digitized and working with holding institutions to digitize them. Among the models for this resource discovery service are PictureAustralia and MusicAustralia, both of which are produced by the National Library of Australia (Ayes 2004).

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 the Network's Web site, as well as being available through journal packages like Project Muse. NEER provides publishing and subscription management services for this electronic version of *Parergon*.

Another area in which the Network has a keen interest is the training of postgraduate students and early career researchers in specialist skills relevant to Early

European research. The Network's digital programme is contributing to this goal by establishing internships in partnership with publishers of electronic text collections. Postgraduate students are given training in text encoding with markup languages like TEI as well as in database management and the development of Web sites. This training is closely linked to the use of editorial and research skills to create and review specialist content for databases in the fields of medieval and Early Modern history and literature.

### **Digital Early European Resources: Stage 2**

While these initial activities in the Network's digital services programme are integral to its success, they are not designed to be technically innovative, nor are they expected to promote transformative cultural change in humanities research. They are intended to harness and integrate a wide range of proven uses of technology. The next stage in the Network's development, however, is designed to use information technologies to reinvent and transform the discipline in a more innovative and fundamental way.

One of NEER's main priorities in this second stage is to develop its own digital repository for the research output of its participants. To date, most digital repositories of this kind have been institutional – that is, they reflect the output of a single university (Lynch 2003). While there has been considerable support and enthusiasm at the level of university policy and in university libraries for establishing these repositories, it has to be said that this approach has not been particularly successful in practice. Most institutions have had difficulty acquiring a sufficient body of material for their repository.

NEER is working with the ARROW Consortium (Australian Repositories Online to the World), based at Monash University in Melbourne, to develop a repository which is discipline-based and national in scope. ARROW has been funded by the Australian Government as part of its Research Information Infrastructure Framework for Australian Higher Education. Its goal is to identify and test software to support best practice institutional digital repositories (Payne 2005). ARROW is using the Fedora architecture and software, initially to manage such materials as article preprints and postprints, theses and other print equivalents. The NEER repository will not be limited to articles and papers, though. It will also aim to collect and make available the underlying research data from these stud-

ies, whether in the form of databases, spreadsheets, correspondence, images, sound files, maps, or other formats. NEER's expectation is that researchers will see the value of contributing their research output to a 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 which is required for such approaches to be successful.

Within this repository, the Network is focusing particularly on metadata content and standards. While various institutions and projects, including ARROW, have worked on metadata schemas and on the mapping of data from one schema to another, comparatively little work has been done on the content and standards for repository metadata. Most of the subject access to repositories is at a fairly simple and summary level, and researchers depositing their papers are given only a limited choice of subject terms for describing the content. NEER is investigating methods for embedding sophisticated subject ontologies into the repository framework, by bringing together the specialist disciplinary knowledge of the participants in the Network and existing vocabularies for subject access in indexing databases like the *International Medieval Bibliography*.

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 this work on metadata and subject vocabularies should link into the broader framework of the Semantic Web.

The Network is also very interested in exploring how the existing repository framework might be used in transforming the way 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. This underlying goal is not significantly different from that of research libraries, though the communication system on which they are based does not depend so much on publishers and journals. The immediacy and availability of research results are improved, but the effect of repositories on the research process itself is not transformative.

NEER is aiming to test the integration of repositor-

ies with other emerging technologies in an effort to design new structures for communicating research in the humanities. It aims to make room for younger researchers especially to experiment with new approaches. Among the components of this structure will be blog-like narratives of research activities, seen as a continual work in progress, but designed to report research results. These will link 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 Early European research group.

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 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. Without this, any new structures for disseminating research are unlikely to rise much above vanity publishing. 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, NEER is aiming to contribute to the design of 'a next generation system for scholarly communication' (Van de Sompel 2004) which goes beyond Open Access, self-archiving and electronic journals. Such a system will need to redefine the 'unit of communication' and provide a more flexible way of 'registering' a communication unit. 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. In the initial stage of the Network, these are well-established and widely used already, but others are new approaches

which are intended to transform the way in which research is communicated in the humanities.

While the activities of the Network include software development and technical innovation, this is not the primary focus. Nor is the Network emphasizing technical solutions to the management and manipulation of large datasets along the lines of the British e-Science programme. Instead, the Network's 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.

NEER offers a unique opportunity to develop and test new technologies for scholarly communication and the dissemination of research, within the framework of a new government approach to funding academic research communities. This digital programme will be crucial to the Network's success in reinventing Early European research and demonstrating the applicability of new technologies to fundamental changes in the humanities.

## References

- ACLS. 2004. Commission on Cyberinfrastructure for the Humanities & Social Sciences. *Charge to the Commission*. New York: American Council of Learned Societies. [http://www.acls.org/cyberinfrastructure/cyber\\_charge.htm](http://www.acls.org/cyberinfrastructure/cyber_charge.htm)
- Atkins, Daniel E. et al. 2003. *Revolutionizing Science and Engineering through Cyberinfrastructure: Report of the Blue-Ribbon Panel*. Arlington, VA: National Science Foundation. <http://www.cise.nsf.gov/sci/reports/toc.cfm>
- Ayres, Marie-Louise, Toby Burrows and Robyn Holmes. 2004. 'Sound footings: building a national digital library of Australian music' in: *Research and Advanced Technology in Digital Libraries*, ed. Rachel Heery and Liz Lyons (*Lecture Notes in Computer Science 3232*) (Berlin: Springer-Verlag, 2004), 282-291.
- Djorgovski, S. G. 2004. 'Virtual Observatory, Cyber-Science and the Rebirth of Libraries', ARL/CNI Forum on E-Research and Cyberinfrastructure. <http://www.arl.org/forumo4/djorgovski.html>
- Hey, A. J. G. and A. E. Trefethen. 2003. 'The Data Deluge: An e-Science Perspective', in: *Grid Computing – Making the Global Infrastructure a Reality*, ed. F. Berman, G. C. Fox, and A. J. G. Hey. (London: Wiley, 2003), pp. 809-824.
- Lynch, Clifford A. 2003. 'Institutional Repositories: Essential Infrastructure for the Scholarship in the Digital Age', *ARL Bimonthly Report* 226. <http://www.arl.org/newsltr/22/ir.html>
- Payne, G. J. 2005. 'Australian Research Repositories Online to the World – ARROW', EDUCAUSE Australasia, Auckland, April 2005.
- Van de Sompel, Herbert, et al. 2004. 'Rethinking Scholarly Communication: Building the System that Scholars Deserve', *D-Lib Magazine* 10 (9), Sept. 2004.