

Collaboration Success in the Dataverse Libraries as Digital Humanities Research Partners

Sue Owen

Deakin University, sowen@deakin.edu.au

Deb Verhoeven

Deakin University, deb.verhoeven@deakin.edu.au

Anne Horn

Deakin University, unilbrn@deakin.edu.au

Sabina Robertson

Deakin University, sabina.robertson@deakin.edu.au

Sue Owen, Deb Verhoeven, Anne Horn, and Sabina Robertson, "Collaboration Success in the Dataverse Libraries as Digital Humanities Research Partners." *Proceedings of the IATUL Conferences*. Paper 1.

<http://docs.lib.purdue.edu/iatul/2014/openaccess/1>

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries. Please contact epubs@purdue.edu for additional information.

COLLABORATION SUCCESS IN THE DATAVERSE: LIBRARIES AS DIGITAL HUMANITIES RESEARCH PARTNERS

Sue Owen

Deakin University, Australia
sowen@deakin.edu.au

Deb Verhoeven

Deakin University, Australia
deb.verhoeven@deakin.edu.au

Anne Horn

Deakin University, Australia
unilbrn@deakin.edu.au

Sabina Robertson

Deakin University, Australia
sabina.robertson@deakin.edu.au

Abstract

At Deakin, the Humanities Networked Infrastructure project (HuNI), has paved new ground for facilitating the effective use and re-use of humanities research data. HuNI is one of the first large-scale eResearch infrastructure projects for the humanities in Australia and the first national, cross-disciplinary Virtual Laboratory (VL) worldwide.

HuNI provides new information infrastructure services for both humanities researchers and members of the public. Its development has been funded by the National eResearch Collaboration Tools and Resources project (NeCTAR) and undertaken by a consortium of thirteen institutions led by Deakin University. A Deakin University Library team with skills in data description, curation, retrieval and preservation is exploring with digital humanities researchers and developers effective means to support and maintain the HuNI project.

HuNI ingests and aggregates data from a total of 31 different Australian cultural datasets which cover a wide range of disciplines in the humanities and creative arts. The HuNI VL also provides a number of online research capabilities for humanities researchers to discover and work with the large-scale aggregation of data. The HuNI VL enables researchers to create, save and publish selections of data; to analyse and manipulate the data; share findings and to export the data for reuse in external environments.

In a major innovation, HuNI also enables researchers to assert relationships between entities in the form of 'socially linked' data. This capability contributes to the building of a 'vernacular' network of associations between HuNI records that embody diverse perspectives on knowledge and ramify avenues for research discovery beyond keyword and phrase searches.

This paper reports on key milestones in this project, the future role of Libraries as digital humanities research partners and the challenges and sustainability issues that face national digital humanities research projects that are developed in strategic library settings.

1. Digital Humanities: Libraries and the Academy

The Digital Humanities (DH) had their inception nearly 70 years ago, when in 1946, Padre Roberto Busa planned to compile *Index Thomisticus*, a concordance to the works of Thomas Aquinas (Winter, 1999). He successfully persuaded IBM to sponsor his work in 1949. A thirty year project ensued, producing a printed index of 56 volumes. This early scholarship was referred to as 'humanities computing' and offered a significant new approach to humanities research over the manually driven research of the past (Adams and Gunn, 2012).

Libraries and the academy have a shared philosophy.

We value

the rights of citizens to access information and the human record, and infrastructure and systems that are openly accessible, collaboratively built and ensure long-term preservation.

We understand

the serendipitous linkages that occur between knowledge assets and how an understanding of e-infrastructure and digital tools enable researchers to discover and exploit previously unimaginable avenues of research.

Digital Humanities was a topic of interest at the 2008 IATUL conference. At this time, research libraries in the United States were actively debating their role in this new arena. Digital tools for the humanities were becoming more widely available (University of Virginia, 2006) and e-science, cyber infrastructure, and e-research were increasingly terms used in relation to the Humanities (Crane, Babeu & Bamman, 2007).

In late 2007 the US Council on Library and Information Resources (CLIR), sponsored a one-day workshop, *Promoting Digital Scholarship: Building the Environment*. The workshop focused on mass digitisation projects (Crane & Friedlander, 2008). The following year, CLIR in conjunction with the National Endowment for the Humanities (NEH) ran a symposium *Promoting Digital Scholarship: Formulating Research Challenges in the Humanities, Social Sciences, and Computation* (Friedlander, 2009).

In preparation for the 2008 symposium, CLIR commissioned a survey of Digital Humanities Centres to establish the extent of these centres and also their financing, structure, products and services and sustainability (Zorich, 2008). Two general categories of Digital Humanities Centres were identified:

- Centre-focused, organised around a physical location with a wide range of resources and many diverse projects and programs undertaken by diverse audiences
- Resource-focused, organised around a primary resource located in a virtual space, serving a specific group of members.

Infrastructure for Digital Humanities at that time was viewed largely in terms of physical entities - buildings and databases.

Also in 2008, the Australian Academy of the Humanities, was exerting pressure on the Australian government for the inclusion of the humanities and social sciences in its consideration of funding for collaborative research infrastructure (Byron, 2008). The Academy was scoping infrastructure "to enable the transformation of humanities and interdisciplinary research practices and outcomes" (Turner, 2008, p 4.) University libraries were key allies in this endeavour.

Over the last seven years, Australia has developed an e-infrastructure ecosystem that has been nourished by government funding and nurtured by government policy. The Digital Humanities has found its place within the 'big science' model of research processes and a number of major infrastructure projects have been funded. New communications technologies and research data infrastructure are now appreciated by humanities researchers, enriching the connections within the academy and powering the linkages of content and data.

One of these projects is the Humanities Networked Infrastructure (HuNI) service. This paper describes an initiative led by Deakin University, in collaboration with national and international partners to advance the Digital Humanities. The project has been funded through Australia's Education Investment Fund (EIF) which invites competitive bidding for infrastructure development for research (Australia. Department of Finance, 2008). Project partner countries include Finland, UK, US, the Netherlands and international partner organisations include Europeana, The European Library and Data Analysis and Networked Services (DANS).

2. The contribution of research libraries

Making information accessible, upholding the principles of freedom of information and harnessing the affordances of technology, underpin the library and information science profession. Web scale discovery services recognise that academically useful sources of information emanate from wide ranging locations, not simply library-acquired collections (Cossham, 2013). Academic libraries have liaised with vendors to address the challenges of federated searching, meta-searching and most recently webscale discovery to ensure the university community benefits from all quality sources whatever their provenance. Vendor products which provide access to distinct siloed databases are not satisfying the complex and unique needs of university researchers. They neither enable cross-disciplinary discovery nor provide digital tools for addressing complex research problems.

Library 'research data' discussions focus on description, storage, access, preservation and re-use of data sets, including big data outputs. A recent US initiative is the SHared Access Resource Ecosystem (SHARE) a project involving a coalition of libraries and universities to make research publications and associated research data sets from US digital repositories more discoverable and more accessible and for the research community to productively create new knowledge from these assets (Association of Research Libraries, 2014). SHARE, according to Walter and Ruttenberg (2014) will

...comprise a notification system to alert interested stakeholders of new research results releases; a distributed registry able to accommodate publications and research data; a discovery layer across multiple repositories; and a content aggregation layer facilitating data and text mining in addition to curation and discovery.

While many libraries are embracing new roles in relation to research data, an equally important question is the library's contribution to useful applications of linked data, creating connections across databases and leveraging inherent connectedness. Linked data uses 'triples' to assert connections. Triples comprise three elements, a subject, predicate and object which together define the relationship between concepts. OCLC's WorldShare platform has been an early adopter of linked data, along with the Swedish, French and German National Libraries, the British Library, the Library of Congress and Europeana. (OCLC, 2012). The BBC's Nature website actively uses linked data, assembling media, image, sound and textual resources from many authoritative sources into one place (www.bbc.co.uk/nature/).

A recent \$1million grant from the Andrew W Mellon Foundation is funding Cornell University Library, Harvard Library Innovation Lab and Stanford University Libraries to investigate the use of linked data and the semantic web as common formats to bring together disparate library systems and improve discovery, access and knowledge about how content is being used (Cornell University Library, 2014).

To be of most value to the community the linked data also needs to be open. In 2011, the first International Linked Open Data for Libraries Archives and Museums (LODLAM) Summit was convened. Reporting on the second LODLAM summit in 2013, Weinberger (2013) recognised that Linked Open Data is 'the future' for libraries archives and museums data – the challenge is to breach the current divide between traditional 'schema' thinking and Linked Open Data

opportunities. These approaches are seen as vital to ensuring serendipitous discovery pathways for researchers.

The importance of serendipity to the research process is widely recognised, and is particularly profound for the humanities and creative arts researchers (Hoeflich, 2007; McClellan, 2005). The role of contemporary research libraries, and more particularly the use of digital technologies, for enabling serendipitous discovery in the humanities is still a relatively nascent field of enquiry. Many researchers for example, believe that the digitisation of research methods has limited rather than enhanced serendipity (Quan-Haase & Martin, 2011).

One possibility afforded by the digitisation of library catalogues and resources is the ability for researchers in the Humanities to use and create linked data through open systems and through socially constructed linkages which are driven by the perspectives and understandings of individuals. The opportunity for researchers to move from simply accessing and sharing data to sophisticated methods for sharing their knowledge as linked data is a step closer to ensuring serendipitous discovery in a digital information environment. To do this without predetermining the thinking of researchers into restrictive ontologies or categorisations further opens new realms of understanding (Burrows, 2013).

Information specialists and libraries are already contributing to significant advances in the range and reach of Digital Humanities tools, infrastructure, approaches and collaboratories. In addition to those mentioned earlier, well established programs and innovative services include:

- University of Virginia Library Scholars' Lab www.scholarslab.org/
- Harvard Library Innovation Lab <http://librarylab.law.harvard.edu>
- University of South Carolina Centre for Digital Humanities, Division of Libraries <http://cdh.sc.edu/>
- The Humanities and Technology Camp (THATCamp): participants include many DH practitioners, including librarians. THATCamp 2014 was held 24-25 April, University of Florida, Gainesville, USA <http://gainesville2014.thatcamp.org>
- National Library of Australia: Trove provides a single access point to over 393 million Australian and online resources located in the deep web <http://trove.nla.gov.au/>
- Victoria. Parliamentary Library (Neish, 2014)
- Humanities, Arts, Science and Technology Alliance and Collaboratory (HASTAC) Listing of Digital Humanities Centers and Institutes (mostly US-based) <http://hastac.org>
- The HuNI research infrastructure includes the world's first national cross-disciplinary virtual laboratory, designed to meet the specificities of the humanities research methodologies and to increase capacity for the next developments in the Digital Humanities.

These programs offer valuable insights into the progress and contributions already made by research libraries to digital humanities research.

3. Opportunities for Digital Humanities with NeCTAR

Significant infrastructure development supporting research in all disciplines, including the Digital Humanities, is now underway in Australia. As part of the Australian government's Education Investment Fund, the National e-Research Collaboration Tools and Resources (NeCTAR) program in 2011 provided funding opportunities to develop and extend e-Research infrastructure (<https://www.nectar.org>).

One of the major NeCTAR funding streams focused on the development of discipline-based 'virtual laboratories'. The aims were to integrate existing tools/data/resources; support data-centred workflows in research; and build virtual research communities which could better address existing research problems.

The type of 'virtual laboratory' envisaged by NeCTAR was based firmly on a 'big science' model of research processes, requiring a cross-disciplinary approach (Burrows, 2013). It was found that existing approaches to building digital resources for the humanities would not be sufficient to meet NeCTAR's parameters – neither an aggregated collection from a variety of sources, (similar to Europeana www.europeana.eu/) nor an aggregated database of metadata records from Australian collecting institutions (similar to the National Library of Australia's Trove service <http://trove.nla.gov.au/>).

Humanities researchers at Deakin University, in collaboration with research colleagues nationally and internationally, successfully submitted an innovative proposal for infrastructure funds to support further developments of NeCTAR. Three-year funding for the proposal was granted and in May 2012 Humanities Networked Infrastructure (HuNI) was launched.

4. Creating HuNI from NeCTAR

HuNI (<http://huni.net.au>) is a major new enabling infrastructure service for humanities researchers. Deakin University is the lead partner working with a consortium of thirteen institutions. HuNI ingests and aggregates data from a total of 31 different Australian datasets which cover a wide range of disciplines in the humanities and creative arts, including literature, biography, performing and visual arts, media studies and linguistics. HuNI offers cultural industry researchers both simple and advanced discovery tools that comb across these curated academic collections. It does this without itself being in the 'content business' thereby upholding content neutrality in search results.

It is however the intention of HuNI to influence research practices in other, more transparent ways. For example, HuNI has been developed in order to expand the kinds of questions that can be asked of Australian culture and to produce new discoveries. And to achieve this HuNI has forged an innovative approach to data aggregation and data-centred workflows for humanities research.

Underlying the HuNI initiative is the recognition that cultural data is not insular and in order to explore its dimensions fully, researchers need to collaborate across disciplines, institutions and social locations (Verhoeven 2014). To this end HuNI provides relatively simple access to multivariate data derived from different research domains and initially developed to solve different research questions through a single easy to use website. The datasets and services at the heart of the HuNI aggregate cover a variety of disciplines across the humanities and creative arts, include (but are not limited to) the AustLit database, the Australian Dictionary of Biography (ADB), Design and Art Australia Online (DAAO), the Encyclopaedia for the History of Australian Science (EHAS), Australian Indigenous Languages Database (AUSTLANG) and media history databases like Cinema and Audiences Research Project (CAARP) and bonza Cinema Studies Database. HuNI provides an environment for harvesting, ingesting, matching, aligning, and linking the entity data from these disparate sources.

HuNI enables researchers for the first time the opportunity to search and browse across datasets drawn from a wide range of humanities and creative arts disciplines and sources. HuNI also allows researchers to work with the results of the discovery process; to create, save and publish selections of data from HuNI by constructing personal or shared virtual collections; to manipulate these and export them into other software environments for further analysis. The collections of annotations and result sets created by researchers using the aggregated HuNI data will be stored within the HuNI virtual laboratory for sharing and reuse. It will also be possible for casual collectors to enhance their data for ingestion into the HuNI aggregate.

One of the most innovative features of HuNI is the way it enables researchers to link and define the relationships between entities within the HuNI aggregate. If, for example, researchers search HuNI and identify two entities in their result set which are related, they can add a link between the two records and nominate the nature of the relationship. The linking statement might be drawn from a suggested vocabulary of relationships, or the researcher might simply use free text. This feature

has also been developed to allow a researcher to state that two entities are not related, in recognition that this kind of assertion is also a key characteristic of humanities research. These crowdsourced or “social” linking assertions are visible in HuNI. To help display the proposed relationships between entities, HuNI produces a network graph, displaying up to six degrees of separation. In this way, the “social linking” of data forms the basis for researchers to both create and browse network graphs within the HuNI Virtual Laboratory, enhancing opportunities for serendipitous discovery.

The opportunity to socially link data encourages HuNI users to share their knowledge and research findings in the form of specific assertions, and to evaluate, discuss or debate these statements with each other. Crucially, the source of the social linking statements is captured, enabling subsequent researchers to see who created each assertion. HuNI users can annotate previously produced links with their own comments and assessments.

By not relying on a pre-determined mapping to a detailed ontology, HuNI depends on the generous participation of researchers and community users to establish most of the connections within the data aggregate. This has the additional benefit of enabling HuNI to capture the different disciplinary perspectives of users, rather than trying to fit them into a single normative framework.

This socially linked data technology was developed in consideration of the need to balance the disciplinary imperatives for specific vocabularies and data structures with designing a service that is explicitly intended to transcend disciplinary boundaries and link related data effectively and meaningfully. Socially linked data as proposed by HuNI suggests that ‘linked open data’ might equally be explored through ‘open linking’ as much as the more conventional emphasis on ‘open data’.

HuNI demonstrates that there is value in developing shared common systems that can enhance use of distributed collections and open up new audiences for them. In developing HuNI, strong relationships have been developed for sharing humanities and creative arts data and resources across a multitude of organisations. HuNI provides an important common infrastructure for both institutions and individuals with humanities and creative arts collections to share their information with each other and a wider group of users. It reveals the value of open data, the development of APIs (automated data sharing systems) and the use of new analysis tools for researchers (both within and outside the academic community).

5. Challenges and opportunities for Library involvement

Ongoing and perpetual access to, development and use of HuNI represents a strategic issue for the HuNI project team and for humanities and creative arts researchers in Australia. This is a challenge for all e-research projects and the challenge of sustainability is a shared experience of many researchers in the digital humanities. The need for reliable and enduring services was recognised early in the National Collaborative Research Infrastructure Strategy Roadmap (NCRIS) funding discussions. For humanities researchers, libraries are valued as partners due to their reputation for providing robust services that are ‘always on’, and for their stewardship in providing trusted safe keeping of the human record. It is not surprising, therefore, that Deakin University Library is exploring with HuNI Project leaders how the library and the academy can assure access to and development of HuNI into the future.

Reflecting trends in research libraries across the globe (ACRL, 2012), Deakin University Library’s workforce capability is evolving, with the core skills underpinning our profession being re-imagined and re-aligned to keep pace with the affordances of new technologies and the information requirements of researchers. Research repository and webscale discovery developments initiated by the Deakin University Library in the last five years have sharpened our skills in data description, curation, retrieval and preservation. Deakin University Library’s Research Services have also

aligned us closely with data requirements for research funding applications and the data needs of researchers at all stages of the research cycle.

Challenges for HuNI in the future include:

- Ongoing maintenance of the HuNI aggregate and technical services within existing university infrastructure systems
- Assurance of HuNI database and virtual laboratory access now and in the longer term.
- Effective and efficient use of the HuNI VL through user training and awareness programs
- Ingestion and aggregation of new records as they are added to the existing 31 databases (and potentially more in future).
- Developing analytics and evaluation measures to inform ongoing research and development of the HuNI VL.
- Successful grant applications to further develop HuNI VL functionality, data visualisation and socially linked data.

As observed by Zorich (2008), many of the early Digital Humanities Centres suffered from structural and service issues. This led to the risk of many Digital Humanities Centre being 'orphaned' as they were unable to take advantage of developments in web technology and participate in large-scale projects that would have ensured long term survival and relevance. Our aim is for that not to be the fate of HuNI.

6. Conclusion

The premise for HuNI, which ingests and aggregates data from many Australian datasets in the humanities and creative arts, was the recognition that cultural data is not insular. Researchers need to collaborate across disciplines, institutions and social locations. The aims of HuNI are ambitious. The type of virtual laboratory envisaged is based on a big science model. The Library is gaining a deeper understanding of humanities research in a digital world through its association with HuNI.

Many libraries are embracing new roles in relation to research data and offer a suite of research services building the institution's e-research capabilities. The effectiveness of the Library's contribution can be enhanced as we develop our understanding of the humanities research process and the preferences and thought processes underpinning humanities research methods (Nowviskie, 2014). Questions around governance and sustainability must be faced by all e-infrastructure projects and are particularly challenging within the context of the new capabilities that are required and the uncertainties of government funding. It is not surprising that alliances between libraries and digital humanities research initiatives are being formed and there are many opportunities emerging. At Deakin University that opportunity is now.

References

- ACRL Research Planning and Review Committee. (2012). 2012 top ten trends in academic libraries: A review of the trends and issues affecting academic libraries in higher education. *C&RL News*, (June), 311-320.
- Association of Research Libraries. (2014). SHared Access Research Ecosystem (SHARE). Retrieved April 6, 2014 from <http://www.arl.org/focus-areas/shared-access-research-ecosystem-share>
- Adams, J.L. & Gunn, K.B. (2012). Digital Humanities: where to start. *C&RL News*, (October) 536-9, 569
- Australia. Department of Finance. (2008). *Nation-building Funds Act 2008*. Retrieved April 6, 2014 from <http://www.finance.gov.au/investment-funds/nation-building-funds/act/>
- Burrows, T. (2013). A data-centred 'virtual laboratory' for the humanities: Designing the Humanities Networked Infrastructure (HuNI) service. *Literary and Linguistic Computing*. 28(4), 576. doi:10.1093/lc/fqt064
- Byron, J. (2008). *Submission to the review of the National Collaborative Research Infrastructure Strategy Roadmap*. The Australian Academy of the Humanities. Retrieved from http://www.humanities.org.au/Portals/0/documents/Policy/Submissions/text/POL2008_3.pdf
- Cornell University Library. (2014) Extending the semantic web. Retrieved 19 April 19, 2014 from <http://www.library.cornell.edu/node/3945>
- Cossham, A. F. (2013). Bibliographic records in an online environment. *Information Research* 18(3), 33-52.
- Crane, G. Babeu, A.& Bamman, D. (2007) eScience and the Humanities, *International Journal on Digital Libraries*, 7, 117-122.
- Crane, G. & Friedlander, A. (2008). More than a million: building the digital environment for the age of abundance. Retrieved from <http://www.clir.org/activities/digitalscholar/Nov28final.pdf>
- Friedlander, A. (2009). Asking questions and building a research agenda for digital scholarship. In *Working together or apart: promoting the next generation of digital scholarship*. Report of a workshop co-sponsored by the Council on Library and Information Resources and The National Endowment for the Humanities. Washington, DC: Council on Library and Information Resources Retrieved from <http://www.clir.org/pubs/reports/pub145/pub145.pdf>
- Hoeflich, M. H. (2007). Serendipity in the Stacks, Fortuity in the Archives. *Law Library Journal*, 99(4), 813-827.
- Linked Open Data for Libraries Archives and Museums (LODLAM). Retrieved from <http://lodlam.net/>
- McClellan III, J. E. (2005). Accident, Luck, and Serendipity in Historical Research. *Proceedings of the American Philosophical Society*, 149(1), 1-21. Retrieved from <http://www.jstor.org/stable/4598905>
- Neish, P. (2014). Linked Data: thinking big, starting small. Streaming with possibilities, VALA2014 Conference 3-6 February, Melbourne Convention and Exhibition Centre. Retrieved from <http://www.vala.org.au/vala2014-proceedings/vala2014-session-10-neish>

- Nowvskie B. (2014). Asking for it. [blog post] Retrieved from <http://nowvskie.org/2014/asking-for-it/>
- OCLC. (2012). Linked data for libraries [video] Retrieved from <http://m.youtube.com/watch?v=fWfEYcnk8Z8>
- Qan-Haase, A., & Martin, K. (2011). Rethinking Tradition: the Impact of Technology & the Loss of Serendipity on the Historical Research Process. In *Proceedings of the American Society for Information Science and Technology*, 48(1), 1–2.
- Turner, G. (2008). *Towards an Australian Humanities Digital Archive*. The Australian Academy of the Humanities. Retrieved from http://www.humanities.org.au/Portals/0/documents/Policy/Research/Towards_An_Australian_Digital_Humanities_Archive.pdf
- Verhoeven, Deb. (2014). Doing the Sheep Good: Facilitating Engagement in Digital Humanities and Creative Arts Research. Chapter forthcoming in Paul Arthur and Katherine Bode (eds), *Advancing 2 Digital Humanities: Research, Theory, Methods*, Palgrave Macmillan, 2014.
- Walters T. & Ruttenberg, J. (2014). SHared Access Research Ecosystem. *Educause Review* March/April, 56-7
- Weinberger, D. (2013). What I learned at LODLAM [blog post] Retrieved from <http://www.hyperorg.com/blogger/2013/06/22/what-i-learned-at-lodlam/>
- Winter, T. N. (1999). *Roberto Busa, S. J. and the invention of the machine-generated concordance*. Faculty Publications, Classics and Religious Studies Department, Paper 70, University of Nebraska-Lincoln Retrieved from <http://digitalcommons.unl.edu/classicsfacpub/70>
- Zorich, D.M. (2008). A survey of Digital Humanities Centers in the United States. Washington, DC: Council on Library and Information Resources Retrieved from <http://www.clir.org/pubs/abstract/pub143abst.html>