



AUSTRALIAN ACADEMY OF THE HUMANITIES
NATIONAL RESEARCH INFRASTRUCTURE CAPABILITY ISSUES PAPER
SEPTEMBER 2016

Submission Template

Name	Dr Christina Parolin
Title/role	Executive Director
Organisation	Australian Academy of the Humanities

INTRODUCTION

The Australian Academy of the Humanities (AAH) welcomes the opportunity to respond to the National Research Infrastructure Capability Issues Paper and recognises this as the first step in working towards a shared long-term vision of the conceptual foundations and strategic directions for Australia's research infrastructure capabilities.

The humanities together with the arts and social sciences (the HASS sector) form a sizeable part of Australia's research and innovation system. HASS researchers comprise 43% of the university-based research system, and HASS contributed 44% of the total number of units of evaluation in the Excellence for Research in Australia (ERA) initiative in 2012.¹ The HASS sector is not only sizeable but diverse, comprising some 50 disciplines at the four-digit field of research level. Researchers in these fields are therefore integral to the research system in Australia, as well as being vital partners with their colleagues in the STEM fields in interdisciplinary and cross-disciplinary research collaborations.

We welcome the paper's recognition of humanities research infrastructure needs and the critical role of cultural and collecting institutions. The order of magnitude of digital and 'big data' developments are transforming research practice in the humanities. The scale of existing data, together with expectations of future data creation, retention, storage and sharing presents both opportunities and challenges to future planning.

Three of the biggest challenges for research in the humanities include: data availability, skills development in digital data curation, and research translation. Research infrastructure needs to now achieve the scale required to transform research in these disciplines, in innovative interdisciplinary research, and drive global research collaborations around world-leading nationally significant infrastructure.

CONSULTATION QUESTIONS

Question 1: Are there other capability areas that should be considered?

While the Issues Paper is to be applauded for its 'whole of sector' approach, it has not fully addressed cross capability capacity. While the 'Underpinning Research' and 'Data for Research and

Discoverability' capabilities have been designed from the outset as cross-cutting capabilities, the same principle should apply across the other capabilities if the infrastructure roadmap is to enable advanced multi-disciplinary collaboration as the basis for addressing complex research problems, for example, in health, the environment, or social cohesion.

Question 2: Are these governance characteristics appropriate and are there other factors that should be considered for optimal governance for national research infrastructure.

The governance characteristics identified in the paper are broadly appropriate.

The AAH draws attention to four additional points:

- The governance model will need to have national reach and supra-institutional authority, and its members must enjoy the confidence and respect of the various stakeholder communities: HASS and STEM scholars, the e-research community, and the range of 'supply side' organisations, including the collecting and cultural and research institutions.
- Governance mechanisms should be consultative and collaborative and seek to build links and connections between the capability domains.
- There is great unmet demand for research infrastructure in the humanities sector. The scale of existing data, together with expectations of future data creation, retention, storage and sharing presents both opportunities and challenges to future planning. A clear and transparent priority setting process is essential to ensure efficient and value-for-money allocation of resources.
- Greater clarity about the purpose, scope and life cycle of infrastructure investments is also needed. Capability development should start with *a priori* strategic questions about what the infrastructure is intended to achieve. In developing a long-term strategy, Australia needs to get better at understanding the 'how' and 'why' and not simply the 'what' and 'who', which is where our evaluative mechanisms focus.

Question 3: Should national research infrastructure investment assist with access to international facilities?

Question 4: What are the conditions or scenarios where access to international facilities should be prioritised over developing national facilities?

Where appropriate, it will be vital to link to international projects and consortia. Cost, quality and scale are factors that should be taken into account in making determinations about when to invest in overseas access/facilities or develop national capabilities. These decisions should be made case-by-case in capability areas. We agree with the paper's assessment of the value that international collaboration on infrastructure can bring: global networks, a seat at the table, perspective on national and global problems, meeting Australia's obligations, and researcher access.

In terms of the 'Understanding Cultures and Communities' Capability, much of the research content maintained by collecting and cultural institutions, government entities and local societies and communities is uniquely Australian, so this is not something we can outsource or access overseas. Providing the platforms to make this content available, discoverable and researchable is **a national responsibility**. It is also our point of difference and the basis by which we contribute

more broadly to the global research effort. Further an important additional consideration for the 'Understanding Cultures and Communities' Capability in terms of international facilities is the critical need to address the issue of digital access to the many thousands of culturally significant Australian records held in overseas libraries, museums, archives and galleries.

Question 5: Should research workforce skills be considered a research infrastructure issue?

Question 6: How can national research infrastructure assist in training and skills development?

It is critically important that research workforce skills are identified as a research infrastructure issue. Our HDR graduates must be equipped with the skills to operate in an increasingly digital research environment, to understand and work with digital data, tools and structures. Infrastructure support for research training needs to focus not only on the production of research but also on the development of a researcher.

There is cross-sector consensus on the need to take into account the development and continuity of skills, to factor in remediation of skills gaps, as well as the enabling capacity of infrastructure to drive skill development across the sector. A sustainable funding environment is essential to building and retaining capacity and skills.

Two of the biggest challenges across the system are data capability and digital literacy. These are not confined to STEM. As the British Academy states, 'big data makes statistical and numerical understanding relevant across all disciplines'. With HASS researchers comprising 43% of the university-based research system, and with HASS teaching 65% of students,² there is a significant opportunity for a large scale, nationally coordinated research infrastructure to drive and enable digital literacy and data capabilities of this half of the higher education system in Australia.

E-research infrastructure is transforming the way researchers across the disciplines undertake and analyse research. Some of the exemplary research infrastructure projects in the humanities have been noted in the Issues Paper and in this submission: digital skills and capabilities are being developed right across the humanities disciplines. For example, the tools needed to analyse large data sets are already being generated from developments in the digital humanities. A visualization tool prototyped by humanities scholars at Stanford University's Centre for Spatial and Textual Analysis (CESTA) led to the commercialised software that allowed the analysis of the 11.5 million records in the Panama papers.³ Closer to home, text-mining methodologies developed at the University of Newcastle for analysing Shakespeare plays have contributed to biomedical research into cancer treatments.⁴

Research infrastructure investments will drive the skills and capability developments at both the discipline-specific and more generic training level. A national coordination role is needed to ensure there are mechanisms to support cross-capability conversations around skills and training for the 'demand' side. Consideration should also be given to leveraging the skills and expertise resting in 'supply' side organisations. This will both optimise the investment in research infrastructure by developing specialist skills across the sector, and importantly, facilitate new collaborations and knowledge exchange across and between the disciplines. This is the type of skills mixing that has been shown to drive innovation.⁵

Question 8: What principles should be applied for access to national research infrastructure, and are there situations when these should not apply?

Public investment in research infrastructure is a public good so the fundamental principle of open access should apply. There will be situations in which ethical and legal considerations take precedence. Data employed in humanities research is often culturally sensitive. Protocols for the use and re-use of data will need to be implemented, noting there is already some clear progress in this with the work of AIATSIS and ATSIDA.

Understanding Cultures and Communities

Question 24: Are the identified emerging directions and research infrastructure capabilities for Understanding Cultures and Communities right? Are there any missing or additional needed?

The AAH agrees with the broad directions and capabilities outlined here.

The importance of achieving national-scale research infrastructure to support and drive transformative research on Australian cultures and communities cannot be overstated. Humanities research, like that across the sector, is becoming increasingly technologically and data driven. The data – or ‘big content’ – that humanities researchers need for advanced research is generally a combination of public sector data and data held and managed by collecting institutions, and not primarily, as is the case for the majority of science data, by universities and other research institutions.

Data availability and discovery is one of the most pressing challenges for humanities researchers.

Since the advent of mass-data digitisation in the 1990s, Australian humanities researchers working in universities and other public and private sector organisations, as well as a growing community of citizen researchers, have seized on opportunities presented by access to digital data (evidence) on Australia’s history, identity and cultural life made available by national and local collecting institutions. The extensive use of the National Library of Australia’s Trove facility by researchers working on Australian society and culture is direct evidence of this – as the Issues Paper notes, it has ‘enabled a paradigm shift for humanities researchers’.

The available data across the nation’s cultural and collecting institutions, however, represents just a fraction of actual data stored. The bulk of the national cultural record including oral recordings, visual and written material, objects, plans, maps, celluloid film and art housed in libraries, museums, galleries, archives, community organisations, Indigenous communities, research institutions, government departments and private collections at the local, state and national level is neither conserved digitally nor accessible in digital format. Much of it is vulnerable to deterioration or destruction. Moreover, decisions around digitising content have generally not been developed primarily according to research priorities but rather by broader community and institutional responsibilities, including preservation requirements.

Data availability needs and challenges are also already increasingly extending to born digital content. Web archives and social media data, for example, will be significant sources of evidence for cultural and social research but there are major challenges around discoverability, access and storage that need urgent attention. Digitisation and the technology to create tools for digital research have also led to the development of exemplary discipline-specific digital resources and projects in the humanities and social sciences. Examples include those in archaeology (FAIMS), linguistics and ethnomusicology (PARADISEC), literary studies (AustLit), history (Founders & Survivors and the Prosecution Project) and law (AUSTLII) and public policy (APO). They are vitally important sources of data for Australian research but the ability to share and re-use data across platforms is limited and all are facing storage issues. Investment in the design and provision of ‘virtual laboratories’ or federated data sets (like the NECTAR-funded HUNI) have signalled a broader ambition of sharing data but are not yet integrated in the practices of the HASS sector more broadly.

Alongside Trove, there is also much to learn from the success of the Atlas of Living Australia (ALA). These exemplary data aggregators will be critical building blocks for the next generation facility described above which is needed to underpin the UCC Capability. Trove and the ALA provide both proof of concept and proof of demand.

What researchers working across the humanities and social sciences need now, and for the decade ahead, is a researcher-driven national-scale facility that can expand exponentially the discovery, access, data mining, curation, analysis and interpretation that is crucial for modern research, new forms of scholarship and innovation, and research translation. This will require a next-generation approach to big data availability and curation, enabling new research across diverse datasets in the HASS sector and driving cross-disciplinary collaborations with the science sector. It will require an open infrastructure platform which:

1. delivers prioritised **digitised research content** from Australia’s dispersed and diverse cultural record, allowing the discovery of records of Australian life which document who we are as a nation – our people, places and communities – or shows how Australians have experienced and interacted with our environment, or that underpins our understanding of cultures and communities here and around the world. Discovery of records impacts on all Australians but is a particular issue for Indigenous communities.
2. allows for **content development capacity** by providing researchers with the ability to make connections across a variety of data types (newspapers, diaries, oral recordings, archives, film and sound, material objects and artistic works) and data sets, including existing, and future discipline-specific digital resources and projects developed for historical and cultural research.
3. allows access to and **development of tools** – including high end analytical tools that will support semantic text analysis, data mining and analysis to manage big data curation process for a wide variety of research needs, along with the facility to build, share, enhance and re-use data sets.
4. involves active participation of both **researchers and communities** – further building our cultural record by contributing stories, artefacts and objects of our history and heritage, digitising and correction (cf: the volunteer portals of the Atlas of Living Australia and Trove,

and the community engagement models of PARADISEC, Founders & Survivors, and the Prosecution Project) and supporting the translation of research.

Given the volume and array of material available for digital capture, the research community needs a national facility that can:

- prioritise content digitisation in accordance with the needs of researchers
- support prioritised content digitisation to optimise data availability
- develop data capture methodologies and standards appropriate to different content types, and national interests and international collaboration
- address the ethical and legal issues surrounding copyright, ownership and access, including with reference to Aboriginal and Torres Strait Islands communities
- define data curation models and ensure these are widely adopted
- deliver digital data to research devices, including desktops/ mobile through high-capacity connections and cloud based storage
- provide a platform for the development of and access to tools to allow big data curation, sharing and re-use of data
- enable and drive cross-disciplinary research – both across HASS and with STEM, and
- mobilise citizen-researchers to assist with highly-targeted research projects.

A national-scale facility for the UCC capability would exponentially increase the speed by which the research process occurs (replacing time consuming searches of a variety of individual institutional databases and records). It will also enable as-yet unimagined connections to be drawn across datasets and platforms, allowing researchers to frame and investigate advanced research questions about Australian society, history, identity and our relationships with other cultures and societies. It would offer researchers unprecedented discoverability and access of untapped data archives; would be informed by research on issues of national priority and significance, including research on Indigenous cultures, traditions, languages and health; allow researchers to examine trends across time and place; and reach and obtain valuable new insights from local knowledge and local historical societies.

The Issues paper rightly identifies that Australia has a strong digital humanities field. The digital humanities are a 'broad church' encompassing a diverse set of disciplines and research practices. The UCC Capability is applicable across the full spectrum of humanities disciplines. Sophisticated digital tools are redefining how we understand past cultures and communities, and events such as WW1 (looking at injuries and trauma to the body in new ways, for example).⁶

The following is just one example of the type of research which would be enabled by the new national facility envisaged above:

The economies, peoples and states of the Asia-Pacific are increasingly related through dense and complex relational networks of institutions, including businesses, NGOs, cultural and educational institutions, and government agencies among others. All produce public data embedded in digital documents. But the complexity of relations

and diversity of sources inhibit efficient and effective research. A mechanism for identifying research priorities (eg focussing on Chinese and Australian institutional networks) to inform digital data resource building to meet these priorities would enable research into international trade, security, public diplomacy ('soft power'), cultural engagements, educational exchanges, people-movements and other critical areas of HASS research in the priority domain.

The UCC capability has purchase well beyond the humanities – it will also support and connect with social sciences and STEM research and drive cross-disciplinary collaborations, in areas requiring Australian data such as climate science, health, demography, sociology, anthropology, economic history, criminology, botany, ecology, geology, palaeontology, ancient DNA science and oceanography.

International and interdisciplinary collaborations are also increasingly being forged around these new technologies, such as the project to reconstruct an ancient Egyptian woman drawing on expertise in medicine, science, history, archaeology, and art using 3D printing and computerised scanning.⁷

HASS data presents unique challenges for computer scientists, in computer vision and deep learning, but it is only by facilitating access that computer science will be able to drive research agendas forward in collaboration with HASS—for shared research outcomes.

Although designed initially to meet the needs of researchers, such a facility has the potential to reach new communities and raise awareness of Australia's cultural heritage nationally and internationally. It also opens possibilities for advanced education infrastructure ready to be leveraged by educational institutions, contributing to the wider digital literacy agenda, and broadening the skills and capacity development for students and researchers who are not directly involved in the STEM fields (note our earlier comments on research training).

The UCC capability will require a governance model connecting data-users with data-providers, a *data prioritisation* mechanism to reflect active researcher demands within the framework of national priorities, a *co-resourcing* model supporting long term investment in data digitisation and high-speed delivery to meet researcher and other end user needs at device/desktop level, and a *reflexive feedback mechanism* alerting collecting institutions to changing researcher demands, drawing new data developments and content development capacity to the attention of researchers, and informing development of citizen-researcher participants in priority research projects.

Question 25: Are there any international research infrastructure collaborations or emerging projects that Australia should engage in over the next ten years and beyond?

The facility envisaged above would have vast international reach – it will open Australia's cultural record to the world, encouraging and enabling research on Australian topics by international researchers, and creating the conditions for comparative and collaborative research efforts. Importantly, it should also allow remote connection with thousands of culturally significant Australian records held in overseas libraries, museums, archives and galleries. As world-leading infrastructure, such a facility will generate international interest and opportunities for collaboration.

To our knowledge, there is nothing that currently exists on the scale and scope imagined for the national-scale, researcher-driven facility envisaged above.

There are some recent and exciting developments in research infrastructure internationally which are looking to more directly connect the needs of researchers with data residing in cultural and collecting institutions and local communities. Of the seven infrastructures in the social sciences and humanities in Finland's *Strategy and Roadmap for Research Infrastructures 2014-2020*, two are open libraries (FinELib – the Finnish National Electronic Library; and Finna – the Public Interface of the Finnish National Digital Library).⁸ The social sciences and humanities infrastructure is the only infrastructure to service all disciplines in Finland.

In terms of exemplary digital infrastructure for research on culture and society, the Digital Public Library of America (DPLA) and Stanford University collaboration is of particular reference for the UCC Capability.⁹

Another international facility that will provide opportunities for close engagement for the UCC Capability is Europeana – pan European infrastructure which brings open source standardisation to the digitisation of cultural heritage across the European Union.¹⁰ It is clearly an internationally significant resource with potential for close collaboration, particularly given the future plans outlined in 'strategy 2020' which aims to 'build a bigger set of high quality material and a shared infrastructure that enables re-use and creates value for all stakeholders.'¹¹

Another European initiative with collaborative potential is the new project 'Elaboration of a Concept for a European Academies Internet Gateway for the Social Sciences and Humanities (AGATE)' led by the Union of the German Academies of Sciences and Humanities, in close collaboration with ALLEA, the federation of All European Academies. It is bringing together participants from the 'European science academies, infrastructure projects in the fields of digital humanities and social sciences, and related infrastructures, such as DARIAH, CLARIAH(NL)/CLARIN, Europeana, OpenAire, and the German Council for Scientific Information Infrastructures' to conduct a feasibility study for the project.¹²

Question 26: Is there anything else that needs to be included or considered in the 2016 Roadmap for the Understanding Cultures and Communities capability area?

Of all the capabilities, 'Understanding Cultures and Communities' potentially has the largest disciplinary remit. Responding to today's global, social, cultural, health, environmental and economic challenges requires specialist knowledge of people, societies and cultures. In the table that maps the capabilities to the National Science and Research Priorities (SRPs) (at Attachment C in the paper) the assessment that the 'Understanding Cultures and Communities' capability only addresses the Transport priority requires particular remedy. At a minimum, it is applicable to Food; Soil and Water; Cybersecurity; Health; and Environmental Change.

The AAH also notes that the range of priorities for Australia's future prosperity extends beyond the existing SRPs into areas of social and cultural transformation, such as tackling income and social inequality, domestic violence, and educating for the future.

Underpinning Research Infrastructure

Question 30: Are the identified emerging directions and research infrastructure capabilities for Underpinning Research Infrastructure right? Are there any missing or additional needed?

Question 31: Are there any international research infrastructure collaborations or emerging projects that Australia should engage in over the next ten years and beyond?

Question 32: Is there anything else that needs to be included or considered in the 2016 Roadmap for the Underpinning Research Infrastructure capability area?

E-research infrastructure and digitisation are the big areas of growth for the humanities, like other sectors of the research system. The data intensive research needs of the HASS sector will need to be kept in view in the development of this capability. Further, the world-leading innovations in digital infrastructure that have been developed in the cultural and collecting sector in Australia must be properly recognised and leveraged – and underpinning national research infrastructure extended to this vital sector of the Australian research landscape.

Data for Research and Discoverability

Question 33: Are the identified emerging directions and research infrastructure capabilities for Data for Research and Discoverability right? Are there any missing or additional needed?

Question 34: Are there any international research infrastructure collaborations or emerging projects that Australia should engage in over the next ten years and beyond?

Question 35: Is there anything else that needs to be included or considered in the 2016 Roadmap for the Data for Research and Discoverability capability area?

Clearly this Capability will be of direct relevance to the Understanding Cultures and Communities Capability given the importance of data availability, discoverability and research in the HASS sector.

In 2008 the Academy completed a scoping study on the development of a national scholarly digital archive in the humanities. The aim of this scoping study, conducted by Professor Graeme Turner FAHA, was to compile a sound evidence base for a prioritised, scalable list of proposed resources for such an archive in order to guide any potential future investment in this area. The principles guiding the selection and prioritisation of materials for digitisation which that report identified are still valid: demonstrable demand; academic or research significance; and public interest.¹³

We would be happy to elaborate on any of the feedback in this submission. Please direct your initial enquiries to our Executive Director,

Dr Christina Parolin.
Professor John Fitzgerald FAHA
President

The AAH is one of Australia's four Learned Academies, established to advance knowledge and the pursuit of excellence in the humanities for the benefit of the nation. A key role of the AAH is to provide independent expert advice to government and policy makers, promoting the social significance of humanities scholarship and its vital importance in shaping effective public policy.

¹ Turner, G. and Brass, K. (2014) *Mapping the Humanities, Arts and Social Sciences in Australia*, Australian Academy of the Humanities, Canberra, p.2

http://www.humanities.org.au/Portals/0/documents/Policy/Research/MappingProject/txt/Mapping_HASS_Aust_Fina|Report_All_Oct2014.pdf

² Turner, G. and Brass, K. (2014) *Mapping the Humanities, Arts and Social Sciences in Australia*, Australian Academy of the Humanities, Canberra.

³ <http://news.stanford.edu/thedish/2016/06/03/visualization-tool-prototyped-by-stanford-humanities-scholars-aids-the-investigation-of-panama-papers/>

⁴ Craig, H (2015), 'Enlisting Shakespeare to help fight cancer', *The Power of the Humanities*, Australian Academy of the Humanities, <http://www.humanities.org.au/Publications/PoweroftheHumanities/EnlistingShakespeare.aspx>

⁵ Cunningham, S., Theilacker, M., Gahan, P., Callan, V. and Rainnie, A. (2016) *Skills and capabilities for Australian enterprise innovation*. Report for the Australian Council of Learned Academies, www.acola.org.au

⁶ <http://www.army.gov.au/our-work/unrecovered-war-casualties-army/fromelles/the-fromelles-project>

⁷ <http://mdhs.unimelb.edu.au/news-and-events/brought-to-life>

⁸ The Finnish Research Infrastructure Committee (2013) *Finland's Strategy and Roadmap for Research Infrastructures 2014-2020*, p. 16.

http://www.aka.fi/globalassets/awanhat/documents/firi/tutkimusinfrastruktuurien_strategia_ja_tiekartta_2014_en.pdf

⁹ <https://dp.la/info/>

¹⁰ <http://www.europeana.eu/portal/en>

¹¹ <http://strategy2020.europeana.eu/>

¹² <http://www.sagw.ch/sagw/laufende-projekte/digital-humanities.html>

¹³ Turner, G. (2008) *Towards and Australian Digital Archive*, Australian Academy of the Humanities, Canberra.

http://www.humanities.org.au/Portals/0/documents/Policy/Research/Towards_An_Australian_Digital_Humanities_Archive.pdf