

BEYOND SOUTH SEAS: MAKING HISTORY IN NETWORKED DIGITAL TECHNOLOGIES

PAUL TURNBULL, SCHOOL OF HUMANITIES, GRIFFITH UNIVERSITY
p.turnbull@griffith.edu.au

with

MARK FALLU, RESEARCH COMPUTING SERVICES, GRIFFITH UNIVERSITY
m.fallu@griffith.edu.au

Paper for eRA AHCH Workshop, 3 October 2008

In this paper, I discuss some of the more salient intellectual and technological dimensions of work over the past year, focused on developing an open source knowledge creation, management and publication system. In key respects, our work seeks to anticipate developments in national collaborative e-research infrastructure over the next five or so years.

Especially in view of recent statements on innovation policy by the Australian government, we can expect the next five or so years will see significant advances in the development of online knowledge repositories for not only more complex kinds of quantitative research data, but also for qualitative data in rich and diverse media forms that will offer new possibilities for humanities research. We will also see improved or new middleware, allowing Australian research communities in the humanities collaboratively to create, share and interrogate new knowledge of cultural and social phenomena. However, if humanities researchers are to exploit these and other possible advances in digital research infrastructure, then what they will also need are 'tools' enabling the creation, reception and use of knowledge that these infrastructural advances can put into intellectual circulation. They will need the means of using networked digital technologies as primary media for research, and to publish their findings as complex multimedia artifacts.

Currently, I am working with Mark Fallu, of Griffith's Research Computing Services unit, on building one such 'tool' for humanities researchers. The system we are creating could be characterized as a RESTful web publishing platform operating in conjunction with a collection of lightweight web services and visualization tools.¹ Both the tools and the content that they support will be capable of being incorporated into other websites and applications, independently of their use as part of our system. All of the components of the system consist of open-source tools that may be freely implemented and modified by the academic community. The most critical component of the system is the Plone content management system. Plone is built on top of the Zope web application framework using the Python programming language.²

Other major specifications that have been selected for implementation as part of the system include the Object Reuse and Exchange (ORE) specification, recently promulgated

¹ Here, we are particularly indebted in our thinking to Fielding, Roy, *Architectural Styles and the Design of Networked-based Software Architectures*, PhD Dissertation, UC Irvine, 2000, <http://www.ics.uci.edu/~fielding/pubs/dissertation/top.htm>

² <http://plone.org/>

by the Open Archives Initiative, the Open Access Initiative Protocol Metadata Harvesting (OAI-PMH), the Resource Description Framework (RDF), and the International Committee for Museum Documentation (CIDOC) Conceptual Reference Model (CRM). Together with these specifications that are relatively well known to the e-research community, the project is also implementing a number of minor specifications to encode specific data types including, 'micro-formats', GeoJSON and COINS.

Plone was chosen for being standards-based, technically mature, having inbuilt advanced functionality, an extensive user community and a wide range of well-designed ancillary programming modules, notably for enabling 'user-friendly' web-based authoring and scholarly editing in an XML environment, and geo-temporal knowledge representation (a critically important capability for researchers in many humanities disciplines). We were also attracted to Plone because of its use and adaptation by various research communities in the sciences, humanities and creative arts around the world. Importantly, in Australia, it has been used with promising results to prototype ways of supporting virtually collaborative knowledge creation, management and that emulate research practices in many humanities disciplines.³

THE LESSONS OF SOUTH SEAS

Our research and development path draws heavily on my experiences in creating online scholarly resources since the mid-1990s. In particular, we are seeking to build on the successes, and also overcome the shortcomings, of *South Seas* - a web-based resource focused on James Cook's momentous first Pacific voyage aboard the Endeavour of 1768-1771.⁴ Consequently, what follows takes the form of my recounting in some detail how the intellectual aims and technological underpinnings of *South Seas* evolved between late 1999 and early 2004 - interspersed with brief descriptions of the key functionalities of the knowledge creation, management and publication we are currently building. This may seem a cumbersome way of proceeding. However, it seems the best way to underscore a vital point: while there is much to be learnt from the uses to date of digital technologies in the development of e-research infrastructure and its uses in disciplinary communities beyond the humanities, the integration of digital technologies within the humanities is in its infancy; and if the integration of these technologies is to be successful it must be grounded in a rich ethnographic understanding of not only the commonalities, but also the peculiarities of disciplinary practice in the humanities.⁵ Building humanities e-research capability is not simply a matter of developing tools and other infrastructural elements enabling the creation, dissemination and re-use of digital 'data'. That infrastructure must simulate, as far as is feasible, the intellectual aims and

³ For example, the Data Acquisition, Accessibility and Annotation e-Research Technologies (DART) project undertaken at Monash, the University of Queensland and James Cook University.

⁴ *South Seas* resides on a server in the National Library of Australia's network at: <http://www.southseas.nla.gov.au>.

⁵ My thinking on this point is informed by my research over the past decade on the intellectual practices and products of racial science research communities in nineteenth-century Britain and colonial Australia. However, recently, I have become aware of fascinating work by Susan Leigh Star. See, for example, her article, 'The Ethnography of Infrastructure,' *American Behavioral Scientist*, 43: 377-391, <http://abs.sagepub.com/cgi/content/abstract/43/3/377>

practices of Australian humanities researchers within the networked environment. In other words, what the humanities require are what the late Rob Kling nearly a decade ago termed 'human-centered' systems – that is, systems that 'work well for people and help support their work, rather than make it more complicated.'⁶

That was essentially the thinking informing the development of *South Seas*. More specifically, I sought to go some way towards developing a key capability for historians wanting to work in digital media - scholarly editing standards for the creation and interpretation of historical texts and images with networked digital technologies.

INTO SOUTH SEAS

South Seas was created over some four years by myself and co-researcher Chris Blackall, with the assistance of Dr Christine Winter.⁷ All three of us were then associated with the Centre for Cross-cultural Research at the Australian National University (ANU). The Centre was established with Australian research Council (ARC) funding in the late 1990s with a brief to share the outcomes of research into various aspects of anthropology and cross-cultural history with wider publics in a range of media, including exhibitions, film and multimedia. I was appointed a senior research fellow to develop the Centre's capability to use networked digital technologies as media for research and the dissemination of research findings.⁸ Blackall joined the Centre as a graduate student and has since been closely associated with the development of Australia's national collaborative electronic research infrastructure at the ANU.

I am a historian with long standing interests in the evolution of eighteenth-century European ideas of societal development.⁹ Around 1997, I had begun studying the ethnographic reportage produced during Cook's Pacific voyaging, and was particularly intrigued by the many obvious and also the many subtle differences in how places, people and events occurring during the course of the *Endeavour* voyage were reported in the various journals kept by its leading participants. The most important of these journals were the daily record of occurrences that Cook kept for submission to the Admiralty on his

⁶ Rob Kling, 'What is Social Informatics and Why Does it Matter?' *D-Lib Magazine*, vol. 5, 1 (1999), <http://www.dlib.org/dlib/january99/kling/01kling.html>

⁷ Dr Winter is now a researcher associated with the ANU and the Australian War Memorial whose work analyses the legacies of the German empire in the Asia-Pacific region.

⁸ I am currently a professor of history at Griffith University, where I continue my interests in the theory and practice of making history in networked digital media.

⁹ Stemming from research interests in the historical sociology of Edward Gibbon (1737-1794). See Turnbull, Paul. 'Gibbon's Exchange with Joseph Priestley', *British Journal for Eighteenth-Century Studies*, 14 (1991); "'Une marionnette infidèle': The Fashioning of Edward Gibbon's Reputation as the English Voltaire", in David Womersley, ed., with John Burrows and John Pocock, *Edward Gibbon: Bicentenary Essays, Studies on Voltaire and the Eighteenth Century* 355 (Oxford: the Voltaire Foundation, 1997); 'A Forgotten Cosmogony: William Hull's 'Remarks on The . . . Aboriginal Natives'?' *Australian Historical Studies* 24, no. 95 (1990): 207-20.

return,¹⁰ the private journal kept by Joseph Banks,¹¹ who accompanied Cook on the voyage, afterwards soon becoming one of the most important figures in late eighteenth century British scientific circles; and there was also the journal of Sidney Parkinson, an artist and draftsman employed by Banks who was to die on the return leg of the voyage.¹² I was also struck by how the official account of the voyage, produced by John Hawkesworth, a minor London writer, and published in 1773, purports to tell the story of the voyage through Cook's eyes, but is actually a narrative woven from passages in both Cook and Banks's journals, supplemented in various parts by testimony deriving from other personnel on the voyage.¹³

Given the extensive length and detailed nature of these sources, plus the fact that in manuscript and print form, they were only readily accessible to a handful of well resourced scholars, I came up with the idea of presenting them in an online edition that would enable easy comparison of Cook, Banks and Parkinson's differing observations of occurrences on the voyage in the immediate aftermath of their experiencing them. Not only this, I could see it would be valuable to interrelate these records with an electronic version of the complete text of Hawkesworth's account of the voyage, allowing readers to compare what was recorded during the course of the voyage with what the public were subsequently told.

I envisaged this online edition of the principal records of the *Endeavour* voyage as having scholarly notes, annotations and commentaries comparable to those of print-based scholarly editions of important historical texts. But soon after I began exploring the creation of an electronic transcript of the holograph manuscript of Cook's journal, held at the National Library, I came to think that digital technologies might allow more adventurous things to be done by way of historically contextualizing the journal and other key documents of the *Endeavour* voyage.

By the late 1990s, the National Library of Australia was committed to making digital surrogates of important images and maps relating to the history of Pacific exploration

10 Cook, James, J. C. Beaglehole, and R. A. Skelton. *The Journals of Captain James Cook on His Voyages of Discovery*. 4 vols. Cambridge: Published for the Hakluyt Society at the University Press, 1955, vol.1: *The Voyage of the Endeavour, 1768-1771*. Cook's holograph manuscript is preserved as National Library of Australia MS 1. See http://www.nla.gov.au/collect/treasures/mar_treasure.html

11 Banks, Joseph, and J. C. Beaglehole. *The Endeavour Journal of Joseph Banks : 1768-1771*. 2 vols. Sydney: The Trustees of the Public Library of New South Wales in association with Angus and Robertson, 1962. The State Library of New South Wales electronic edition of the journal can be found in their online collection of the Papers of Joseph Banks at http://www2.sl.nsw.gov.au/banks/sections/section_01.cfm Again, we acknowledge all the support given to *South Seas* by Alan Ventress, then Mitchell Library and now Director of the New South Wales Records Authority.

12 Parkinson, Sydney. *A Journal of a Voyage to the South Seas, in His Majesty's Ship, the Endeavour Faithfully Transcribed from the Papers of the Late Sydney Parkinson, Draughtsman to Joseph Banks, Esq. On His Late Expedition, ... Round the World. Embellished with Views and Designs*. Place Published: printed for Stanfield Parkinson, the editor: and sold by Messrs. Richardson and Urquhart; Evans; Hooper; Murray; Leacroft; and Riley, 1773. A facsimile edition of the journal was published in 1984 by Caliban Books in 1984.

13 Hawkesworth, John. *An Account of the Voyages Undertaken by the Order of His Present Majesty, for Making Discoveries in the Southern Hemisphere, and Successively Performed by Commodore Byron, Captain Wallis, Captain Carteret and Captain Cook, in the Dolphin, the Swallow, and the Endeavour : Drawn up from the Journals Which Were Kept by the Several Commanders, and from the Papers of Joseph Banks, Esq.* London: Printed for T. Strahan and T. Cadell, ... 1773. There is no modern edition of this work.

freely accessible via the internet. Already many digital copies of key images relating to Cook's voyaging held by the library had been put online, and there were plans to digitize many more.¹⁴ This led me to explore with library staff the feasibility of creating a digital edition of the *Endeavour* journals that would virtually incorporate these pictorial resources.

For its part, the National Library was keen to investigate how new kinds of digital scholarly artifacts - such as an online edition of the *Endeavour* journals - might use online surrogates of images and maps deriving from the voyage held in their collections. As I have discussed elsewhere, the library was one of the first of the world's major libraries to explore the potential of the networked digital communications for overcoming barriers to accessing its collections.¹⁵ Its involvement in exploring the potential of networked digital technologies actually predates the creation of the World-Wide-Web by several years. Indeed, having kept an informed eye on the growth of Australian online publication and web-sites, the library in 1996 established PANDORA, a digital archive enabling the collection and long-term public access to those of these new kinds of information artifacts judged to be culturally significant knowledge resources for Australians.¹⁶

By 1998, the library was also aware of early experiments in e-learning and ventures that we have now come to term e-research. It could see that it and other major Australian libraries would soon need the capability to collect and provide ready access to complex web-based resources created by Australian researchers in the humanities, creative arts and social sciences. And this posed numerous challenges, especially if the resources were to be curated and made available in ways that did not see their value as knowledge resources diminish as technologies for the creation, dissemination and preservation of knowledge in digital forms evolved.

This was the background to myself and specialist staff of the library agreeing to collaborate on producing a web-based edition of the key *Endeavour* voyage texts. They sought and eventually secured funding for the project from the ARC through the precursor to its current Linkage Projects Scheme.¹⁷ Providing free public access to these important historical texts would be a major contribution to knowledge of Australian and Pacific history. However, the venture was also seen as providing a focused means of identifying what technical approaches and standards historians would need to employ should they wish to create scholarly editions of historical documents that, much like books in conventional libraries, were to be the basis of future online repositories of historical knowledge.

¹⁴ As can be seen from searching on Pacific themes in the National Library of Australia's online Pictures Catalogue, <http://www.nla.gov.au/catalogue/pictures/index.html>; and online maps collection at <http://www.nla.gov.au/digicoll/maps.html>

¹⁵ Turnbull, Paul. 'The Network and the Nation: the Development of National Bibliographical Resources', in P. Cochrane (ed.), *Remarkable Occurrences: the National Library of Australia's First 100 Years, 1901-2001*, Canberra: National Library of Australia, 2001, pp. 255-71.

¹⁶ Preserving and Accessing Networked Documentary Resources of Australia (PANDORA), <http://pandora.nla.gov.au/about.html>

¹⁷ ARC Strategic Partnership with Industry - Research and Training (SPIRT) scheme grant: 0002126: Dr PG Turnbull Dr J Pearce Mr P Gatenby Mr C Law: The Endeavour Project: creating and implementing scholarly standards for the preparation and publication of historical editions in digital form (National Library of Australia and Australian National University).

NAVIGATING DIFFICULT WATERS

An initial ARC application was unsuccessful. It probably did not inspire confidence in the assessing panel that one anonymous expert reader bluntly put it that the team did not have a clue what they were talking about: there could never be ways of replicating scholarly standards for online editions of scholarly works in the virtual environment.¹⁸ The best that could be achieved they argued was perhaps some refinement of electronic editing techniques for the preparation of print-based historical documents; and while conceding this might be a useful outcome, this reader couldn't see how this entailed original research - the primary criteria for funding by the ARC through its linkage scheme.

This distinction drawn between research and research infrastructure may have made sense in the world of print-based scholarly communication, but in the digital world it has become a problem yet to be squarely addressed, let alone resolved. As a recent planning paper released by Project Bamboo observes, the evolution of e-research infrastructure for the humanities is currently at the point that scholars who want to work with digital technologies as research media can anticipate spending two-thirds of their time in data creation and its translation into forms enabling them to spend the remaining third of their time on things of intellectual value.¹⁹ This predicament is now recognized by peak scholarly bodies such as the Australian Academy of the Humanities;²⁰ and over the past four or so years there have been a number of initiatives aimed at securing researchers in the humanities the infrastructure to shift the balance in favor of intellectual engagement with research data.²¹ However, this shift will continue to be slow and risks being impeded by disciplinary communities in the humanities proving unreceptive to rethinking what constitute research practices and outcomes in the post-Gutenberg academy.

For my part, it seems puzzling that it has been so hard for colleagues to treat significant digital resources as equivalent for the purposes of research funding and professional advancement to scholarly articles and books. The test of their value is surely whether peer appraisal finds they constitute an original contribution to knowledge. I continue to encounter colleagues in my discipline ready to point out that treating a database as a research outcome is akin to rewarding the creation of a card index system, rather than the book it helped create. As yet they fail to be persuaded that many of the 'databases' that humanities researchers are now creating are complex knowledge objects designed to be used by other researchers in Australia and internationally to generate new insights into various social and historical phenomena. Moreover, while they are quick to argue that humanities disciplines have much to contribute to scientific and technical innovation by

¹⁸ Two other reviewers were much more positive, but were clearly historians with little understanding of developments in digital librarianship and electronic scholarly text editing.

¹⁹ Project Bamboo is Bamboo is a new international effort to advance the humanities through the development of shared digital technologies. The Bamboo Planning Project Paper can be accessed via <http://projectbamboo.org/>

²⁰ The Academy is actively seeking to assess the e-research needs of humanities researchers. See <http://www.humanities.org.au/Policy/HumTech/> and its submission to the 2008 National Collaborative Research Infrastructure Strategy (NCRIS) Roadmap Review, <http://www.humanities.org.au/Policy/default.asp>

²¹ These have included several annual national conferences bringing together humanities researchers using digital technologies, and the creation of more focused networks, such as the Australia New Zealand Digital Encyclopedias Group (ANZDEG). More recently there has been the Humanities, Arts and Social Sciences Working Group of the NCRIS Roadmap Review. See <http://ncris.innovation.gov.au/>

enhancing public understanding of its benefits and risks, they have little sense that digital resources, made freely available via the internet, might prove as much if not more of a public good than a book.

Currently, there are few signs that the status of digital work is being discussed within Australian disciplinary communities in the humanities. The only discussion relating to digital technologies currently exercising the minds of Australian historians, for example, is what ranking various print-based journals should have in the database of the Australian government's proposed Excellence in Research for Australia (ERA) initiative. Indeed, discussions around ERA have so far ignored the question of whether an evaluative framework for judging the quality of Australian research should encompass the appraisal of excellence in humanities research taking the form of digital artifacts.

Even so, as previously mentioned, the value of embodying digital technologies in humanities research is now recognized by bodies like the Australian Academy of Humanities; and meeting the infrastructural needs of the small but growing community of humanities scholars working with these technologies is accepted as being an integral part of the National Collaborative Research Infrastructure Strategy (NCRIS). Thus it may well be that the combination of continued advocacy by the Academy and the creation of key elements of humanities e-research infrastructure will prove to be catalysts for disciplinary discussions of how to evaluate the worth of research projects that exploit these advances in e-infrastructure.

While many things remains uncertain, the environment for humanities researchers integrating digital technologies in their practice is markedly different than it was nine or so years ago. Then, there were only a few Australian academic colleagues who could understand why, in the case of *South Seas*, a historian would want to create a web-based resource, to say nothing of spending no small amount of time assessing developments in the fields of digital librarianship and information science, and then training themselves in the use of various softwares and programming tools. Indeed, by the time a second ARC application was successful, I was finding it difficult to convince senior colleagues at the Centre for Cross-cultural Research that its investment in supporting the creation of an online edition of the *Endeavour* journals would be only be realized if the edition employed what was emerging by way of international consensus on techniques and standards for the creation and management of digital editions of historical texts. In fact his ability to bring the venture to a successful conclusion began to be questioned.

Colleagues accustomed to working solely in print-based media could not see that a good deal had in fact been achieved prior to gaining ARC funding. Selected parts of the *Endeavour* journals and specimen multimedia objects had been used in the trial phase of TransAct, Canberra's public high-speed broadband network. Several prototypes of the edition had been mounted on open test websites, and I was being invited to speak about this work at several international conferences. Indeed, the fact that prototype work could be viewed by assessors, and several papers on the venture had been published, probably had significant weight in the ARC's decision to fund the venture. And this, incidentally, raises a further issue needing to be addressed in current discussions on Australian humanities e-research: if, in the future, there is greater recognition by the ARC funding agencies of the principal outcomes of humanities research taking digital forms, then there is much to be said for allowing prototyping work to be submitted as part of the assessment process. Indeed, my experiences over the past decade lead me to think that as funding agencies are seeking to maximize the public benefits of research, it is worth considering

making it a prerequisite for funding major digitally-based projects that they should employ - as far as is practicable - programming and information standards best ensuring the wide circulation and use of the knowledge they offer through their integration within national collaborative e-research infrastructure.

But to return to *South Seas*. Despite much being learnt during the first two years of the project, progress was uncomfortably slow. Prototyping raising as many problems as pointing to workable solutions to presenting large amounts of complex historical information in the web-based environment. Progress was also slowed by the fact that Centre support was limited to the provision of research assistance in the preparation of electronic transcripts of key texts. Prior to securing ARC funding, there was neither Centre personnel nor funding to employ people with the necessary expertise in programming or information management. Indeed, even after ARC funding was won it proved extremely difficult to find anyone with suitable expertise who was not employed in creating information infrastructure for the private sector and government on a salary way beyond what a university could offer. The situation today, incidentally, is little better; and any strategic development of e-infrastructure for the humanities will depend heavily on resolving this problem.

Further complicating matters, Chris Blackall and I had come to see the best approach to presenting the *Endeavour* journals online would be to develop some way in which these texts and accompanying annotations could be marked up according to the extensible markup language (XML) based document type description (DTD) developed by the Text Encoding Initiative (TEI), an international consortium of scholarly editors and librarians established to develop and maintain standards for representing texts in digital forms.²² What is more, feedback from users of early online prototypes of the edition not only confirmed the intellectual value of interrelating these texts on a chronological basis, but also led us to think about also giving users the ability to explore these texts in space as well as time. And here, Blackall was aware of how XML-based ways of representing geo-spatial data were being promisingly developed by researchers in geography and environmental sciences. It seemed possible to use GIS data for the regions sailed by Cook to create a series of XML-based maps on which the track of the *Endeavour* was plotted, using Cook's daily measurements of latitude and longitude and best estimates of what we now know his actual location was. The maps could then be interlinked via XML processes with the *Endeavour* texts, scholarly annotations and also external resources such as the digital collections of voyaging imagery and maps held by the National Library.

However, in 2000, XML-based programming was still in its infancy in various important respects. Tools for creating versions of the *Endeavour* journals marked up in TEI XML were little better than line editors with simple macros for inserting commonly used tags. There also appeared to be little consensus about the process of interrelating content in different documents. Various ways of transforming XML documents into XHTML were being championed; and trials of these translation processes revealed that they were all cumbersome in serving users large numbers of richly interlinked web-pages. So, in the end, our use of XML text encoding in *South Seas* was confined to its use as a preservation strategy. As a result of our collaboration with AUSTHC (discussed below), the system ultimately used to create, edit and publish what currently exists on the *South Seas* website has the capacity to export the editions of the *Endeavour* journals and other full transcripts of historical texts in a rudimentary TEI XML format.

²² See <http://www.tei-c.org/index.xml>

XML and TEI standard has of course now greatly matured. Early versions of the TEI specification were primarily concerned with the production of faithful digital surrogates of documents. TEI was used to transcribe the content of a document and describe how that content was structured within a document.

As the standard has evolved to allow richer and more comprehensive descriptions of documents the TEI specification has seen the addition of elements that allow for the encoding of information not just about the document, but about the world. The inclusion of the person element (TEI P5, sec. 20.4.2) is an example of the new ability to create a semantic mapping between the contents of documents and the world.

Hence a key goal of our current research and development is to build a system allowing for the storage, annotation, processing and retrieval of historical documents, together with associated scholarly commentaries and essays that are formatted in (ideally) version 5 of the TEI XML-DTD. This gives the system a powerful ability for researchers to make assertions about the historical or other meanings of materials within it.

The 'user front-end' of the system will be web-based 'tools' for collaborative authorship, editing, analysis and citation of network accessible resources.

Essentially we are developing a process by which texts can be marked up in accordance with the TEI guidelines to a high degree of structural and/or semantic granularity, then ingested into the system after being prepared using one of numerous currently available XML editors, or exported from word-processing software such as Open Office using TEI XML plug-ins. What is more, we are working on using XML micro-formats to enable unstructured texts to be ingested into the system via a web-interface so that they are automatically encoded into TEI, and added to the system as discrete objects, or incorporated into the structure of an existing document already marked up in TEI - without compromising that document's structure. Our ultimate goal is to enable texts to be 'round-tripped' within a scholarly editing work-flow by one or more users in different locations with the necessary version controls. We learnt recently that New Zealand's Electronic Text Centre (NZETC)²³ has already made significant progress on achieving this, using somewhat different programming techniques; and we will be seeking either to adopt or emulate elements of their approach to the task.

Experimentation with XML during the course of building *South Seas* also led us to appreciate that few humanities researchers beyond those preparing scholarly editions of historical or literary texts would want to invest time and intellectual energy in becoming adept at fine-grained TEI-style markup. So our aim now is not only to devise ways in which text with little or no structure can be automatically marked up at the point of being cut and pasted in our system, but also to provide some simple means by which common, structurally or semantically important features of a document can be marked-up and, if required, interrelated to relevant parts of other documents.

²³ See <http://www.nzetc.org/tm/scholarly/tei-NZETC-About-technology.html>

COLLABORATION WITH AUSTHC

Early in 2000, Chris Blackall and I fell into conversation with Gavan McCarthy and Joanne Evans at the Australian Science and Technology Heritage Centre (AUSTHC) at the University of Melbourne. AUSTHC had been established in 1999, to create and provide online resources for the study of Australian science and technology. Indeed AUSTHC came about in large part because since 1994, McCarthy, an archivist, and Evans, a skilled programmer, had proven the scholarly value of web-based resources by building Bright SPARCS, somewhat modestly described as a register of people involved in the development of Australian science and technology.²⁴ Within a year or so of existing, AUSTHC had begun to create what is now a wealth of authoritative knowledge resources of Australian scientists and scientific institutions that are not only routinely used by scholars, but also by university and school students and many members of the public. Taken together, these resources, built by AUSTHC personnel in collaboration with various Australian research communities, now form an indispensable part of the nation's emerging e-research infrastructure for the humanities, social sciences and creative arts in Australia.

Perhaps the most important aspect of AUSTHC's research and development of online resources has been the creation and gradual refinement of a suite of database systems for knowledge management and web publishing for archivists, museum curators and researchers. Information about these systems and examples of their use can be found on the AUSTHC web-site and also several papers published by McCarthy and Evans.²⁵ Suffice it to say here that conversing with McCarthy and Evans resulted in us deciding to collaborate with AUSTHC on what was essentially customized development of two of their suite of tools, the Online Heritage Resource Manager (OHRM) and the Web Academic Resource Publisher (WARP), into a knowledge management and publication system for the online edition of the *Endeavour* journals - which by this time we had become accustomed to calling 'South Seas'.

Looking back on *South Seas*, the mastery over creating richly interrelating 'content' provided by the OHRM / WARP system enabled a favorable shifting of the 'Bamboo balance' previously mentioned towards intellectual work.

Obliged to re-transcribe the entire text of the Cook journal, I had gained an intimate appreciation of the scholarly achievements of J.C. Beaglehole in producing what remains the definitive scholarly edition of Cook's journals; but it also led me to think the Beaglehole's editorial work could be usefully supplemented in two respects. Firstly, I was struck by how Beaglehole's edition of the Cook journals appeared to take for granted that its readership possessed at least some knowledge of eighteenth-century sailing and navigational practices. It seemed worth providing a fairly extensive glossary of common

²⁴ <http://www.asap.unimelb.edu.au/bsparcs/> What is especially impressive about Bright SPARCS has been its use in school science and history curricula. See http://www.asap.unimelb.edu.au/bsparcs/guides/t_teachers.htm

²⁵ See links relating to tools at <http://www.austehc.unimelb.edu.au/> and various papers at <http://www.austehc.unimelb.edu.au/ohrm/>. Also, importantly, McCarthy, Gavan and Evans, Joanne. 'The Open Resource Scholarly Network: new collaborative partnerships between academics, libraries, archives and museums' at <http://www.vala.org.au/vala2002/2002pdf/15McCEva.pdf>.

published in 1767 by William Falconer (1732-1770),²⁶ a Scots seafarer and poet best remembered for his lengthy poem, *The Shipwreck* (1762). Rather than compile a new glossary, I decided to create an electronic version of the complete text of Falconer's encyclopedia, put online so that readers of any of the *Endeavour* journals could search this work from whatever page of the journals they were reading.

Secondly, Beaglehole said relatively little in notes to his edition of the *Endeavour* journal about the life-ways and culture of the Indigenous peoples of Oceania whom Cook encountered. Partly, this was due to the limitations on the amount and scope annotation and commentary in any print-based edition of the Cook journals; but more importantly, ethnographically informed research into the life-ways and cultures of Pacific peoples prior to European contact only began in earnest after Beaglehole completed his edition. Douglas Oliver's great work on pre-contact Tahitian society, for example, was not published until 1975.²⁷ Similarly, the re-evaluation of Indigenous Australian history from the time of European Invasion did not begin in earnest until the 1970s.²⁸ And in Aotearoa New Zealand, Maori history in the era of Cook's voyaging was to remain dimly understood in many respects until the first installment of Anne Salmond's remarkable re-evaluation of exchanges between Maori and Europeans during the long eighteenth century appeared in 1991.²⁹ Any online edition of the *Endeavour* journals would do well to draw upon this wealth of cross-cultural scholarship.

The best way to do this, given what this scholarship had to say about the complexities of cultural phenomena encountered during the Cook voyages, was to provide a companion series of essays - a virtual equivalent to the encyclopedic guides to various aspects of social and cultural history that many scholarly publishers were producing by the late 1990s. These essays could sit within *South Seas* as a stand-alone 'Companion' yet be made readily accessible to readers via links in the relevant pages of the *Endeavour* journals.

Moreover, as the OHRM system had been originally designed to create biographical entries on scientists and scientific institutions to which additional 'pages' could be added, in which images could be displayed and important bibliographical sources listed. It required little additional programming so that *South Seas* could offer essays on various aspects of eighteenth century European voyaging and Oceanic Indigenous cultures to which pages were linked offering recommended further readings and, importantly, stable virtual pathways to images and maps either digitized as part of *South Seas*, or put online by the National Library or other major cultural institutions.

26 Falconer, William. *An Universal Dictionary of the Marine or, a Copious Explanation of the Technical Terms and Phrases Employed in the Construction ... Of a Ship. Illustrated with Variety of Original Designs of Shipping, ... To Which Is Annexed, a Translation of the French Sea-Terms and Phrases, Collected from the Works of Mess. Du Hamel, Aubin, Saverien, &C.* Printed for T. Cadell, London, 1769.

27 Oliver, Douglas L. *Ancient Tahitian Society*. 3 vols. Canberra: Australian National University Press, 1975.

28 With the pioneering work of Noel Loos's, *Invasion and Resistance : Aboriginal-European Relations on the North Queensland Frontier, 1861-1897*. Canberra: Australian National University Press, 1982; and Henry Reynolds's, *The Other Side of the Frontier : An Interpretation of the Aboriginal Response to the Invasion and Settlement of Australia*. Townsville: History Dept., James Cook University, 1981.

29 Salmond, Anne. *Two Worlds : First Meetings between Maori and Europeans, 1642-1772*. Auckland, N.Z.: Viking, 1991.

During the course of writing these companion essays, Chris Blackall and I came to sense that one limitation of the OHRM / WARP system was the lack of scope for distributed online content creation and editing. New material could be added, or existing content edited, and the relevant pages on the *South Seas* website easily updated. However, new or amended content had to be entered into one or more interrelated tables of a relation database (Microsoft Access) located on one PC. Hence additional content by authors other than myself or Chris could only be added by new material being sent to him in a form enabling it to be added to the relevant data table. As test materials went online with rich metadata (another valuable feature of the system), these quickly became indexed by Google and other search engines, resulted in us being contacted by historians and literary scholars in the US and Europe asking whether we would be interested in them contributing editions of eighteenth-century voyaging texts that they had created electronically but could not find a publisher willing to produce them in a print-based edition.

At the same time, both of us had been keeping an informed eye on the emergence of the first wiki-style softwares, and the beginnings of online collaborative knowledge creation - notably the pioneering work of John Willinsky in creating what was to become the Public Knowledge Project.³⁰ However, at this advanced stage in the venture, there was no way that the OHRM / WARP system could easily be modified to allow for distributed web-based authoring and editing. Also, the pre-processing we would have to undertake to put these editions into the existing system and publish them on the *South Seas* site was a burden we were in no position to take on.

So while these overtures from colleagues were encouraging, it left us with the uncomfortable feeling that *South Seas* was in a key respect outdated well before its scheduled official launch date. The future of e-humanities appeared to lay with web-based collaborative authoring and editing; and this would entail any future re-engineering of *South Seas* developing workable ways of managing through to publication what could potentially be a substantial number of external contributions in ways that would have to allow for online peer review. And this in turn raised questions about how the IT and human costs of developing projects like *South Seas* could be sustained beyond project-based funding - a question that still needs to be addressed in the context of current strategic initiatives to foster humanities e-research.

The hybrid version of the OHRM and WARP offered workable solutions to the complexities of interrelating texts and annotations and visual material within *South Seas*, while providing the means for users to discover and explore relevant resources in the online collections of the National Library and other cultural institutions employing persistent naming schema in online collections. It meant forsaking the idea of employing XML beyond archiving our editions of the journals and other major texts in TEI format to enable their future use and preservation in future XML-based systems. It also meant working with XML-based mapping no further than ensuring as best we could that the process of creating base-maps from raw GIS data enabled their replication as scalable vector graphics (SVG) as some point in the future. However, as mentioned above, it was already clear that the immaturity of XML in various key respects rendered it too risky to do anything much beyond this. Indeed, discussions with McCarthy and Evans confirmed our suspicions that an XML-based system rivaling the functionality of a hybrid version of OHRM / WARP was still quite some way off in the future.

³⁰ Willinsky, John. *If Only We knew: Increasing the Public Value of Social Science Research*. New York: Routledge, 2000. On the Public Knowledge Project, see <http://pkp.sfu.ca/>

So too did my conversations with the late Mike Fagan, the programming genius responsible for various major online projects created at the Matrix Centre for Humane Arts, Letters and Social Sciences Online, at Michigan State University.³¹ Fagan was greatly impressed by the AUSTHC approach to knowledge management and online publication. It enabled automated creation of linked XHTML pages with rich metadata from content stored as discrete 'records' in a relation database (Microsoft Access). Most importantly any number of relations could be made between these 'records' which would then be expressed as hyperlinks when the records were output as XHTML pages. Similarly, the system rendered it technically easy, if still labor intensive, to present maps of Cook's track as a series of objects in the Flash format with embedded links to the relevant entries of the *Endeavour* journals and Hawkesworth text, and to digital facsimiles (also in the Flash format) of published versions of Cook's own charts (see Fig. 1 below). By 2000 this proprietary format was a *de facto* standard, due to the vast majority of personal computer users using Microsoft's Internet Explorer installed with an updatable Flash viewer. Given the computer skills of prospective users, Flash seemed the sensible way to go given the immaturity of SVG and that the process of creating these maps had been designed to minimize as much as possible the work of regenerating them in an open format at some future point.

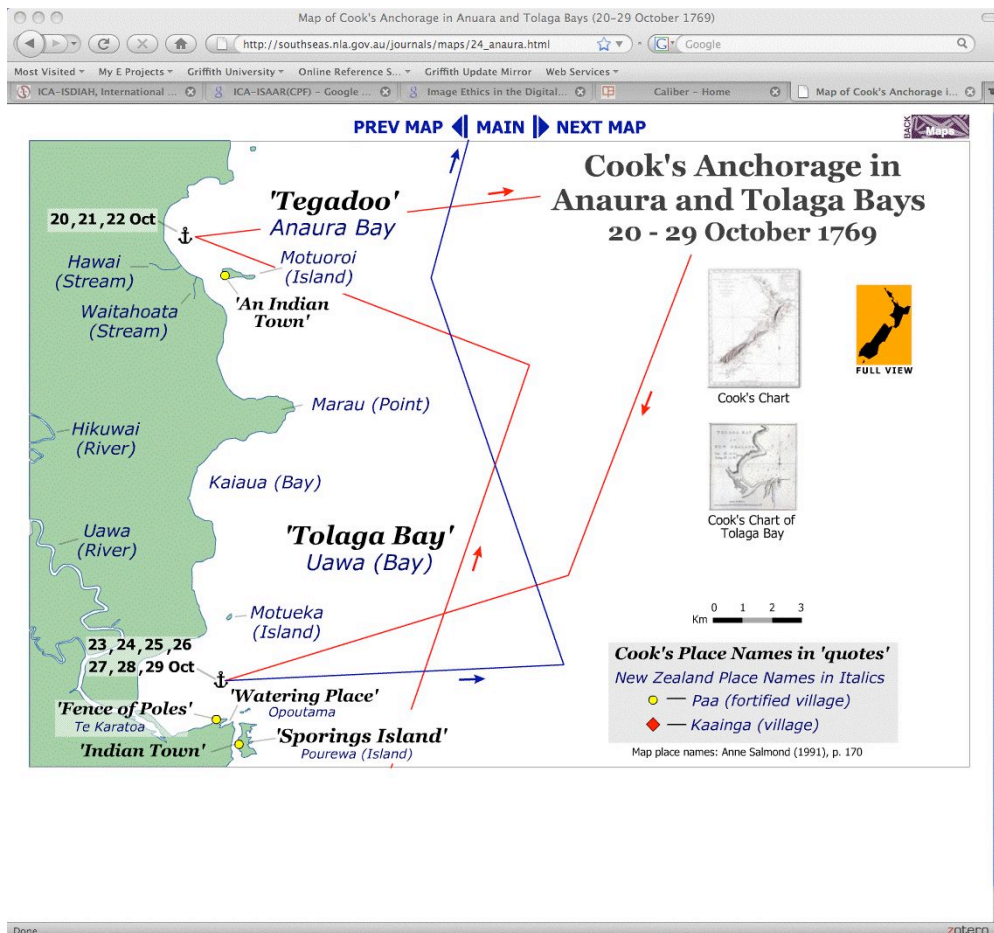


Fig. 1: Screenshot of typical Flash format map on South Seas website

³¹ <http://www2.matrix.msu.edu/>

That future point has now arrived, and another focus of our current work is on the representation of historical knowledge using advances in open source GIS software components. It is relatively easy to develop a web-based authoring and editing environment within the Plone system that gives users the ability to geo-reference knowledge through little more than pointing and clicking on a map. The maps on which they can situate this knowledge can be generated from raw GIS data, using a middleware stack consisting of PostGIS,³² MapServer,³³ together with custom products serving GML encoded data and spatial indexing tools.³⁴ This give us among things the ability to allow for searches based on geographic regions or the intersection of geographical features. Some geographic analyses such as distance calculation can also be undertaken by online users, while the underlying data sets can also be made accessible for more comprehensive analysis using various client-side GIS tools. Moreover, this stack is currently incorporated into the system's authoring and editing environment using the OpenLayers javascript mapping tool³⁵ and the Kupu XHTML editor.³⁶ OpenLayers has allowed us to implement a sophisticated web browser based interface for working with geo-spatial data. We are confident of being able to position historical content on both modern and historical maps and to analyze the correspondence between them. We are also able to use this information to undertake 'rubber-sheeting' - i. e. where digitized historical maps are 'stretched' to match contemporary mapping data.

We are also using OpenLayers as the means of interacting with high resolution digitized scans and photographic images. The same ability to drag, slide and zoom in on an annotatable map is used to enable interaction with images. Regions in an image can be highlighted, annotated and persistently linked to any other content in the system, or to digital surrogates of historical materials offered online by cultural institutions employing a system of persistent identification - as exemplified by the National Library of Australia.³⁷

We have also looked at exploiting the Google Maps API³⁸ for geo-spatial knowledge representation. However, we are rapidly moving away from using Google Maps (such as appears in Fig. 2 below) to rely on our own custom generated maps.

³² PostGIS <http://postgis.refractions.net/>

³³ MapServer, <http://mapserver.gis.umn.edu/>

³⁴ On GML and tools, see <http://www.galdosinc.com/education/gml-tools-2/>

³⁵ OpenLayers, <http://openlayers.org/>

³⁶ Kupu, <http://kupu.oscom.org/>

³⁷ See the National Library of Australia website at: <http://www.nla.gov.au/initiatives/persistence/PIcontents.html>

³⁸ Google Maps Api, <http://code.google.com/apis/maps/>

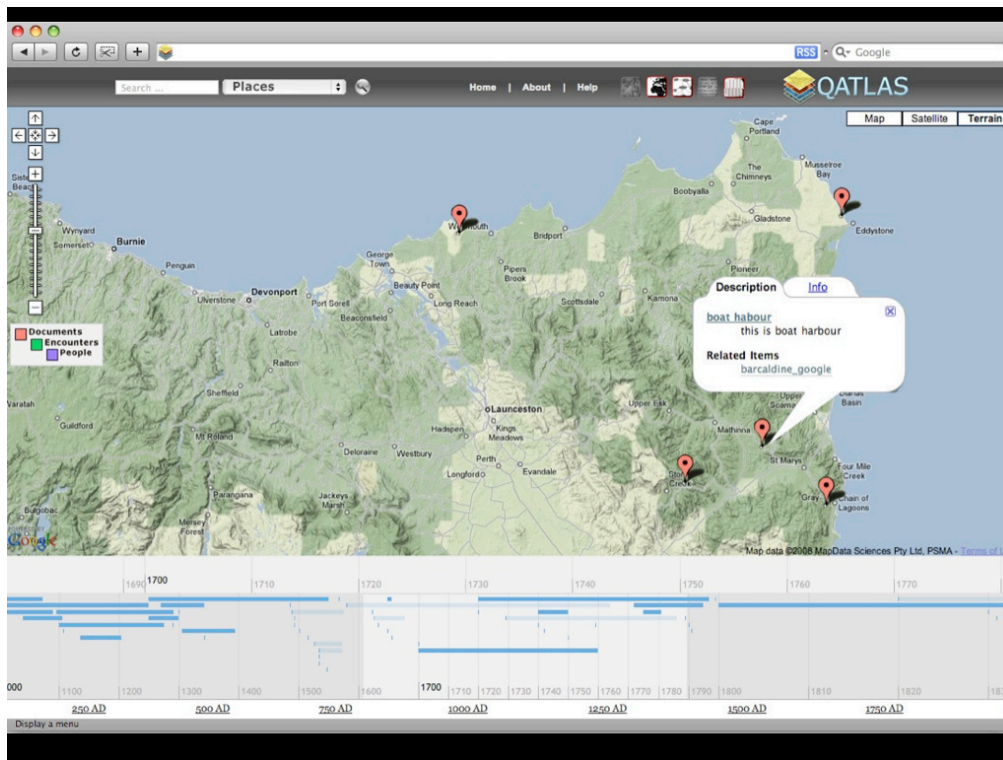


Fig. 2 Screenshot of prototype geo-temporal interface.

This is partly because of concerns about reliance on an external service arising out of its failure during demonstrating a prototype of our system to colleagues with whom we are exploring using the system as the basis for an online historical atlas of north-eastern Tasmania. Though more significant is that using our own maps allows them to represent geo-spatial features of historical significance absent from Google Maps. In this regard, the OpenLayers javascript component offers functionality absent from the public Google Maps api.

Historians are of course interested in investigating the causal connections between things in time, not just space. So the functionality of our system extends, as can be seen above (again Fig. 2), to representing things geo-spatially over time. Currently our prototype of the system interrelates maps of specific localities with a time-line, which allows users to see things in places during the time they existed. We are currently using the SIMILE Timeline³⁹ and Timeplot⁴⁰ javascript components to represent time based information.

Another crucially important feature of the OHRM / WARP system employed in creating *South Seas* has informed our current work. This is the mastery over the management and representation of knowledge provided by the simple yet powerful ontology which lies at the heart of this system. Gavan McCarthy has been recognized as one of Australia's leading theorists and practitioners of archival science. Archivists are fundamentally concerned with the systematic preservation, cataloguing and retrieval of knowledge of past human activity. McCarthy drew upon his expertise to develop the OHRM and other related AUSTHC systems on an ontology derived from the ISAAR (CPF) standard developed by

³⁹ SIMILE Timeline, <http://simile.mit.edu/timeline/>

⁴⁰ SIMILE Timeplot, <http://simile.mit.edu/timeplot/>

the International Council on Archives (ICA).⁴¹ The standard was designed, as ICA points out, to ‘make it possible to collect any important information on the records creators, corporate bodies, persons or families’ but to do so such that archivists can ‘develop dynamic and multidimensional descriptive systems.’⁴² McCarthy’s genius was to use the standard so as to create with the OHRM a cheap and easy to use dynamic and multidimensional descriptive system that could be employed to create a wide variety of online knowledge resources. Indeed, of the online resources that currently comprise e-humanities infrastructure in Australia, a remarkable number employ the OHRM or its variant, the Heritage Documentation Management System (HDMS).

In the case of *South Seas*, it was a relatively simple exercise to extend the archival science-based ontology of the OHRM to provide a remarkable degree of control over the management of diverse kinds of knowledge sources and their digital representation so as not to diminish the ability of researchers and other audiences to grasp the intellectual and cultural complexities of Cook’s first Pacific voyage.

Even so, over the past year we have come to think that online resources such as *South Seas* could benefit from being built on a more descriptively powerful ontology, especially if they seek to move beyond providing a matrix for the presentation and scholarly interpretation of a blend of historically significant documents and visual imagery, to explore the meanings and values of things such as artifacts in museum collections, built heritage and aspects of intangible cultural heritage captured by means of audio-visual and sonic softwares.

WORKING WITH COMPLEX AND CONTESTED KNOWLEDGE

Humanities research involves working with complex knowledge, about which there is often disagreement as to its meanings and values. In designing a knowledge creation and management system for the humanities, it is vital that such a system be able to capture, process and display in suitably nuanced ways the contested nature of this knowledge. It also goes without saying that in the case of creating digital resources exploring aspects of cross-cultural histories or Indigenous histories, it is vital to ensure that such ventures abide by appropriate ethical protocols concerning ownership and control of how the past experiences of the relevant peoples are interpreted in the online environment.⁴³ Ontological modeling must recognize and respect the meanings and values that surrogates of historical and cultural artifacts have for the peoples whose history is being told.⁴⁴

⁴¹ ICA, <http://www.ica.org/>

⁴² ICA, ICA-ISDF, International Standard for Describing Functions, <http://www.ica.org/en/node/38665>

⁴³ See Hart’s Cohen excellent essay, ‘“Moral Copyright”: Indigenous People and Film’, in Gross, Larry *et. al.* (eds.), *Image Ethics in the Digital Age* (University of Minnesota Press, 1993), pp. 313-326. Also online via Google Books. Cohen writes about moral copyright of photographic images and film, but much of what he has to say has become more pertinent with the evolution of digital reproductive and communication technologies.

⁴⁴ As is exemplified in the work of Jane Hunter in connection with the Indigenous Knowledge Management Project. See <http://www.culture.gov.au/conference2/hunter/hunter.pdf>

Any system that sought to exclude what was potentially contestable knowledge would obviously be of limited use. Indeed, the inclusion of contested knowledge is core to the kinds sort of analytical and scholarly activities that many humanities researchers using a digital historical resource would want to undertake - such as, for example, differently organizing and interrelating knowledge in the system to see new connections between phenomena that have been understood in differently enculturated and possibly incommensurate terms.

One could use a traditional relational database consisting of tables of rows and columns to represent the relationships expressed between contested and non-contested knowledge; but this would risk incurring significant administrative overheads. These overheads would likely increase unpredictably as there became a need for increasing precision in the resulting data descriptions, and queries upon those descriptions. After investigating a variety of approaches, we decided that a system that attempted to partition contested and non-contested knowledge was unsustainable and unsuitable for our long term needs.

We needed a system that would allow us to manage all knowledge contained in the system as separate and potentially contested entities. That would allow us to specify arbitrarily complex and possibly contradictory organizations of the knowledge, whilst managing the complexity in such a fashion as to maximize academic productivity and to ensure the ongoing navigation, retrieval and citation of those resources.

We decided upon a a system architecture that consists of primary source ‘resources’ marked up as TEI XML, annotated with knowledge ‘nodes’ and propositional statements relating those nodes to each other. An individual resource might be annotated by many nodes, with each node referring to a specific section within the primary source by way of offset markup⁴⁵. Nodes can be conceptualized as belonging to a flat structure in that they only contain a reference to the primary source, all information indicating how the nodes might be organized is stored externally to the node as propositions about them.

The use of offset markup means that we can usefully annotate ‘read only’ primary sources, such as network accessible external resources, for example material maintained in institutional repositories. Using external resources in this fashion avoids unnecessary duplication of resources and ensures that all scholars are ‘working from the same page’.

The simplest annotation of a resource involves the association of a node with an element of the primary source. Nodes have three properties:

- ID: every node has a unique and persistent identifier.
- Metadata: every node is associated with a limited amount of metadata. This includes, for example, a human-readable name for the node, information about what category of node it is .
- Resources: each node can be associated via offset markup with other documents, maps, images, external resources and perhaps most importantly, newly created scholarly commentaries and essays.

In the system architecture as it currently stands, we have enabled the creation of three categories of nodes:

⁴⁵ Offset markup is described in section 16.9 of TEI P5 see, <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/SA.html#SASO>

- Place Nodes: for elements corresponding to geographic entities (geographic, political and cultural)
- People Nodes: for entities corresponding to the individuals, groups and organisations.
- Event Nodes: for entities that occur at a specific time.

We may add additional node categories as we develop new front end tools to create, display and manipulate them.

To create a node, it is not important whether or not its existence as a ‘real’ entity, is contested, stable or of indeterminate status. Clearly all of the nodes described above are of a type that may have disputed properties. All that is required is a primary source to be annotatable with a reference to the entity. So, for example, any place mentioned in the Cook journal could be treated as a node in the system, irrespective of whether it currently exists with the same name. A node is, therefore a point of abstraction from a primary source that allows us to make propositions about them and their relationship to each other, providing us with unlimited flexibility in classification and analysis. We would not have this same flexibility if we were attempting to work directly with the primary source.

While any kind of relationship can be made between nodes, it is beneficial for technical and analytical reasons to restrict the relationship to a defined ontology. To date our ontological modeling has focused on whether we can develop a ‘user-friendly’ way of describing and managing digital surrogates of historical documents, images and cultural artifacts using the CIDOC Conceptual Reference Model.⁴⁶

The CIDOC-CRM is a complex standard to work with, but has the obvious advantage of having evolved over the past decade, under auspices of the International Council of Museums (ICOM), to become a widely used international framework by which virtually all cultural heritage information in digital forms embodying the standard held by museums, libraries and other cultural institutions can be meaningfully interrelated with this project. Moreover, the standard appears ideally suited to our aims not just because it extends to describing things of both tangible and intangible cultural significance, but has the potential to allow us to employ GIS softwares to geo-spatially and temporally map relations between knowledge sources described using CIDOC-CRM.

We are also evaluating the use of other ontologies to assess their ability to express the sort of relationships between nodes that would be of most use to historians. Among those we are examining are the ABC Harmony ontology⁴⁷ and HISTO⁴⁸.

An illustration of nature of the relationships in the database amount to a collection of statements like: ‘Actor node X was a participant in Event node Y’, ‘Event node Y occurred at Place node Z’. By expressing this kind of knowledge as a relationship we are able to use the CIDOC-CRM to map knowledge from our primary sources to real world entities. However, because we are using relationships to model this knowledge, rather than annotations to an objects metadata, we are not locked in to the CIDOC-CRM. At some point in the future we may decide that one of the other ontologies we are evaluating above

⁴⁶ CIDOC-CRM, <http://cidoc.ics.forth.gr/scope.html>

⁴⁷ ABC Harmony Ontology, <http://jodi.tamu.edu/Articles/v02/i02/Lagoze/lagoze-final.pdf>

⁴⁸ HISTO is the Finnish History Ontology, created as part of the Finnish Culture on the Semantic Web project, <http://www.seco.tkk.fi/ontologies/histo/>

may be found to be more appropriate for modeling history. A simple mapping from one ontology to another can be performed to achieve a new organization of material in the system without substantial modification to the underlying database structure.

We believe the ability to map from one ontology to another delivers benefits by ensuring the system has three crucial features required for modern humanities cyberinfrastructure: flexibility, longevity and interoperability:

- Flexibility is inherent in the design of the system's with arbitrarily complex organisations of nodal relationships capable of being entered into the database.
- Longevity is achieved by nodal content not being dependent on any one set of relationships to encode, store and retrieve nodal content.
- Interoperability is an indirect benefit of both of the above. Knowledge stored in the system can always be made accessible to other systems mapped to their native ontology via a simple mapping. This is can be done for both import of external data into our system and the export of data from it.

To give a concrete example of the benefits flowing from the flexibility of the system's database structure, consider the situation where a researcher using the system wishes to test the hypothesis that there was a qualitative difference between subsets of 'events' that have been classified as 'encounters' between Cook and indigenous Australians reported in the Cook's journal. After using the visualization tools built into the system, the researcher can see that the Cook journal details journeys along the Australian coast that pass the same (or adjacent) coastal areas more than once. Using the system the researcher is able to visualize the results of a query that organizes events by their relationship to time (when the encounters occurred) and to geo-location (where encounters occurred). The researcher can then analyze the nature of those events when seen from the perspective of these different organizations and possibly make new interpretations of events on the basis of this analysis.

The example above raises the question. How do we disambiguate proposed nodes which may correspond to the 'same' real-world entity, but might not be exactly the same thing? For example, is the coastal named by Cook the 'Dove-Cot' the same thing as the geographical co-ordinates recorded in his journal and log-book? Unfortunately the system is not able to provide an easy (and definitive) answer to this question. What the system can do is assist researchers to create 'place nodes' for each of the references in the primary source and to use web browser based tools to make propositions about their relation to each other via examinations of the correspondence between Cook's account of the phenomenon, how it is described in other documents or Indigenous oral traditions, and modern topographic details stored in the system.

While we are looking at implementing automatic proper name detection software such as the General Architecture for Text Engineering (GATE)⁴⁹ and Leximancer⁵⁰ to enable semi-automatic identification of possible place and person nodes, critical evaluation of the type described in the example above will always be required to evaluate and argue for the validity of possible propositions about nodes.

⁴⁹ GATE, <http://www.gate.ac.uk/>

⁵⁰ Leximancer, <http://www.leximancer.com/>

It may be asked why have we chosen our system for historical work over other systems expressing node-like elements and relationship-like statements such as Heurist Scholar.⁵¹ The answer is firstly, our relationships are expressed using the Resource Description Framework Standard (RDF)⁵², a core technology of the semantic web. Secondly, and flowing on from the use of RDF, is that we are able to incorporate a variety of specialist tools for querying and analyzing semantic data, such as RDFLib⁵³. This allows us to offer functions such as a 'live view' of knowledge placed in the system supported by up to date semantic search and navigation. (e.g. using a graph of linked nodes to discover resources associated with individual nodes). A common use of this function might be to ascertain where in historical documents assertions about the nodes have been made.

Avoiding periodic batch processing to produce the 'latest' knowledge greatly assists in the processes of collaborative and distributed authoring. There are also technical benefits in that the RDF processing libraries we use only merges the relationships necessary to facilitate specific tasks. For example to display all the relationships linking to 'Actor node X'. This is a much more computationally and memory efficient approach than merging all relationships in the database upon each request especially in the case of large data sets or use of the system as an institutional repository.

Finally, use of RDF enables us to use the SPARQL Query Language for RDF (SPARQL).⁵⁴ This allows for the construction of complex SQL like queries over the data. Development of front end interface tools to construct and refine SPAQL queries through the web will allow researcher to analyse large volumes of semantic data and to store custom queries as content objects that may be annotated and referenced like any other content in the system.

As previously mentioned, the system and its web output employ the open-source Plone content management system, which is itself built on the open-source Zope application server. As a result we are also able to use Zope's built in database the Zope Object Database (ZODB) for all of our resource, object, and RDF graph storage.⁵⁵

As opposed to a relational database such as MySQL⁵⁶, a ZODB is an object-oriented database. Such a database can have largely similar functionality to a relational database. However, rather than store its data in the form of tables, it stores it in the form of objects. These objects can be assigned properties. We can thus model the ZODB representation of a knowledge node as follows

NODE OBJECT

Properties

ID: encounter_1769_May_2

Metadata:

⁵¹ Heurist Scholar - developed by a team at the Archaeological Computing Laboratory, University of Sydney, under the direction of Ian Johnson, <http://heuristscholar.org>

⁵² RDF, <http://www.w3.org/RDF/>

⁵³ RDFLib, <http://rdflib.net/>

⁵⁴ SPARQL, <http://www.w3.org/TR/rdf-sparql-query/>

⁵⁵ See <http://www.zope.org/Products/StandaloneZODB>

⁵⁶ MySQL, <http://www.mysql.com/>

Name: Theft of Astronomical Qaudrant

Type: Event

Resources:

Annotation1, Annotation2..., AnnotationN

The properties include a persistent ID, a human-readable title for the node, an indication that this is a event node, and a list of references to specific sections of a TEI XML resources in the form of offset markup. Depending on the nature of the resources, they may be included directly in the ZODB. Or, if they are of a type which impractical for storage in the ZODB for various technical reasons (e.g. high resolution scanned image files) an external or filesystem reference.

Unlike the nodal portion of the system which has been developed as a custom Plone product, the storage format for the relationship propositions has been implemented . using the formal mechanisms of Resource Description Format (RDF).

The basic principal of RDF is that a semantic representation of the content of a set of data can be expressed in the form of a special class of statements referred to as triples. A triple consists of a subject, a predicate, and an object. The triple states that the relationship referred to by the predicate holds between the subject and the object. Triples can be expressed in a number of ways. The most intuitive form is as a directed graph. Indeed this is the form we store them in.

An example from our current work is expressed below.

E39.Actor:http://atlas.griffith.edu.au/people/James_Cook --->p11b.participated_in-----
>E5.Event:http://atlas.griffith.edu.au/event/meeting_with_Joseph_Banks

The directed graph above expresses that there is an entity with URI http://atlas.griffith.edu.au/people/James_Cook. The entity referred to is the person James Cook. This URI serves as the unique ID which allows nodal content and relationships to be merged. The predicate in the triple is represented as an arrow pointing from the subject to the object. The prefix used in the triple refers to the XML namespace defined in the CIDOC-CRM ontology. This triple, therefore, expresses the claim that in English would be rendered as 'The person James Cook was a participant in the event of meeting with Joseph Banks'. Much of the power of RDF comes from the fact that triples can be joined to form larger networks of triples. This allows knowledge to be inferred from nodes, between whom, there is no direct relationship.

A network of triples could be used to represent that the event 'Meeting with Joseph Banks' had other participants, including Banks himself and that Banks was born on 14 February 1743 and died on 19 June 1820. And of course further assertions can be made by creating relations between networks of nodes and detailed annotations of digital editions of documents such as Cook and Banks's *Endeavour* journals.

We are currently working to develop interface tools that allow annotation of the triples. We anticipate that researchers will appreciate being able to indicate who was responsible for the creation of the triple and more substantively, being able to associate a triple with an annotation that provide justification for putting a triple in the database. This justification may be as simple as a georeferenced point or it may take the form of a scholarly essay.

The ability to annotate triples brings us back to the beginning of our discussion of the technical features of our system. By implementing a system of TEI xml encoded primary sources, annotated via offset markup via nodes and networks of annotated RDF propositions we are able to allow researchers to easily explore competing hypotheses concerning a particular set of historical data. In the context of our historical interests, it would allow researcher to include contradictory statements regarding the location of events within the same database, annotating each proposed relationship for what sort of evidence was used to justify it and where the evidence can be found in primary sources. Users could then choose to extract proposed relationships based on different kinds of evidence or of different provenance in order to compare them, and in turn perhaps refine extant knowledge claims.

SOUTH SEAS AND THE MIGRATION OF HISTORICAL PRACTICE ONLINE.

During the course of creating *South Seas*, I was anxious that our investment in developing the underlying technology did not jeopardize its primary aim, which was to show that digital technologies and related information standards could be employed to bring new dimensions to historical practice. And with good reason. *South Seas* was recently described by an anonymous expert in humanities e-research as ‘...one of the most valuable resources produced within the Digital Humanities within Australia in both its significant content and technical sophistication.’ This is an over-statement; but one heartening to encounter given that *South Seas* was created in an environment in which many colleagues still saw the web as at best ephemeral to the real business of historical scholarship - the writing of scholarly books and articles. Many stories could be told. One of my worst memories is being told that a selection committee considering me for a more senior position had supposedly been told I was unsuitable because ‘he’s not interested in ideas, but with playing around with computers.’ But comments made me all the more determined to bring the venture to a good outcome. They reflected at best ignorance of what by 2000 a small but increasing number of historians - especially in the United States - were attempting to do with digital technologies, as well as patrician disinterest in the spectacular growth through the 1990s of new web-based audiences for historical scholarship.

I have written elsewhere about this neglect of how the integration of networked communication within everyday life has led to increasing numbers of Australians creating and consuming online information about history and heritage.⁵⁷ Here, it seems worth rehearsing the point that no historian now using networked digital media has to my knowledge ever argued that the historical profession should forsake conventional print-based modes of communication for the internet. Rather, they have argued that the visual and sonic possibilities of digital communication technologies make them media for undertaking and presenting research in ways that can elucidate aspects of past phenomena that are extremely difficult, and in some instances impossible, to explain satisfactorily through the medium of print alone. In this sense, digital technologies promise to be an important, integral future dimension of historical research. Moreover, as the eminent US historian Orville Vernon Burton has argued, if historians continue to judge the book as the sole ‘gold standard’ of professional worth, they risk retarding the development of technical and conceptual solutions for making the web a more robust medium for historical scholarship. Burton, a leading figure in the historical application of computing since the

⁵⁷ Turnbull, Paul. ‘Australian Historians and the World-Wide-Web’, *Public History Review*, 8 (2000), pp. 17-30

late 1970s and head of the US National Super-computing Authority's Humanities program, has rightly pointed out that

unless work done in digital forms is...rewarded in the same ways as work done in more traditional forms of history...those interested in digital history must either abandon or limit those interests to create traditional scholarship worthy of tenure or promotion or remove themselves from history departments and move to more supportive academic units such as Digital History Centers or Departments of Information Sciences or Library and Information Sciences, where an increasing number of historians and humanities scholars reside within universities.⁵⁸

There are indeed advantages in historians being located in academic units supporting research collaborations between historians, digital librarians and information scientists; the increasing sophistication of networked communication technologies makes it crucial that practitioners of digital history engage in collaborative research with librarians and information scientists. However, as Burton points out, if digital history is not done within history departments it will make it harder for historians using digital technologies to have their aspirations understood and valued by disciplinary peers working in other fields of historical inquiry. They also risk losing the benefits of enriching their work through interaction with peers who are engaging in researching the same or related phenomena, but continue to favor print-based modes of scholarly communication.

In hindsight, the disinterest in fostering the integration of digital technologies into historical practice that Chris Blackall and I encountered in the early 2000s proved, on balance, an invaluable stimulus. It impressed on us that if change was to occur, it would happen through resources like *South Seas* being things that even more conservative minded historians of the eighteenth-century Pacific might want to use in the course of their research. Prototype work suggested that the planned correlation of entries in the key *Endeavour* journals on a chronological basis was something that would tempt any scholarly user. Indeed, it has since proved one of the most used features of *South Seas*.

With relatively little publicity, *South Seas* has in fact become a resource extensively used by historians, anthropologists and literary scholars. However, its most visible use has been in university courses on the history of voyaging and cross-cultural exchange in the eighteenth-century Pacific. Leading Canadian and US colleagues in history and anthropology have reported informally on using *South Seas* with graduate students, having discovered the resource through George Mason's online guide to world history sources. The guide describes *South Seas* as follows:

In teaching world history courses, [South Seas]...contributes to an understanding of a key moment in Pacific exploration and cross-cultural encounters. It provides useful comparison to other frontier situations in a global context, particularly to themes of "first contact" and the representation of non-Europeans in European cultural and scientific thought. The site's organization makes it particularly suited to exercises that compare differences between...texts in order for students to engage with questions of evidence and interpretation.⁵⁹

⁵⁸ Burton, Orville Vernon. 'American Digital History', CHM Essays, reprinted from, vol. 23 no. 2, Summer 2005, pp. 206-220, <http://chnm.gmu.edu/resources/essays/d/30>

⁵⁹ <http://chnm.gmu.edu/worldhistorysources/r/73/whm.html>

It has been difficult to gauge how far South Seas has been used as a teaching resource in the United Kingdom and other EU countries, but some measure of the value placed on the resource can be taken from positive assessments of the resource in online catalogues of digital resources in the arts and humanities.⁶⁰

Research use has been harder to judge; and in retrospect some further formal investigation should have been attempted. However, what we have learnt on this score has come to inform the work we are currently undertaking on creating a new knowledge creation, management and publication system for e-researchers in the humanities. Perhaps the most important feature of research use of *South Seas* to date is that it has gone unacknowledged. Typically a researcher will compare, for example, what Cook and Banks said about a particular incident, then write about its significance citing only Beaglehole's published editions of the two journals. As far as can be ascertained, John Gascoigne is the only prominent historian of the eighteenth-century Pacific to acknowledge their indebtedness to *South Seas* in providing him with the means to undertake extensive comparative analysis of the most important first-hand accounts of the *Endeavour* expedition. In the preface to his 2007 study of Cook, he writes:

Any student of Cook is also much indebted to invaluable reference works on the subject... [including] ...the great amount of useful material drawn from Cook's day and our own made readily available on the National Library of Australia-sponsored 'South Seas' website.... Such works have ever been at my (physical or cybernetic) side while writing this book.⁶¹

This preference of historians to consult *South Seas* but cite print versions of the *Endeavour* journals is not something I've decried. Quite the opposite; I think it serves to underscore that while significant progress was achieved in making the web a medium of historical practice, *South Seas* has been at best a useful adjunct to print-based scholarship.

This is because it does not inspire the same investment of trust that historians routinely give put in scholarly artifacts such as the book version of Beaglehole's edition of the Cook and Banks journals. Nor, in truth, does *South Seas* warrant the same degree of trust. Its electronic transcripts of the first two volumes of Hawkesworth's accounts of the Carteret, Wallis and Cook voyages were meticulously prepared by Dr Christine Winter. The version of Banks's journal it offers uses the electronic text prepared by the State Library of New South Wales. But all of the other texts were created solely by me while actively engaged in all other aspects of the project. These texts are generally accurate, but over the years various minor errors have been discovered. There was simply not the resources available to the project to ensure my complete fidelity to the original documents.

Bodies of knowledge in the natural and human sciences have evolved through their circulation, interpretation and re-appraisal in forms of reportage that are communally agreed to warrant trust. Historical practice, for example, relies on the ability to trust whether evidence licenses how past phenomena are explained, generally through an infrastructure of interconnected knowledge residing in books, peer-reviewed articles and other print genres of scholarly discourse. While practitioners of history with shared or related interests may challenge the interpretation of evidentiary sources on the basis of

60 For example, the UK's *intute* research discovery service, supported by the Economic and Social Research Council and the Arts and Humanities, <http://www.intute.ac.uk/cgi-bin/search.pl?term1=South+Seas&submit.x=9&submit.y=5&submit=Go&limit=0&subject=All>

61 Gascoigne, *John. Captain Cook : Voyager Between Worlds*. Hambledon, London, 2007, p. vi.

first-hand examination of those sources, more commonly interpretations of evidence are accepted as trustworthy for the purpose of developing new or variant arguments. What makes them trustworthy is their embodying agreed procedures and conventions for judging the value of historical scholarship - things such as conventions for producing editions of historical texts, for accurately citing evidence and argument in books and articles, and for providing bibliographical information confirming that publications have undergone some forms peer scrutiny.

These points may seem obvious and trite; but it is important to see that these key infrastructural elements of printed-based scholarly communication need to be replicated in the online environment. This is happening. The development of things such as internationally agreed standards for metadata and digital object identification have allowed researchers, especially in the sciences, to engage in knowledge production with little or no reference to print-based papers and journal articles. But it is important to see that what has so far been integrated within research practice are digital surrogates of print-based genres of scholarly communication possessing much the same form they have had since the 1860s. We have yet to develop the means of working with more complex kinds of 'born digital' knowledge resources.

What is more, the infrastructure of scholarly communication shapes and is in turn shaped by the ambitions, assumptions and practices that have evolved within specific cultural geographies. In the case of scientific communities, electronic communication has become overtly integral to scientific practice, due in large measure to the ability to access, aggregate and critically appraise relevant knowledge speedily and comprehensively long being an essential element of that practice. By way of contrast, in humanities disciplines such as history, analysis of footnotes to articles in recent print-based history journals reveals that while these authors are almost certainly making extensive use of online resources such as JSTOR and Project Muse, they overwhelmingly cite in writing the original artifact, rarely if ever its online surrogate. This raises the question of whether the provision of more extensive e-research infrastructure alone will benefit historians using digital technologies in innovative ways. There is a deeply enculturated privileging of print-based communication shaping what it means to 'aspire to the character of an historian' that is yet to be usefully unpacked.

FUTURE VOYAGING

In this paper, we have described key aspects of a knowledge creation, management and publication system we are developing which we believe will be well-suited to migrating historical practice into the online environment. Through its combining a sophisticated yet easy to use TEI XML authoring and editing environment with advanced visualization techniques and tools for multi-description and management of knowledge, the system, we believe, will have value not only for historians, but for many researchers across the spectrum of the humanities. Importantly, this system is being built on the basis of experience gained in the creating of *South Seas* - an earlier attempt to integrate digital technologies within historical practice. It is informed by a deep concern to ensure that in developing this new system it has the ability to replicate essential elements of historical practice in the digital environment.

Many challenges remain. It is work that like so much else in Australian universities is meagerly funded. It is also work, the significance of which is still not understood by many

disciplinary colleagues; and it may be that shifting disciplinary culture will prove as much if not more of a challenge than making good use of digital technologies. The reality is that research students currently wanting to incorporate networked digital outcomes into their doctoral studies are still finding it difficult to locate themselves in Australian history departments.

Yet, on the other hand, history and humanities subjects are being increasingly taught in our schools through what is often remarkably innovative use of digital technologies - as is strikingly evidenced by the number of sessions on the use of these technologies in the program of the 2008 History Teachers of Australia Association's national conference.⁶² Moreover, government has recognized the research and broader educational possibilities of digital technologies for the humanities. This may still not convince colleagues who are comfortable and relaxed in the world of print-based scholarly discourse to 'go digital.' But that should not be the goal. Rather, we need to focus on translating what is good and valuable in our established disciplinary practice into the online environment. This seems the best way to ensure a future in which humanities researchers who want to use digital technologies have not only the technical resources, but also the peer recognition, to enable them freely to do so.

⁶² Program available from the Queensland History Teachers Association's website, <http://www.qhta.com.au/conferences.htm#153962>