



# Mapping International Research Infrastructures for the Humanities, Arts and Social Sciences

A REPORT FOR THE COMMONWEALTH DEPARTMENT OF EDUCATION

FEBRUARY 2020



This report has been prepared by the Australian Academy of the Humanities using multiple sources of data. The analysis and findings are subject to the limitations of the data used. While every effort has been made to ensure the accuracy of that information, the Australian Academy of the Humanities does not make any warranty, express or implied, regarding it.

## ADVISORY GROUP

Professor Duncan Ivison FAHA (Chair), Deputy Vice-Chancellor Research, University of Sydney

Professor Peter Anderson, Director, National Indigenous Research and Knowledges Network

Dr Marie-Louise Ayres, Director General, National Library of Australia

Anne-Marie Lansdown, Deputy CEO, Universities Australia

Dr Merran Smith, Director, Population Health Research Network

Professor Mark Western FASSA, Director, Institute for Social Science Research, University of Queensland

## PROJECT MANAGER

Dr Kylie Brass, Director Policy and Research, Australian Academy of the Humanities

## CONTRIBUTING AUTHORS

Kylie Brass

Peter Elford

Ingrid Mason

Christina Parolin

# Contents

<b>1</b>	<b>Executive Summary</b>	<b>4</b>
1.1	Introduction.....	4
1.2	Scope .....	4
1.3	Key findings – International Mapping .....	5
1.4	Key lessons – International Mapping .....	5
1.5	Findings from the Gap Analysis .....	6
1.6	Models to consider for HASS scoping study .....	6
1.7	Potential pathways and priorities for Australia.....	7
<b>2</b>	<b>International HASS Research Infrastructures</b>	<b>8</b>
2.1	Introduction.....	8
2.2	International HASS infrastructures selected .....	8
2.3	Selection rationale.....	10
2.3	National profile: The Netherlands .....	16
2.4	Findings .....	19
2.5	Lessons .....	22
2.6	Opportunities for collaboration .....	30
2.7	Models for further analysis for HASS scoping .....	32
2.8	Conclusion .....	34
<b>3</b>	<b>Gap Analysis</b>	<b>35</b>
3.1	Introduction.....	35
3.2	The Australian landscape .....	35
3.3	Findings from the gap analysis .....	38
<b>4</b>	<b>Potential Priorities</b>	<b>41</b>
4.1	Australian HASS research infrastructure entity.....	41
4.2	HASS Research Data Commons .....	43
4.3	Social Sciences and Languages Data Hub .....	43
4.4	Indigenous Data Framework .....	44
4.5	National Digitisation Capability .....	44
4.6	Digital HASS Peak Capability.....	45
4.7	Implementation.....	45
	<b>Appendix A – Research phases</b>	<b>48</b>
	<b>Appendix B: HASS Research</b>	<b>49</b>
	<b>Appendix C: HASS Data</b>	<b>51</b>
	<b>References</b>	<b>54</b>

# 1 Executive Summary

---

## 1.1 Introduction

Australian humanities, arts and social sciences (HASS) researchers are working on pressing social and cultural challenges: from maximising the benefits of digital technologies for young people to providing the evidence base for decisions about education, employment, public health, and social and cultural policy.

The *2016 National Research Infrastructure Roadmap* (the Roadmap) identified the need for national-scale infrastructure for humanities, arts and social science (HASS) research to “drive transformations in the way researchers discover, access, curate and analyse social and cultural data”.

The Australian Academy of the Humanities was engaged by the Department of Education to map key international HASS research infrastructures – across humanities, arts, social sciences and Indigenous research – to inform the development of Australian HASS national research infrastructure (NRI).

This report outlines key findings from the international mapping and Australian gap analysis, drawing on lessons from international investments and specifically to identify common patterns in nationally or regionally significant research infrastructures, timeframes and investment levels required to establish, maintain and scale up national research infrastructures, and what may be required to collaborate and compete in data and technology-intensive HASS research globally.

Although the research infrastructures mapped for the project may fall outside what would traditionally be defined as ‘national research infrastructure’ as part of Australia’s National Collaborative Research Infrastructure Strategy (NCRIS) program, they are nevertheless national or pan-national scale infrastructures. This provides an opportunity to reconsider the definition applied in Australia in terms of international comparators.

## 1.2 Scope

The analysis in this report is primarily informed by a detailed mapping of research infrastructures in Europe, the UK, the Netherlands, USA, Canada and New Zealand – with the Netherlands serving as a case-study of best-practice national approach to research infrastructure development across the HASS community, including stakeholder organisations.

The infrastructures highlighted in the report are diverse in terms of scope, type and organisation. The infrastructures surveyed meet a variety of research needs and development capabilities – from large-scale, integrated and sustainable data services – including storage and use; digitisation of printed text for analysis purposes; digital platforms for discovery; high performance computing; interoperability across platforms and/or data sets; data sovereignty and privacy.

### 1.3 Key findings – International Mapping

International investments in HASS research infrastructure have shown pathways for infrastructure investment in:

- > heritage research, data and technologies – building capabilities in characterisation technology and pattern recognition;
- > language research, data and technologies – building capabilities in informatics, semantics and AI; and
- > social research, data and technologies – building capabilities in internet of things, civic technologies and precision services.

HASS national research infrastructures mapped for this project show staged investment building on existing research practices, peer networks, investments in research and infrastructures, and capabilities, and are heavily oriented to improving existing data assets and processes.

The international mapping identifies the emergence of new and unique HASS research infrastructures that are heavily data, information and computer science enabled and provide different arrangements of data custody, new data and large scales of data that lead to radically new research techniques.

In last ten years, international HASS research infrastructures have focused on equipping HASS researchers with the tools and techniques to take on big data challenges. In the social sciences, there have been major investments to align data archives, survey instruments and methods. Humanities and arts investments have focused on experimentation, building capacity, and platforms and digitisation technologies for cultural heritage.

In Europe, there are new large-scale text and audio-visual data collections curated and co-located with cloud and high-performance computing at a national level. They are oriented to enabling novel exploitation of new, curated, diverse and large data assets, utilising cloud and peak computing facility specialists and technologies.

### 1.4 Key lessons – International Mapping

International mapping tells a story of what can be achieved through national research infrastructure stimulus, indicating where Australia might seek to ‘leapfrog’ and what pitfalls to avoid. The report identified seven broad lessons for the development of HASS NRI in Australia:

1. **Laying the foundations:** There is evidence of staged investment and pathways for the development of infrastructure over time. Infrastructures have transitioned from program funding to ‘landmark’ status.
2. **Capacity-building and community-building:** supporting and encouraging researchers undertaking digitally enhanced research is a fundamental part of the process of research infrastructure development and operation.
3. **Consolidation:** funding programs are moving from seeding infrastructure build to focus on cross-cutting initiatives designed to accelerate collaboration and common infrastructure at scale, advancing interoperability, comparative research and support for big data research.

4. **Fragmentation is an issue:** this has been identified as an area to address in humanities, arts and heritage, with programs funded to address the need for stronger coordination.
5. **Balancing domain-specific and opportunities for common infrastructure development:** In Europe domain-specific development is giving way to a trend for consolidation and common infrastructure development. Yet there are indications that while some of the challenges will be solved by common approaches, there is still a need for specialised, domain-specific infrastructure build.
6. **Developing complex governance models and partnerships:** the strong trend towards consolidation across the HASS research infrastructures is leading to increasingly sophisticated governance and co-investment models.
7. **Alignment of strategies:** National strategies also recognise the critical importance of participating in, and contributing to, international and regional research infrastructures.

## 1.5 Findings from the Gap Analysis

Australia does not have any nationally comparable HASS research infrastructures, and this represents a significant gap in national research capacity. Consequently, research data assets currently enabling HASS research either do not exist in Australia, or are institutionally hosted and uncoordinated, and largely not FAIR (findable, accessible, interoperable and reusable).

Patterns in the Australian landscape look like the pre-stimulus phase in Europe in which Framework Program funding seeded key HASS strengths. Australia has made some investments via ANDS/NeCTAR/RDS and through the Australian Research Council's Linkage Infrastructure, Equipment, and Facilities scheme and there are Australian comparators to many of the European research infrastructures operating at an institutional level that are well placed to be evolved into national capabilities.

Much of Australia's public sector data – both social and cultural – is locked up and/or underutilised in HASS research, and in other disciplines. It is clear from the mapping survey that in Europe many of the cross-government jurisdiction challenges that are typical of the Australian federation have been addressed through formal Memoranda of Understanding (MoUs) and EU incentives (fiscal and regulatory).

## 1.6 Models to consider for HASS scoping study

The report suggests seven key national-scale infrastructures warrant further analysis to address gaps in Australian national research infrastructure, as part of a HASS scoping study. These include:

1. IMPACT Centre of Competence, providing state-of-the-art tools, services and facilities in document imaging, language technology and the processing of historical text.
2. Europeana, Europe's digital platform for cultural heritage.
3. SAIL Databank – providing robust secure storage and use of anonymised person-based data for research to improve health, well-being and services.

4. CLOSER, hosting longitudinal research and brings together eight world-leading social longitudinal studies.
5. HathiTrust, digitisation, text and data mining services.
6. CLARIAH, providing access to large collections of digital data and to innovative applications for the processing of data.
7. ODISSEI, a platform for the survey and collection of social science data, designed to draw the efforts of research more closely to feeding into social policy directions and outcomes.

## 1.7 Potential pathways and priorities for Australia

Australia is in a position to plan out the HASS NRI ecosystem based on the lessons identified in this international mapping report. The following priorities for Australia have emerged from the mapping and gap analysis:

1. Given the strong trend internationally towards consolidation across the HASS research infrastructures, Australia should consider a coordinating entity to create focus, clarify responsibility and reduce complexity for the development of requirements and roadmaps for Australian HASS research infrastructures.
2. Adopt a portfolio approach to program and infrastructure delivery to foster the capacity and capabilities of the Australian HASS sector over the next decade:
  - a. Deliver a HASS Research Data Commons
  - b. Establish a Social Science and Language Data Hub
  - c. Develop an Indigenous Data Framework
  - d. Establish a National Digitisation Capability
  - e. Develop a Digital HASS Peak capability

## 2 International HASS Research Infrastructures

---

### 2.1 Introduction

Internationally, governments have recognised both the potential and need for intensive digital research infrastructure for humanities, arts and social science researchers for almost two decades. The mapping undertaken for this project sought to identify key international infrastructures to inform Australia's planning for HASS research infrastructure. The aims of the project were twofold: to provide evidence-based advice on the scale and impact of public investment in international HASS research infrastructure; and an understanding of the structure and approaches to international HASS research infrastructure.

The analysis in this report is primarily informed by a detailed mapping of research infrastructures in Europe, the UK, the Netherlands, USA, Canada and New Zealand. A range of infrastructure stimuli have enabled research fields to build capability via targeted calls, such as through successive European Framework programs. Europe, Canada, and the Netherlands have also engaged in strategic roadmapping exercises, which have scoped HASS capabilities, and the UK is currently undergoing the first strategic national roadmapping of its own – which includes HASS.

An in-depth case-study of the approach to HASS research infrastructure development in the Netherlands is provided as an exemplar of a national approach to investment.

### 2.2 International HASS infrastructures selected

Exemplar infrastructures in these jurisdictions were identified with the assistance of a project Advisory Group – experts drawn from key stakeholder communities – HASS research, Indigenous research, the cultural and collecting sector, and universities.

The project adopted a broad definition used by the European Strategy Forum on Research Infrastructures (ESFRI) and the EU Framework Program:

Facilities, resources and services that are used by the research and innovation communities to conduct research and foster innovation in their fields. They include: major scientific equipment (or sets of instruments), knowledge-based resources such as collections, archives and scientific data, e-infrastructures, such as data and computing systems and communication networks and any other tools that are essential to achieve excellence in research and innovation.

The project was also informed by the UK's first national roadmapping exercise (in progress) which takes an 'ecosystem' view to scope the broadest range of 'research and innovation' infrastructure.

The research infrastructures mapped for the project may fall outside what would traditionally be in scope as part of Australia's National Collaborative Research Infrastructure Strategy (NCRIS) program. Infrastructure development in HASS, specifically the question of whether to develop existing infrastructures or build new capability, will require strategic investment decisions from a range of organisations including the government.

The infrastructures selected for analysis span 'humanities, arts and social sciences' domains; 'heritage' infrastructures which have successfully brought together research and cultural institutions; and 'Indigenous' research infrastructures.

Some of infrastructures mapped for this study are relatively small, institutionally based programs, others are pan-European landmark efforts. The mapping and gap analysis were carried out contiguously with the Advisory Group asked to identify key areas of Australia's strengths, gaps and prospects at the outset of the project. The mapping work was not intended to be comprehensive; it was necessarily selective.

#### 2.2.1. Humanities & Arts/ Heritage

- > CLARIN – Common Language Resources and Technology Infrastructure (Pan-national/ European)
- > DARIAH – Digital Research Infrastructure for the Arts and Humanities (Pan-national/ European)
- > CLARIAH – Common Lab Research Infrastructure for the Arts and Humanities (Netherlands)
- > EU E-RIHS – European Research Infrastructure for Heritage Science (Pan-national/ European)
- > Europeana (Pan-national/ European)
- > Impact Centre of Competence (Digitisation) (Pan-national/ European)
- > HathiTrust (USA)

#### 2.2.2. Social Sciences

- > CLOSER – Cohort and Longitudinal Studies Enhancement Resources (Pan-national/ European)
- > CESSDA Consortium of European Social Science Data Archives (Pan-national/ European)
- > ESS – European Social Science Survey (Pan-national/ European)
- > SAIL Databank (Pan-national/ European)
- > SHARE – Survey of Health, Ageing and Retirement in Europe (Pan-national/ European)
- > ODDISEI – Open Data Infrastructure for Social Science and Economic Innovations (Netherlands)

### 2.2.3 Indigenous

- > FNIGC – First Nations Information Governance Centre (Canada)
- > Te Mana Raraunga – Māori Data Sovereignty Network (New Zealand)

### 2.2.4 HPC Facility

- > Compute Canada (Canada)

## 2.3 Selection rationale

All the infrastructures selected for this mapping exercise are focussed on transforming access to and analysis of cultural and social data for research (Table 1).

The work we have undertaken in this report ranges over infrastructures with very different funding sources and involves a range of government stakeholder departments and agencies. We have sought to understand how other jurisdictions have solved problems relating to:

- > Building relationships and partnerships across sectors
- > Enabling research access to government data
- > Bringing together of cultural and research infrastructures to achieve efficiencies for respective stakeholders, particularly around digitisation challenges
- > Timing, staging and scale of investments – specifically, whether capability should lead investment, or whether investment should build capability.
- > The types of investment (is it institutional, program funding, or strategic?) to best meet research needs and achieve impact.
- > Working towards solutions for sensitive data storage and access.

The infrastructures mapped are diverse in terms of scope, type and organisation. Selections were based on success in terms of longevity (indicating recurrent funding based on evidence of impact) and meeting research needs of the respective domain communities.

The infrastructures surveyed meet a variety of research needs and development capabilities – from large-scale, integrated and sustainable data services – including storage and use; digitisation of printed text for analysis purposes; digital platforms for discovery; high performance computing; interoperability across platforms and/or data sets; data sovereignty and privacy.

All infrastructures are operational and well-established. The only infrastructures listed as not being at the operational phase is the new large-scale European Research Infrastructure for Heritage Science (EU E-RIHS) heritage science program.

**Table 1: International infrastructures mapped for this project – at a glance**

<i>Domains</i>	Research Infrastructure	Scope	Infrastructure type	Start	Duration	Investment
<i>Social Sciences</i>	European Social Survey (ESS)	A research-driven cross-national survey measuring the attitudes, beliefs and behaviour patterns of diverse populations of Europe across more than thirty nations. Conducted across Europe since its establishment in 2001. Through the participating countries, ESS operates as a pan-national data discovery service. Targeted at survey methodology and interoperability of data infrastructure to support comparative analysis. <a href="http://www.europeansocialsurvey.org/about/">http://www.europeansocialsurvey.org/about/</a>	Archive or repository Data service Discovery platform Capability building Community building Information sources Research practices and methods Research tools/ platforms	2002	16-20 Years	\$16-20M
	Survey of Health, Ageing and Retirement in Europe (SHARE)	A multidisciplinary and cross-national panel database of micro data on health, socio-economic status and social and family networks of about 140,000 individuals aged 50 or older (around 380,000 interviews). Covers 27 European countries and Israel. Pan-national cooperation around survey methods and data collection of ageing populations in Europe. Support for cross-cutting infrastructure to advance interoperability, comparative research and support for big data research. <a href="http://www.share-project.org/home0.html">http://www.share-project.org/home0.html</a>	Archive or repository Data service Discovery platform Capability building Community building Information sources Research practices/ methods Research tools/ platforms	2004	11-15 Years	\$21-50M
	Consortium of European Social Science Data Archives (CESSDA)	Provides large-scale, integrated and sustainable data services to the social sciences. It brings together social science data archives across Europe, with the aim of promoting the results of social science research and supporting national and international research and cooperation. <a href="https://www.cessda.eu/About">https://www.cessda.eu/About</a>	Archive or repository Data service, Discovery platform Capability building Community building Research practices and methods	2006	11-15 Years	\$101-150M

Social  
Sciences

SAIL Databank	<p>SAIL (Secure Anonymised Information Linkage) provides secure access to demographic, health, social and education data from the resident population of Wales – a flagship for the robust secure storage and use of anonymised person-based data for research to improve health, well-being and services. Now powered by the UK Secure e-Research Platform (UKSeRP).</p> <p><a href="https://saildatabank.com/about-us/overview/">https://saildatabank.com/about-us/overview/</a></p>	<p>Archive or repository Data service Discovery platform Capability building Community building Software service Information sources Research practices and methods HPC service Cloud storage/compute Research tools/ platforms</p>	2007	11-15 Years	\$21-50M
Cohort and Longitudinal Studies Enhancement Resources (CLOSER)	<p>Hosts longitudinal research, brings together eight world-leading longitudinal studies with participants born throughout the 20th and 21st centuries. Its work maximises the use, value and impact of these, and other, longitudinal studies to help improve understanding of key social and biomedical challenges.</p> <p><a href="https://www.closer.ac.uk/home/what-we-do/">https://www.closer.ac.uk/home/what-we-do/</a></p>	<p>Archive or repository Data service Discovery platform Capability building Community building Information sources Research practices and methods Research libraries</p>	2012	6-10 Years	\$11-15M

Humanities & Arts	Common Language Resources and Technology Infrastructure (CLARIN)	Provides access to digital language resources for researchers, students, and citizen-scientists, especially in the humanities and social sciences, through single sign-on access. Offers solutions and technology services for deploying, connecting, analysing and sustaining digital language data and tools. Enables advanced analytics and new types of services and technologies. Links between heritage and language infrastructures, and new industries. <a href="https://www.clarin.eu/content/vision-and-strategy">https://www.clarin.eu/content/vision-and-strategy</a>	Archive or repository Data service Discovery platform Capability building Community building Information sources Research practices and methods Research tools/ platforms Language data and tools	2006	11-15 Years	\$16-20M
	Digital Research Infrastructure for the Arts and Humanities (DARIAH)	Enhances and supports digitally enabled research and teaching across the arts and humanities. DARIAH is a network of researchers, expertise, information, knowledge, content, methods, tools and technologies from its member countries. It develops, maintains and operates an infrastructure in support of ICT-based research practices and sustains researchers in using them to build, analyse and interpret digital resources. <a href="https://www.dariah.eu/about/dariah-in-nutshell/">https://www.dariah.eu/about/dariah-in-nutshell/</a>	Discovery platform Capability building Community building Information sources Research practices and methods Research libraries Research tools and platforms	2006	11-15 Years	\$11-15M
	Common Lab Research Infrastructure for the Arts and Humanities (CLARIAH)	CLARIAH is a distributed infrastructure for the humanities and social sciences. It builds on various infrastructure and research projects carried out both nationally and internationally. Its focus areas are linguistics, social and economic history media studies, and text. These areas function as precursors for other disciplines and together comprise all forms of data: text, image, audio-visual material and structured data (databases). <a href="https://www.clariah.nl/en/">https://www.clariah.nl/en/</a>	Discovery platform Capability building Community building Information sources Research practices and methods Research libraries Research tools and platforms Research pilots Dissemination	2012-2024	12 Years	\$42.3M

Heritage

IMPACT Centre of Competence	A not for profit organisation with the mission to make the digitisation of historical printed text “better, faster, cheaper”. It provides tools, services and facilities to further advance the state-of-the-art in the field of document imaging, language technology and the processing of historical text. <a href="https://www.digitisation.eu/">https://www.digitisation.eu/</a>	Discovery platform Capability building Community building Software service Information sources Digitisation service	2008	11-15 Years	\$21-50M
Europeana	Europe’s digital platform for cultural heritage, Europeana is an exemplar in the advancement of cultural heritage interoperability. An aggregator service, and heavily supported by the European Library, a hub of 48 national and research libraries in Europe. It is one of the European Commission’s Digital Service Infrastructures (DSI). “As a DSI, Europeana’s objectives are to innovate the aggregation infrastructure, boost the distribution infrastructure and work towards long-term financial stability through business model innovation. All of this helps make sure that Europe’s businesses and people reap the full benefits of the technological revolution in digital services in culture.” <a href="https://pro.europeana.eu/our-mission/history">https://pro.europeana.eu/our-mission/history</a>	Data service Discovery platform Capability building Community building Software service Information sources Research tools and platforms	2008	11-15 Years	\$101-150M
HathiTrust	Text and data mining infrastructure partnership of academic and research institutions, offering a collection of millions of titles digitised from libraries around the world. Capacity for digitisation, text and data mining services to be shared across academic libraries. <a href="https://www.hathitrust.org/">https://www.hathitrust.org/</a>	Archive or repository Data service Discovery platform Capability building Community building Software/ Information Research practices and methods Research libraries Digitisation service Cloud storage/compute Research tools/ platforms	2008	11-15 Years	Co-investment through membership and investment through Google deal undisclosed

	European Research Infrastructure for Heritage Science (E-RIHS)	In pre-operational phase. Aims to deliver integrated access to expertise, data and technologies through four platforms (heritage interpretation, preservation, documentation and management) to support heritage science. Mission to deliver integrated access to expertise, data and technologies through a standardized, coordinated approach. <a href="http://www.e-rihs.eu/">http://www.e-rihs.eu/</a>	Discovery platform Capability building Community building Information sources Research practices and methods Digitisation service Research tools and platforms	2016	3-5 Years	\$6-10M
<i>Indigenous</i>	First Nations Information Governance Centre	Supports data sovereignty and the development of information governance and management at the community level through regional and national partnerships. Focused surveys and data capture for a range of health and wellbeing objectives where they impact on First Nations peoples in Canada. It adheres to free, prior and informed consent, respects nation-to-nation relationships, and recognises the distinct customs of nations. <a href="https://fnigc.ca/about-fnigc/mission.html">https://fnigc.ca/about-fnigc/mission.html</a>	Data service Discovery platform Capability building Community building Information sources Research practices and methods	2010	6-10 Years	No national policy driven investment, survey project funding.
	Te Mana Raraunga - Māori Data Sovereignty Network	Established to advocate for Māori rights and interests in data to be protected as the world moves into an increasingly open data environment” <a href="https://www.temanararaunga.maori.nz/">https://www.temanararaunga.maori.nz/</a>	Capability building Community building Information sources	2015	3-5 Years	No national policy driven investment, short-term research project funding.
<i>HPC</i>	Compute Canada	Provides essential ARC services and infrastructure for Canadian researchers and their collaborators in all academic and industrial sectors. Specific support for humanities and social sciences built into existing national computational services. <a href="https://www.computecanada.ca/about/">https://www.computecanada.ca/about/</a>	Capability building HPC service Cloud storage/compute service Local storage Research tools and platforms	2016	3-5 Years	\$3-5M

## 2.3 National profile: The Netherlands

The Netherlands offers an instructive national model for Australia. Although the Netherlands and Australia are not direct comparators in terms of size, there are factors that make the Netherlands a useful template for Australia.

The Dutch national research infrastructure roadmap (2016-2020) developed by the Netherlands Organisation for Scientific Research (NWO) includes two new entries: CLARIAH (Common Lab Infrastructure for the Arts and Humanities – which brings together CLARIN and DARIAH) and ODISSEI, a data facility, observatory, laboratory and hub led by CBS (national statistics agency), SURFSara (NREN), CentERdata (data archive, institutionally based), NWO (national research infrastructure funder) – with 30 participating organisations.

These large-scale infrastructures are intended to meet the needs of humanities and arts, and social science researchers (respectively). In the Netherlands, there is a drive to strengthen and integrate national social science and humanities research infrastructure. In the broader EU context, a Social Science and Humanities Open Cloud (SSHOC) (as part of the European Open Science Cloud initiative) is in development and NWO collaborates internationally and is connected into European research infrastructures, research programs, peer networks and joint programming initiatives.

The national profile for the Netherlands indicates EU subsidies operating prior to HASS research infrastructure investment appearing on the NWO national research infrastructure roadmap; and identifies Data Archiving and Networked Services (DANS) – a data archiving service provider for researchers in the Netherlands – support for HASS (within the national landscape) as a coordinator of EU subsidy and provider of social science data infrastructure services.

Analysis of Dutch spending on digital infrastructure (enabling research) indicates sizeable EU investments, that are leveraged and over time increased structural spending nationally to: include two new HASS research infrastructures the national roadmap and investment plan, and to fund DANS as critical part of that agenda. NWO allocated funds to support existing HASS research practices in two funding streams (as innovation programs) in 2015 and 2016, as funding for HASS, once the EU subsidies wane.

By 2017 more national investment was required to sustain digital research infrastructure for a much wider array of research domains with funding from both NWO and the Royal Netherlands Academy of Arts and Sciences (KNAW). DANS played a key role in drawing in and managing EU research infrastructure funds for HASS research and common/interoperability research and data infrastructures (nationally, regionally, and internationally). The expenditure for the Netherlands from 2012-2024 on HASS research infrastructure equates to: 73.1M AUD (42.3M CLARIAH + 30.8M ODISSEI) (Table 3).

**Table 3: The Netherlands investment in HASS research infrastructure 2012-2024**

CLARIAH – humanities and arts	ODISSEI – social sciences	DANS – humanities and social sciences support
<b>Total: 2012-2024 (12 years) @42.3M AUD</b>	Total: 2016-2024 (8 years) @30.8M AUD	Total: 2012-2019 (8 years) @60M AUD
<b>2012-2014, 1M Euro (1.6M AUD)</b>	2016, 0.5M Euro (0.8M AUD)	2012 ongoing ~4.5M pa, 36M Euro (60M AUD)
<b>2015-2018, 12M Euro (17.8M AUD)</b>	2019, 18M Euro (30M AUD)	[2012-2015 ~4.5M Euro pa, 18M Euro, 2017 5.2M Euro pa, 2018 4.9M Euro pa]*
<b>2019-2024, 13.8M Euro (22.9M AUD)</b>		

\* See KNAW annual reports for DANS annual funding: <https://www.nidi.knaw.nl/en/about/documenst>

### 2.3.1. Humanities & Arts/ Heritage

The Netherlands plays an important regional role in Europe as the lead for CLARIAH. The Netherlands has developed innovative partnerships and governance models (national and global nodes) to facilitate a research infrastructure ecosystem which achieves regional impact through a combination of institutional centres of excellence, government linkages, and cross-cutting investments.

The key infrastructures supported in the Netherlands are CLARIN, DARIAH, CLARIAH, Europeana, Time Machines, DASISH (Data Service Infrastructure for the Social Sciences and Humanities – a cluster project that brings together all five ESFRI research infrastructures), the National Library of the Netherlands’ KB Lab, and the eScience Center (clustered in association with CLARIAH-PLUS).

The National Library of the Netherlands, Koninklijke Bibliotheek (KB), participates in Europeana projects, and has developed services specifically for digital humanities research. CLARIAH provides datalegend, a structured data service. Time Machines in Amsterdam and Utrecht are exemplars of institutional centres of competence and national networks required to enable humanities and arts research based on common data infrastructure, i.e. linked open data, technologies, and expertise, digitised historical heritage materials (2D and 3D). Local Time Machines also operate as part of a European wide Time Machine initiative.

National networks include DEN (Dutch national knowledge institute for culture and digitalisation and a national digital heritage strategy) and CLARIAH.

CLARIAH centres in the Netherlands are: DANS (Data Archiving and Networked Services), International Institute of Social History, Dutch Language Institute, National Library of the Netherlands, Huygens ING, Meertens Institute, Max-Planck Institute, Netherlands Institute for Sound and Vision. The CLARIAH network includes Austria and Germany (where the consolidation of infrastructures has also occurred) and connected into a national eHumanities platform.

eHumanities.nl is the national platform and has partnerships with research platforms, universities, cultural institutions around expertise, research, technology, and collections.

The Netherlands features cross-sector and cross-institutional partnerships via the Centre for Digital Humanities – a partnership between Amsterdam and VU University, and the Royal Netherlands Academy of Arts and Sciences (KNAW), with support from the Netherlands eScience Center.

A major success factor for the Netherlands is the coordination and integration of research data and technologies with knowledge infrastructures across higher education and heritage to support humanities and arts research. This involves national data centres e.g. National Library and Netherlands Institute for Sound and Vision as a critical strategic partnership along with centres of competence in higher education e.g. Max Planck Institute for Psycholinguistics, Meertens Institute, Institute for Dutch Lexicology, Data Archiving and Networked Services (DANS) and Huygens ING.

The National Library of the Netherlands is a partner in Impact Centre of Competence on digitisation. As part of CLARIAH-CORE proposal, major cultural heritage institutions will provide their collections as digitised resources and are referred to as “national data centres”.

### 2.3.2. Social Sciences

Over time there has been cross-sector stakeholder investment around establishing large scale infrastructure building upon existing institution and national capacity in the survey and collection of social science data, and as a national node for global collaboration. Building on institutional competencies and long-term commitments in pan-national and global initiatives in social surveys and data archiving.

The collaboration around surveys commenced very early on within Europe and in global spheres (2002). Stakeholder relationships at the national level were formed through collaborative survey data collection (longitudinal) and then in the establishment of national data infrastructure and more recently the large-scale platform (ODISSEI).

ODISSEI builds upon a pre-existing consortium of multiple research organisations from across the university and public sector. The major aim of that consortium (and ODISSEI) is to increase infrastructure system efficiencies (reduce fragmentation and overlap) and coordinate effort. Australian equivalents of this infrastructure in universities are the Australian Data Archive (ADA), based at the Australian National University, and longitudinal surveys and datasets such as Household, Income and Labour Dynamics in Australia (HILDA) Survey. Public federal research agency equivalents would be the Australian Bureau of Statistics (ABS), Australian Institute for Health and Welfare (AIHW), and departments of Social Services and Education.

A shared aim in ODISSEI that operates across jurisdictions in the Netherlands is to draw the efforts of research more closely to feeding into social policy directions and outcomes. From a practices perspective the coordination of effort is designed to leverage more effectively the data and expertise in that community and share common infrastructure such as surveys and panels.

## 2.4 Findings

### 2.4.1. Investment approaches and timeframes

There are broadly two investment approaches internationally and evidence of a balance between:

1. Augmentation: building upon existing capacity and research practices, in which research domains and capability levels have matured to the point where they drive the need for collaborative infrastructure; and
2. Transformation: building new capacity and establishing new research practices -- more visionary, future-focused infrastructures which are achieved through strategic investments.

HASS national research infrastructures mapped for this project show staged investment, particularly in Europe, over the last two decades. In the main those changes have been widespread and small to medium scale investments in data infrastructure and research technologies, for example European Social Survey (ESS) (established in 2002), Survey of Health, Ageing and Retirement in Europe (SHARE) (2004), and Common Language Resources and Technology Infrastructure (CLARIN) (2004).

These investments, dating back 15 to 20 years, build on existing research practices, peer networks, investments in research and infrastructures, and capabilities, and are heavily oriented to improving existing data assets and processes. Complementary and relevant research technologies, information and expertise, to maintain and extend research skills are provided, and increase researchers' capacity to take up new research techniques.

In the last ten years, international HASS research infrastructures have focused on equipping HASS researchers with the tools and techniques to take on big data challenges.

In the social sciences, there have been major investments to align data archives, survey instruments and methods. This is evident in the cooperative relationships between CESSDA with other infrastructure projects e.g. SSHOC and SERISS and other social science-related infrastructures e.g. ESS and SHARE.

Humanities and arts investments have focused on experimentation, building capacity, and platforms and digitisation technologies for cultural heritage. In Europe, there are new large-scale text and audio-visual data collections curated and co-located with cloud and high-performance computing at a national level. These larger scale infrastructures operate on two dimensions: platforms for use by many in the cloud, and platforms connected to peak facilities for highly specialised research. They are oriented to enabling novel exploitation of new, curated, diverse and large data assets, utilising cloud and peak computing facility specialists and technologies.

There is evidence of 'top down' and 'bottom up' approaches to development. A range of infrastructure stimuli have enabled research fields to build capability via targeted calls, such as through successive European Framework programs. Europe, Canada, and the Netherlands have engaged in strategic roadmapping exercises, which have scoped HASS capabilities. The UK is currently undergoing the first strategic national roadmapping of its own, which includes a map for the 'social sciences, arts and humanities' sector.

### 2.4.2. Organisational structure

Very few research infrastructures examined through this mapping exercise operate as a shared facility with a centralising role. All infrastructures are of necessity operating in multi-stakeholder networks and they are supported by institutional centres of competence as key partners that provide support, and in most cases critical enabling expertise and infrastructure.

The research infrastructures included in the mapping survey are mostly distributed ‘expert networks’ composed of institutional competence and institutional infrastructures, coordinated by an entity that represents their collective, often national or European, interests. HASS infrastructures therefore tend to place a greater emphasis on community and capability building. To that extent, they can be said to be distinct from research infrastructures serving the Science, Technology, Engineering and Mathematics (STEM) sector tend to be more single-sited, compared with those in the HASS sector.

### 2.4.3. Partnerships

For new types of HASS research to be enabled, data needs to be unlocked, coordinated, and curated at much larger scales in new data sourcing arrangements and co-developed infrastructure partnerships with government organisations or industry.

In Europe, critical stakeholder relationships around researchers’ access to data have been effectively negotiated over time. For example, the Europeana and IMPACT cultural data infrastructure programs operate in a parallel stream to that of European research infrastructure.

Recent funding partnerships and investments of interest to this exercise are:

- > DARIAH and CLARIN programs spent two years (2016-2017) liaising with LIBER (the European Research Libraries) to establish a memorandum of understanding. It is on that basis, that the partners and members that constitute DARIAH assert: “We bring the world of cultural data with us” to the European Open Science Cloud (EOSC) initiative.
- > DARIAH and CLARIN partners and members, and collaborators have a place within the EOSC funded Social Sciences and Humanities Open Cloud program (SSHOC) – and – on the E-RIHS (an emerging ERIC).
- > CESSDA has MoU in place with other ERIC social science research infrastructures, and in 2017 tackled a “strengthening and widening” activity and reported on the state-of-the-art, obstacles, models and roadmaps for widening the data perimeter of the data services, through planning and engagement (see also their role in the RAIRD project for access to register data). The expected outcome being to “address new data sources and new actors” including statistics agencies and producers of web, transactional, administrative and historical data and to establish “establish agreements with other institutions/organizations and data producers in order to keep up with researchers’ needs, which change over time”.

#### 2.4.4. Trends and emerging priorities

Internationally, new HASS national research infrastructures have been created in the last decade, and this has affected a kind of HASS research ‘disruption’. In the main those investments have been focused on developing entirely new types of infrastructure cloud computing and platforms and large-scale types of research infrastructure.

The new HASS research infrastructures enable entirely new research practices to be fostered, and they require new peer networks to be created in the infrastructure design and research pilots (informing those designs). New investments are directed toward constructing new data infrastructure, i.e. new types of data are drawn together (administrative, social media, and digital heritage) with cloud and high-performance computing infrastructures.

These infrastructures are oriented to enabling novel exploitation from new curated, diverse and large data assets, utilising cloud and peak computing facility specialists and technologies, the skills and research interests of data, information and computer scientists.

These types of HASS research infrastructures afford the capacity to tackle research, that currently it is not possible to undertake without that scale of infrastructure or investment. Key success factors in their development is balancing the HASS research requirements driving the design, and the data/information/computer scientists, operating as both infrastructure designers and research enabling partners.

The emergence of new and unique humanities research infrastructures is evident in the Time Machine initiatives in Europe (below), the ‘Living With Machines’ initiative in the UK, the HathiTrust in the USA, and global collaboration around digitised cultural heritage materials maintained by large national research and heritage libraries e.g. Global Dataset of Digitised Texts Network (GDD Network) (Arts and Humanities Research Council (AHRC), 2019).

In the case of the GDD Network, the fact that the UK has a separate research council for arts and humanities research has enabled strategic investment in domain-specific infrastructure. The GDD Network is funded by the AHRC, led by the University of Glasgow, in close collaboration with the HathiTrust, based in the USA, and involves key library partners as follows: the British Library, National Library of Scotland, National Library of Wales, and Research Libraries UK (RLUK).

The AHRC is also leading the development of the ‘Living with Machines’ project, which is a collaboration between the Alan Turing Institute and the British Library, together with researchers from a range of universities. Funding of £9.2 million has been awarded from the UK Research and Innovation (UKRI) Strategic Priorities Fund for the infrastructure initiative which “will take place over five years and is set to be one of the biggest and most ambitious humanities and science research initiatives ever to launch in the UK” (Alan Turing Institute, 2019).

In Europe, a pattern is emerging of new large-scale text and audio-visual data collections curated and co-located with cloud and high-performance computing at a national level. e.g. Denmark’s Cultural Heritage Cluster – a collaboration between DeIC (Danish e-Infrastructure Cooperation) and the Royal Danish Library. These larger scale infrastructures operate on two levels: platforms for use by many in the cloud, and platforms connected to

peak facility for highly specialised research. Australia is missing this national domain focused capability to support HASS research.

## 2.5 Lessons

International mapping tells a story of what can be achieved through national research infrastructure stimulus, indicating where Australia might seek to ‘leapfrog’ and what pitfalls to avoid.

Key lessons for Australia are in Table 4, which also includes potential priorities and identifies comparative Australian infrastructures (at project, program or institutional level). Overall, the following patterns have emerged from the research and are significant for Australia:

1. **Laying the foundations.** There is evidence of staged investment and pathways for the development of infrastructure over time. Infrastructures have transitioned from program funding to ‘landmark’ status. For example, DARIAH, which was funded in 2006 through the European Strategy Framework for Research Infrastructure (ESFRI) and reached European Research Infrastructure Consortium (ERIC) status in 2014. In the social sciences, longitudinal survey infrastructure has developed over decades, and is now undergoing further transformation through possibilities offered by technology, compute, and data linkage.
2. **Capacity-building and community-building** is a fundamental part of the process. For example, DARIAH’s national approach to boosting support for arts and humanities in a regional network (member and partner organisations) and strong programs for community building and coordination (ground up first).
3. In Europe, **infrastructure stimulus programs** (Framework Program funding have seeded infrastructure build and are now focusing (via Horizon 2020) on cross-cutting initiatives designed to accelerate collaboration and common infrastructure at scale. For example, the cross-cutting program Synergies for Europe’s Research Infrastructures in the Social Sciences (SERISS), which brings together ESS, SHARE, CESSDA, among others to advance interoperability, comparative research and support for big data research.
4. **Fragmentation in infrastructure investment** has been identified as an area to address in humanities, arts and heritage. Two Horizon 2020 programs have been funded to address the need for stronger coordination: PARTHENOS (cross-cutting to improve data practices across programs); and E-RIHS-PP (ESFRI-funded).
5. **Balancing domain-specific and opportunities for common infrastructure development.** In Europe domain-specific development is giving way to a trend for consolidation and common infrastructure development. The alignment of CLARIN and DARIAH as CLARIAH is one example of this trend, as is the new Social Science and Humanities Open Cloud (SSHOC), which is part of the general shift in data and technology intensive research to using cloud infrastructure and working in virtual environments. There are indications in the new SSHOC that while some of the challenges will be solved by common approaches, there is still a need for specialised, domain-specific infrastructure build. The lesson for Australia is that the planning for this should happen at the outset of NRI development. There is an opportunity to ensure the HASS, Indigenous and Heritage domains can be brought together as part of

the planning for a comprehensive assessment of common needs and to identify where specialist infrastructure will be required.

This type of infrastructure development work could align with the strategic policy agenda in Australia (through the Australian Research Data Commons) for stronger research data management for data enabled and driven research, improvements in the production and reuse of quality research data and software assets, and the development of domain oriented virtual environments, and cloud enabled infrastructures.

6. **Developing complex governance models and partnerships to align the needs of different sectors.** Across Europe there is a strong trend towards consolidation across the HASS research infrastructures, leading to increasingly sophisticated governance and co-investment models. Significant effort has been invested in EU legislation to address data privacy and data rights. There has been a focus also on aligning culture/heritage and research infrastructure. The IMPACT Centre of Competence is an exemplar of aligning investments in digitisation to meet the needs of European culture and society and to meet the needs of research (access to heritage data), and also the capacity to leverage collaborative networks and consolidate investments (for both sectors and institutional members).
7. **National strategies also recognise the critical importance of participating in, and contributing to, international and regional research infrastructures.** Here the Netherlands is particularly instructive. The CLARIAH network, includes Austria and Germany (where the consolidation of infrastructures has also occurred) and connected into a national eHumanities platform. The Netherlands Organisation for Scientific Research (NWO) collaborates with CESSDA, ESS, SHARE as part of the internationalisation of social sciences.

**Table 4: Key lessons for Australia from the International mapping**

Research Infrastructure	Lessons	Trends and priorities for Australia
European Social Survey (ESS)	<p>At a national level through the participating countries, ESS operates as a pan-national data discovery service. Targeted at survey methodology and interoperability of data infrastructure to support comparative analysis (closer to the research process).</p> <p>The impact of different funding levels across time as the mode appears to be €1.5M but ramped up in design phases (€4.4M &amp; €6.4M). This may be an investment trend across ESFRI projects (front loaded during these phases). Investments in cross-cutting program SERISS (ESS, SHARE, CESSDA, GGP, EVS) are designed to advance interoperability, comparative research and support for big data research.</p>	<p>Informing potential Australian priorities: <b>coordinating entity</b> and <b>data hub</b> to provide national focus for identifying and facilitating access to government datasets for research.</p> <p>Australian comparators include: Longitudinal datasets e.g. HILDA, LSAC, LSIC, BNLA (via the Department of Social Services)</p>
Survey of Health, Ageing and Retirement in Europe (SHARE)	<p>Pan-national cooperation around survey methods and data collection of ageing populations in Europe. Notable growth in partners participating in surveys from 2011 onward and a jump in 2017.</p> <p>Support for cross-cutting infrastructure - SERISS (and alignment with work undertaken in ESS and CESSDA) and DASISH (another cross-cutting infrastructure). Cross-cutting program SERISS (ESS, SHARE, CESSDA, GGP, EVS) to advance interoperability, comparative research and support for big data research.</p>	<p>Informing potential Australian priorities: <b>coordinating entity</b> and <b>data hub</b> to provide national focus for identifying and facilitating access to government datasets for research.</p> <p>Australian comparators include: Longitudinal datasets e.g. ALSA (Australian Longitudinal Study of Ageing)</p>
Common Language Resources and Technology Infrastructure (CLARIN)	<p>Enables advanced analytics and new types of services and technologies. Links between heritage and language infrastructures, and new industries.</p> <p>Digitisation is a 'knowledge centre' in the CLARIN network – MoU with LIBER on digital collections (with DARIAH).</p> <p>Has not maximised user uptake and community engagement in design of data infrastructure.</p>	<p>Informing potential Australian priorities: <b>coordinating entity, data hub, digitisation capability, digital HASS peak capability.</b></p> <p>Australian comparators include:</p> <p>Alveo virtual laboratory, Australian National Corpus, PARADISEC, Centre of Excellence for the Dynamics of Language, AIATSIS Collections.</p> <p>Updates to Trove recently with AUSTLANG vocabulary (partnership with AIATSIS).</p>

	<p>Centre certification process to identify useful pathways for infrastructure maturity, investment, growth and scaling up and cost-benefit for participation in CLARIN network.</p> <p>Move to blend CLARIN/DARIAH in Germany as per Netherlands as a CLARIAH.</p>	
Digital Research Infrastructure for the Arts and Humanities (DARIAH)	<p>National approach to boosting support for arts and humanities in regional network (member and partner organisations). Community building and coordination (ground up first).</p> <p>Longer time to get DARIAH operational (may indicate a need to build community capability and capacity all in parallel). By comparison language institutes established earlier and CLARIN became operational quickly (with potentially more data and technology capability in the research system).</p> <p>Strong collaborations built with digitisation.eu and MoU arrangements with LIBER (Association of European Research Libraries).</p>	<p>Informing potential Australian priorities: <b>coordinating entity, digitisation capability, digital HASS peak capability.</b></p> <p>Australian comparators include: Institutional infrastructures arising from LIEF investment e.g. AusStage, AustLit, AustLII, Design and Art Australia Online.</p>
Consortium of European Social Science Data Archives (CESSDA)	<p>Early and tight coupling with national statistics bodies and university data archives. Comprehensive coverage of research data and tool requirements over time.</p> <p>Consortia approach for longitudinal, demographic data and data linkage of benefit for government, higher education and industry. Collaboration around discovery of social science data with government, e.g. Denmark cross-ministry support (culture and higher education and research) for data discoverability and access. Early linkages with national statistics bodies and university data archives.</p> <p>CESSDA 2018-2022 strategy indicates a push to global partnerships and to find "third parties" e.g. with ICSPR (USA) and Research Data Alliance. Emphasis on the breadth and depth (50 years) of experience. Cross over of CESSDA with other infrastructure projects e.g. SSHOC &amp; SERISS; Collaboration within ERICs across domain boundaries e.g. Social Sciences and Humanities cluster; Bundling of CESSDA, ESS, SHARE</p>	<p>Informing potential Australian priorities: <b>coordinating entity and data hub</b> to provide national focus for identifying and facilitating access to government datasets for research.</p> <p>Australian comparators include:</p> <p>Australian Data Archive (an institutional service, that delivers national services).</p>

	<p>together as PROGEDO at national level; Cross-cutting program SERISS (ESS, SHARE, CESSDA, GGP, EVS) to advance interoperability, comparative research and support for big data research.</p>	
SAIL Databank	<p>Secure access to demographic, health, social and education data from the resident population of Wales. Data Linkage services. Supports medical and health research primarily. Could be re-usable for HASS research. Identifies Canada (Ministry of Health and Long-Term Care) and Australia (PHRN) as global equivalents for data linkage.</p> <p>Has a successful trajectory of combined infrastructure services (data linkage, a safe haven databank, tools) and national networks to support local, national and international collaborations. Movement from pilot stage (and design) to establish a shared national facility with formal partnerships with access controls, that serves local and national research infrastructure requirements (whether that approach could be generalised and used to meet the needs of research to access to indigenous heritage and research data in government and higher education collections).</p>	<p>Informing potential Australian priorities: <b>coordinating entity and data hub.</b></p> <p>Australian comparators include:</p> <p>Population Health Research Network (NCRIS).</p> <p>Australian Bureau of Statistics (ABS Datalab remote or onsite).</p> <p>Bioplatforms Australia (NCRIS).</p> <p>Australian Data Archive (ANU)</p> <p>National Centre for Indigenous Genomics (ANU)</p>
IMPACT Centre of Competence	<p>The alignment of resources, expertise and technology (around digitisation of text and language-based heritage resources) and the link to research, and expertise around text and language-based heritage resources (the alignment of the investments in IMPACT - for heritage and in CLARIN - for research).</p> <p>Is a node of CLARIN (as a knowledge centre) also serves as a centre of competence for the heritage community to become a member of and build up a network of members and partners. Notably both heritage and higher education institutions are members of IMPACT to gain the benefits of a centre of competence.</p> <p>Exemplar of alignment of investments in digitisation to meet the needs of European cultural and society and to meet the needs of research (access to heritage data), and the capacity to leverage collaborative</p>	<p>Informing potential Australian priorities: <b>coordinating entity and digitisation capability.</b></p> <p>Australian comparators include:</p> <p>Major cultural institutions (e.g. National Library of Australia, State Libraries), science and research institutions (e.g. CSIRO, University of Sydney and Melbourne) have digitisation facilities</p>

	networks and consolidate investments (for both sectors and institutional members).	
Europeana	<p>World leading in the advancement of cultural heritage interoperability. Aggregates using standards from diverse heritage organisations (and approaches to informatics) and uses linked open data practices to support data enhancement and richer semantic discovery. Community engagement through partnerships with research infrastructure programs e.g. DARIAH and Time Machine.</p> <p>Approaches to digitisation: advantages of EU funding, establishing economies of scale, and standards and quality challenges.</p>	<p>Informing potential Australian priorities: <b>digitisation capability</b> and <b>digital HASS peak capability</b>.</p> <p>Australian comparators include:</p> <p>Trove Australia (provided by the National Library of Australia, with a contributor holdings from 1000 libraries and 300 organisations with heritage collections).</p>
HathiTrust	<p>Text and data mining infrastructure arising through consortia, to meet a common research infrastructure requirement (across domains) for scholarly discovery phase (literature reviews), digital humanities, library and information science and computer science.</p> <p>Capacity for digitisation, text and data mining services to be shared across academic libraries.</p> <p>Established HaithiTrust Research Center to enable computational analysis and help meet the technical challenges researchers face when dealing with massive amounts of digital text.</p>	<p>Informing potential Australian priorities: <b>digitisation capability</b> and <b>digital HASS peak capability</b>.</p> <p>Australian comparators include:</p> <p>Trove Australia (provided by the National Library of Australia)</p> <p>Digitisation partnership.</p>
First Nations Information Governance Centre	<p>Membership comprises First Nation organisations with a focus on health and wellbeing.</p> <p>Provides services across government, higher education, industry, and community.</p> <p>Focused surveys and data capture for a range of health and wellbeing objectives where they impact on First Nations peoples in Canada.</p> <p>Funding at national level for surveys by Health Canada, and Indigenous and Northern Affairs Canada, and Employment and Social Development Canada.</p> <p>The integration of Indigenous knowledge in the research system is part</p>	<p>Informing potential Australian priorities: <b>coordinating entity</b> and <b>Indigenous Data Framework</b>.</p> <p>Australian comparators include:</p> <p>National Aboriginal and Torres Strait Islander Health Survey (NATSIHS), National Aboriginal and Torres Strait Islander Social Survey (NATSISS), National Aboriginal and Torres Strait Islander Nutrition and Physical Activity Survey (NATSINPAS), Australian Bureau of Statistics. Mayi Kuwayu Survey (National Study of Aboriginal and Torres Strait Islander Wellbeing).</p> <p>National Indigenous Languages Survey, Australian Institute of Aboriginal and Torres Strait Islander Studies (AIATSIS).</p>

	of the Canadian science and technology plan (2020-25).	Other organisations involved: National Aboriginal Community Controlled Health Organisation (NACCHO), Australian Institute of Health and Welfare (AIHW), Departments of Social Services (HILDA).
Cohort and Longitudinal Studies Enhancement Resources (CLOSER)	<p>Institutional centres and partnerships with universities, British Library and UK Data Service.</p> <p>Has evolved out of established survey mechanisms and data collection to a centre of competence as a hub in the research network and linking to other key stakeholders outside of higher education. CLOSER is embedded in the university context whereas in Australia, the National Centre for Longitudinal Data (NCLD) operates out of the Department of Social Services, this shifts the way the infrastructure is coordinated, and data can be made accessible for research</p>	<p>Informing potential Australian priorities: <b>coordinating entity</b> and <b>data hub</b> to provide national focus for identifying and facilitating access to government datasets for research.</p> <p>Australian comparators include:</p> <p>National Centre for Longitudinal Data (NCLD), Department of Social Services.</p> <p>Australian Data Archive.</p> <p>Australian Bureau of Statistics.</p>
Te Mana Raraunga - Māori Data Sovereignty Network	<p>Ensures data for and about Maori can be safeguarded and protected. Advocates for Maori involvement in the governance of data repositories. Supports the development of Maori data infrastructure.</p> <p>Strategic digitisation agenda (in relation to Māori) is led by the National Library of New Zealand and Archives New Zealand, in partnership with the Ministry for Culture and Heritage (Internal Affairs).</p> <p>Issues in common with Australia is the lack of Indigenous input and oversight into the use of administrative data and integration of cultural and privacy concerns.</p>	<p>Informing potential Australian priorities: <b>coordinating entity</b> and <b>Indigenous Data Framework</b>.</p> <p>Australian comparators include:</p> <p>Maiam nayri Wingara (Aboriginal and Torres Strait Islander Data Sovereignty Collective), Indigenous Data Network (University of Melbourne)</p>
Compute Canada	<p>Specific support for humanities and social sciences built into existing national computational services.</p> <p>Progress on national HPC research infrastructure with publicly funded research organisations e.g. Statistics Canada. Working relationship between Compute Canada and CARL (Canadian Association of Research Libraries).</p> <p>Mix of community building (INKE), HASS projects needing cloud based research infrastructure versus HPC, and capability building (DHSI) as an</p>	<p>Informing potential Australian priorities: <b>coordinating entity</b> for HASS to enable access and support to existing NCRIS facilities.</p> <p>Australian comparators include:</p> <p>Not yet an Australian equivalent (through e.g. NCI or Pawsey).</p>

---

indicators of growth in research uptake. Gap between higher education and government in the enabling of advanced research and the provision of national research infrastructure.

---

European Research Infrastructure for Heritage Science (E-RIHS)

In pre-operational phase. Designed for trans-disciplinary and interdisciplinary use.

Aims to deliver integrated access to expertise, data and technologies through four platforms to support heritage science: ARCHLAB (physical collections), DIGILAB (digital tools and FAIR data), FIXLAB (material science tools), MOBLAB (mobile digital tools).

Problem addressing: A need to address data infrastructure and the enabling language and image processing technologies and techniques (standardisation) across the range of disciplines in humanities and arts (and the range of heritage collection types).

Fragmentation in infrastructure investment has been identified as an area to address in humanities, arts and heritage. Two H2020 programs: PARTHENOS (cross-cutting to improve data practices across programs); and E-RIHS-PP (ESFRI), reference the need for stronger coordination. Both infrastructures combine training, virtual environments, standardisation, community building; leveraging of a range of data, technologies, and techniques. PARTHENOS appears to focus on FAIR data and standardisation of research workflows for humanities research, as an improvement exercise across infrastructures, whereas E-RIHS is focused on all disciplinary dimensions of heritage sciences and global leadership.

---

Informing potential Australian priorities: **digitisation capability** and **digital HASS peak capability**.

Australian comparators include:

NCRIS Synchrotron (Melbourne) and National Imaging Facility (Brisbane). ARC funded scanning equipment: AustLii and Australian Policy Observatory). Facilities in major cultural heritage institutions (e.g. National Library digitisation) and equipment in universities (e.g. ANU CT Lab and University of Melbourne Digitisation Centre) and scientific organisations with heritage collections (e.g. CSIRO and Geoscience Australia).

## 2.6 Opportunities for collaboration

Opportunities exist for information and expertise sharing, competency benchmarking, and formal linkages with international infrastructure. Potential opportunities identified through this project are listed in the following table alongside Australian comparators (NCRIS, national or institutional).

**Table 5: Opportunities for global collaboration**

<b>AU Comparators</b>	<b>Potential Global Partner</b>	<b>Opportunity</b>
Alveo virtual laboratory, Australian National Corpus, PARADISEC, Centre of Excellence for the Dynamics of Language. AIATSIS Collections.	Common Language Resources and Technology Infrastructure (CLARIN)	Affiliation: "third party" status to connect into global research linkages with Europe, USA and South Africa.
Institutional infrastructures arising from LIEF investment e.g. AusStage, AustLit, AustLII, Design and Art Australia Online.	Digital Research Infrastructure for the Arts and Humanities (DARIAH)	Affiliation: "third party" status to connect into global research linkages with Europe and USA.
Australian Data Archive (an institutional service, that delivers national services) .	Consortium of European Social Science Data Archives (CESSDA)	Affiliation: participate in a global community of practice. Leverage specialised sensitive data technologies.
Longitudinal datasets e.g. HILDA, LSAC, LSIC, BNLA (via the Department of Social Services)	European Social Survey (ESS)	Affiliation: connect into global research linkages with Asia, South Africa, USA, Latin America.
As above: longitudinal datasets.	Survey of Health, Ageing and Retirement in Europe (SHARE)	Affiliation: connect into global research linkages with Asia, USA, Latin America, and Israel.
Major cultural institutions (e.g. National Library of Australia, State Libraries), science and research institutions (e.g. CSIRO, University of Sydney and Melbourne) have digitisation facilities.	IMPACT Centre of Competence	Membership: participate in community of practice. Leverage specialised language technologies and corpora or lexicon.
NCRIS Synchrotron (Melbourne) and National Imaging Facility (Brisbane). ARC funded scanning equipment: AustLii and Australian Policy Observatory). Facilities in major cultural heritage institutions (e.g. National Library digitisation) and equipment in universities (e.g. ANU CT Lab and University of Melbourne Digitisation Centre) and scientific organisations with heritage collections (e.g. CSIRO and Geoscience Australia).	European Research Infrastructure for Heritage Science (E-RIHS)	Partnership: participate in community of practice and connect into global research linkages with Europe, Israel, USA, Latin America.

Trove Australia (provided by the National Library of Australia, with a contributor holdings from 1000 libraries and 300 organisations with heritage collections).	Europeana	Partnership: leverage existing global community of practice with Europe, USA and New Zealand.
Maiam nayri Wingara (Aboriginal and Torres Strait Islander Data Sovereignty Collective), Indigenous Data Network (University of Melbourne)	Te Mana Raraunga - Māori Data Sovereignty Network	Leadership: establish global community of practice and connect into research linkages with Canada and New Zealand.
National Aboriginal and Torres Strait Islander Health Survey (NATSIHS), National Aboriginal and Torres Strait Islander Social Survey (NATSISS), National Aboriginal and Torres Strait Islander Nutrition and Physical Activity Survey (NATSINPAS), Australian Bureau of Statistics. Mayi Kuwayu Survey (National Study of Aboriginal and Torres Strait Islander Wellbeing).  National Indigenous Languages Survey, Australian Institute of Aboriginal and Torres Strait Islander Studies (AIATSIS).  Other organisations involved: National Aboriginal Community Controlled Health Organisation (NACCHO), Australian Institute of Health and Welfare (AIHW), Departments of Social Services (HILDA).	First Nations Information Governance Centre	Leadership: establish global community of practice and connect into research linkages with Canada and New Zealand.
Population Health Research Network (NCRIS).  Australian Bureau of Statistics (ABS Datalab remote or onsite).  Bioplatforms Australia (NCRIS).  Australian Data Archive (ANU)  National Centre for Indigenous Genomics (ANU)	SAIL Databank	Partnership: broaden out engagement around data linkage community of practice and connect into research linkages in UK, Europe, Canada beyond population health and into social sciences.
National Centre for Longitudinal Data (NCLD), Department of Social Services.  Australian Data Archive.  Australian Bureau of Statistics.	Cohort and Longitudinal Studies Enhancement Resources (CLOSER)	Affiliation: leverage existing global community of practice and connect into research linkages in UK and Europe.
Trove digitisation partnership (National Library of Australia).	HathiTrust	Affiliation: leverage existing global community of practice in USA.

In the social sciences, it is worth exploring in more detail CESSDA, ESS and SHARE; and for language-oriented humanities research, CLARIN. These three EU programs have strengths in data infrastructure and expertise, communities of practice, and their technologies are potentially portable into the Australian setting. The longitudinal surveying infrastructures i.e. SHARE and ESS also may have portable technologies.

Regarding wider humanities and arts research DARIAH is an obvious candidate, based on the strength of their communities of practice and their role the emerging ERIC E-RIHS. In addition to this, CESSDA, ESS, SHARE, DARIAH and CLARIN are all participants in the Social Sciences and Humanities Open Cloud (SSHOC) development (part of the European Open Science Cloud).

## 2.7 Models for further analysis for HASS scoping

In terms of models for further analysis, potentially as part of the scoping exercise to aid with addressing gaps in Australian national research infrastructure, it would be fruitful to investigate seven infrastructures further (Table 6): IMPACT, Europeana, SAIL, CLOSER, HathiTrust, CLARIAH and ODESSEI to understand the kind of research being enabled with a range of infrastructure types and capabilities. IMPACT and Europeana represent decentralised data collections, and SAIL, CLOSER, and HathiTrust represent centralised data collections. CLARIAH and ODESSEI build on pre-existing research infrastructure and consortia.

**Table 6: Models for further analysis**

Infrastructure	Trends	Further Analysis Areas
<b>IMPACT</b>	<p>Specialisation of digitisation technologies based on heritage object types and contexts, i.e. IMPACT (text) versus 3D (objects) and heritage institution type.</p> <p>An alignment of investments in digitisation to meet the needs of society and research (access to heritage data), and capacity to leverage collaborative networks and consolidate investments.</p> <p>An alignment of heritage and research resources, expertise and technology – around text and language-based heritage resources, IMPACT and CLARIN.</p>	<p>Connection between Europeana (as a pan-national broker) and other leading heritage institutions in orientation to research for expertise and to delivery heritage data (for research reuse).</p>
<b>Europeana</b>	<p>Co-investment compared to EU contribution has three modes: none, 1/4 or 1/2 of the funds co-contributed to the budget. This may indicate investment for national outcomes. Where the co-investment is negligible this may indicate cross-cutting interests with a pan-national outcome.</p>	<p>Examine Framework funding for humanities and arts against the funding for Europeana to ascertain any crossover. Does this impact the availability of digital material (supply) for research reuse?.</p>
<b>Sail Databank</b>	<p>Clustered infrastructures data, expertise, tools, compute operates as a highly effective means to attract research grants and establish local, national and global networks.</p> <p>Building diverse data processing techniques into platform and suite of services to support medical and health research. Are the tools and techniques to extract domain specific data from unstructured text and characterisation tools and techniques for image</p>	<p>Movement from pilot stage to establish a shared national facility with formal partnerships with access controls, that serves local and national research infrastructure requirements. Can this approach be generalised and used to meet the needs of research to access to Indigenous</p>

	<p>data generalisable and reusable for HASS research?</p> <p>A successful trajectory of combined infrastructure services (data linkage, a safe haven databank, tools) and national networks to support local, national and international collaborations.</p>	<p>heritage and research data in government and higher education collections?</p>
<b>CLOSER</b>	<p>Evolution of research infrastructure out of establishing survey mechanisms and data collection to a centre of competence as a hub in the research network and linking to key stakeholders outside of higher education.</p> <p>CLOSER is embedded in the university context whereas National Centre for Longitudinal Data (NCLD) operates out of the Department of Social Services, this shifts the way the infrastructure is coordinated, and how data can be made accessible for research.</p>	<p>How CLOSER liaises with the Australian DSS and ABS equivalents in the UK, to ascertain what network links are established and collaborations enabled.</p>
<b>HathiTrust</b>	<p>Text and data mining infrastructure arising through consortia, to meet a common research infrastructure requirement (across domains) for scholarly discovery phase (literature reviews), digital humanities, library and information science and computer science.</p> <p>Capacity for digitisation, text and data mining services to be shared across academic libraries, e.g. CADRE, California Digital Library.</p> <p>HathiTrust Research Center Advanced Collaborative Support Program is used as a means of addressing computational analysis challenges of the corpus.</p>	<p>What role has computer science played in large complex unstructured and structured data architectures for humanities?</p> <p>What research needs drove the delivery of digital library material as datasets for research. What factors relating to industry partnership and copyright have constrained data accessibility (for research).</p>
<b>CLARIAH</b>	<p>CLARIAH consolidates existing infrastructure around language based research in HASS in the Netherlands and leverages research technologies already developed. Provides coordination platform and expertise in common infrastructure development, interoperability, as well as domain informatics. CLARIAH community of practice has broadened to include more universities and heritage and public institutions, including parliament and university libraries as partners.</p>	<p>Australia's opportunity to achieve the right clustering, cooperation and coordination from the outset of HASS NRI planning. Achieving interoperability with international infrastructure to link Australian-based records with the world is major area for further investigation</p>
<b>ODISSEI</b>	<p>ODISSEI builds upon a preexisting consortium of multiple research organisations from across the university and public sector. The major aim of that consortium (and ODESSEI) is to increase infrastructure system efficiencies (reduce fragmentation and overlap) and coordinate effort.</p>	<p>Achieving sustainability for social sciences infrastructure. How the 'open data infrastructure' model has enabled researchers to answer new, cross-disciplinary research questions or investigate existing questions in new ways.</p>

## 2.8 Conclusion

In Australia there is an opportunity to align the development of HASS NRI with the 2016 Roadmap and with key policy agendas for digital infrastructure, data sharing and citizen/consumer data rights.

International investments in HASS research infrastructure have shown pathways for infrastructure investment in:

- > heritage research, data and technologies – building capabilities in characterisation technology and pattern recognition;
- > language research, data and technologies – building capabilities in informatics, semantics and AI; and
- > social research, data and technologies – building capabilities in internet of things, civic technologies and precision services.

In general, the ‘use-value’ for Australia from these programs are in the following areas:

- > Identifying known ‘good’ (and known ‘bad’) approaches to establishing research infrastructures in HASS, which can be integrated with the existing Australia programs of collaborative infrastructure (e.g. Australian Research Data Commons – ARDC) and evolved from institutional capacity (e.g. Australian Data Archive – ADA).
- > Gaining access to technical expertise and frameworks that benefit from global adoption (e.g. IMPACT for digitisation).
- > Leveraging governance and funding (co-investment) models, some of which are at large scale (e.g. European) which could be applied in an Australian context (e.g. federated states in the case of government data, bring together research and collecting institutions).
- > Avoiding duplication of effort in newer areas through partnering and sharing of best practises, notably in indigenous research infrastructures and digital and data-intensive humanities.
- > Where new types of HASS research are to be enabled, this requires looking closely at where data needs to be unlocked, coordinated, and curated at much larger scales in new data sourcing arrangements and co-developed infrastructure partnerships with government organisations or industry.

## 3 Gap Analysis

---

### 3.1 Introduction

In contrast to the countries and regions surveyed for this study, there are no nationally comparable HASS research infrastructures in Australia, nor a national domain-focused capability to support computationally transformative HASS research. This represents a significant gap in Australia's national research capacity relative to other countries.

In Australia, existing HASS infrastructures are largely operating at an institutional or project-based level, and the state of play may be characterised as uncoordinated. Existing platforms separately work to standardise, harmonise and provide single points of access to data. There is not a combined set of tools to power innovation in the way researchers analyse combined datasets, and support for new research methodologies. Australia does not have a coordinated HASS NRI system which can connect data hubs, facilities and the wider research enabling ecosystem, including NCRIS funded entities and centres of excellence.

No other nations examined for this project explicitly highlight research into Indigenous health, wellbeing and culture as a national priority or key enabling feature of their large-scale research infrastructure. This is a gap in the international landscape that Australia could seek to lead and fill.

An opportunity exists for Australia to take a significant regional leadership or partnership role in a potential Indo-Pacific wide approach to research infrastructure.

### 3.2 The Australian landscape

Patterns in the Australian landscape look like the pre-stimulus phase in Europe in which Framework Program funding seeded key HASS strengths. Australia has made some investments via ANDS/NeCTAR/RDS and through the Australian Research Council's Linkage Infrastructure, Equipment, and Facilities scheme into institutional infrastructure and is in a good position to take advantage of strategic NRI investment.

Over a twenty-year period and particularly over the last decade, national systems within Europe have achieved transformations in HASS research, generating new skills, industries, and technology. In a country the size of Australia, with a federated system, we can take lessons from European investments in particular, which have now moved into a period of consolidation, at domain level (e.g. the collaborative agenda for CLARIN and DARIAH – (CLARIAH) and with underpinning infrastructures – the Social Sciences and Humanities Open Cloud (SSHOC) development (which is part of the European Open Science Cloud).

The project mapped select research projects and programs in Australia against international exemplars to identify areas of alignment and potential opportunity or collaboration (Table 7).

**Table 7: A closer look at Australian comparators**

<b>Infrastructure Name</b>	<b>Associated HASS Projects</b>	<b>AU Comparators</b>
<b>Consortium of European Social Science Data Archives (CESSDA)</b>	See UK Data Archive <a href="#">Data Impact Blog</a> and <a href="#">Impact and Innovation Lab</a> .	Australian Data Archive (an institutional service, that delivers national services)
<b>European Social Survey (ESS)</b>	See <a href="#">bibliography of publications based on ESS data</a> .	Longitudinal datasets e.g. HILDA, LSAC, LSIC, BNLA (via the Department of Social Services)
<b>Survey of Health, Ageing and Retirement in Europe (SHARE)</b>	See <a href="#">publications based on SHARE data</a> .	Longitudinal datasets e.g. ALSA (Australian Longitudinal Study of Ageing).
<b>Compute Canada</b>	<a href="#">Canadian Writing Research Collaboratory</a> <a href="#">Web Archives for Longitudinal Knowledge</a>	--
<b>IMPACT Centre of Competence (digitisation)</b>	See <a href="#">CLARIN Knowledge Centre for Denmark – publications from the Dept of Nordic Studies and Linguistics (NorS)</a>	Major cultural institutions (e.g. National Library of Australia, State Libraries), science and research institutions (e.g. CSIRO, University of Sydney and Melbourne) have digitisation facilities. None specialise in language technologies or provide corpora or lexicon.
<b>E-RIHS</b>	See <a href="#">scientific publications from IPERION-CH (precursor to E-RIHS)</a> .	NCRIS Synchrotron (Melbourne) and National Imaging Facility (Brisbane). ARC funded scanning equipment: AustLii and Australian Policy Observatory). Facilities in major cultural heritage institutions (e.g. National Library digitisation) and equipment in universities (e.g. ANU CT Lab and University of Melbourne Digitisation Centre) and scientific organisations with heritage collections (e.g. CSIRO and Geoscience Australia).
<b>Europeana</b>	<a href="#">Venice Time Machine</a> , <a href="#">Naturalis Biodiversity Center – Butterfly Species Identification</a> , <a href="#">Golden Agents</a> , <a href="#">Life of Newspapers</a>	Trove Australia (National Library of Australia, with contributor holdings from 1000 libraries and 300 organisations with heritage collections.
<b>CLARIN</b>	See Danish node for <a href="#">peer reviewed papers</a> and Swedish node for <a href="#">Tilltal</a>	Alveo virtual laboratory, Australian National Corpus, PARADISEC, Centre of Excellence for the Dynamics of Language. AIATSIS Collections.
<b>DARIAH</b>	See CLARIAH (Netherlands node) for <a href="#">pilot research projects</a> .	At an institutional level there are comparators arising from LIEF investment e.g. AusStage, AustLit, AustLII, DAAO, CoEDL.

<b>Te Mana Raraunga</b>	See <a href="#">research activities</a> .	Maiam nayri Wingara (Aboriginal and Torres Strait Islander Data Sovereignty Collective), Indigenous Data Network (University of Melbourne).
<b>First Nations Information Governance Centre</b>	See <a href="#">publications</a> .	<p>National Aboriginal and Torres Strait Islander Health Survey (NATSIHS), National Aboriginal and Torres Strait Islander Social Survey (NATSISS), National Aboriginal and Torres Strait Islander Nutrition and Physical Activity Survey (NATSINPAS), Australian Bureau of Statistics.</p> <p>Mayi Kuwayu Survey (National Study of Aboriginal and Torres Strait Islander Wellbeing, National Indigenous Languages Survey, Australian Institute of Aboriginal and Torres Strait Islander Studies (AIATSIS).</p> <p>Other organisations involved: National Aboriginal Community Controlled Health Organisation (NACCHO), Australian Institute of Health and Welfare (AIHW), Departments of Social Services (HILDA).</p>
<b>SAIL Databank</b>	See <a href="#">uses for SAIL data</a> and <a href="#">projects using SAIL data</a> .	Australia Population Health Research Network (NCRIS), Australian Bureau of Statistics (ABS Datalab remote or onsite), Bioplatforms Australia, Australian Data Archive (ANU), National Centre for Indigenous Genomics (ANU).
<b>CLOSER</b>	See <a href="#">blog for research news</a> .	<p>National Centre for Longitudinal Data (NCLD), Department of Social Services. Longitudinal data is maintained by survey managers for HILDA, LSAY etc. Longitudinal datasets are accessible for reuse from the Australian Data Archive and the Australian Bureau of Statistics.</p> <p>Education and training around longitudinal data analysis provided by the Australian Consortium for Social and Political Research Inc.</p> <p>Commonwealth Accredited Integrating Authorities have expertise, advice provided via the National Statistical Service on data integration.</p>
<b>HathiTrust</b>	See HathiTrust Research Center <a href="#">Advanced Collaborative Support Projects</a> awarded in 2019.	<p>Trove Digitisation partnership. Large aggregation collection APIs: Trove</p> <p>Access to corpora: Trove Australia; Tinker (Humanities, Arts and Social Science Data Enhanced) &amp; Alveo (Human and Communication Science) Virtual Labs, ARDC funded.</p> <p>Institutional collection APIs: State Library of New South Wales, ACMI, Museum Victoria, National Museum of Australia, State Library of Queensland</p>

### 3.3 Findings from the gap analysis

#### 3.3.1. Consolidation and leadership

The mapping survey highlighted a strong consolidation trend across the international HASS research infrastructures initially established as small projects with a discipline focus, towards larger facilities with a broader Humanities or Social Science focus, or indeed a ‘HASS wide’ focus.

This is particularly evident in Europe where the role of the EU as a regional facilitator and catalyst ensures that a critical mass of expertise and resources from larger nations can enable smaller nations to gain access to research infrastructures that they otherwise might have been unable to sustain themselves.

Data curation expertise and skills and services play a significant role in the development of national (and international) capability, which in turn underpins the integration of research enabling capabilities, such as data mining, analysis (including text analysis), informatics, modelling and visualisation for the HASS community.

This trend has led to increasingly sophisticated governance models being established to support these consolidated research infrastructures exemplified by the European Research Infrastructure Consortium (ERIC), a specific legal structure that facilitates the establishment and operation of Research Infrastructures with European interest (2009).

Funding models have evolved in parallel with these governance structures, often based on a flat fee plus subscription (as a percentage of GDP) from participating nations. These co-investment models are key to ensuring such research infrastructures are sustainable. Some infrastructures, recognising the importance of broad participation, provide guidelines and costs estimates to join, e.g. CLARIN.

#### 3.3.2. Systematic HASS research capacity

Many of the European HASS research infrastructures included in the mapping survey have been developed through project and expansion funding to build upon institutional capabilities in the first instance, and then on aligning and aggregating national capabilities.

This evolution extends back more than 15 years in the social science domains, and over a decade in the arts and humanities domains. Significantly, these research infrastructures have received central (EU) and/or co-contribution (national) funding over that entire period.

Building on existing research practises and peer networks, they are heavily oriented toward improving existing data sets and research methods, particularly when complemented by programs to raise awareness and augment researcher skills. At the European level, five of these research infrastructures are considered landmark facilities (out of a total of 37).

Australia does not have any nationally comparable HASS research infrastructures, and this represents a significant gap in national research capacity. Consequently, research data assets currently enabling HASS research either do not exist in Australia, or are institutionally hosted and uncoordinated, and largely not FAIR (findable, accessible, interoperable and reusable).

The mapping highlights that there are Australian comparators to many of the European research infrastructures operating at an institutional level (such as ADA, the Pacific and Regional Archive for Digital Sources in Endangered Cultures (PARADISEC), and APO), or as a project within a larger program or institution (Trove, or Household, Income and Labour Dynamics in Australia (HILDA)) that are well placed to be evolved into national capabilities.

In Europe there have been two separate streams: humanities and arts, and social sciences. Europe is now embarking on how these can be better connected, which is something Australia has an opportunity to address from the outset. One of the major lessons from Europe is that disconnection between investments has necessitated a later phase ‘harmonisation’ of research infrastructures, both within and across HASS. Harmonisation is now a priority between the current ERIC infrastructures in the social sciences (European Social Survey (ESS), Survey of Health, Ageing and Retirement in Europe (SHARE)) and humanities (CLARIN – European Research Infrastructure for Language Resources and Technology) and DARIAH – Digital Research Infrastructure for the Arts and Humanities) as well as the ESFRI Landmark CESSDA (Consortium of European Social Science data Archives).

### 3.3.3. Access to government data

Much of Australia’s public sector data is locked up and underutilised in HASS research, and in other disciplines. It is clear from the mapping survey that in Europe many of the cross-government jurisdiction challenges that are typical of the Australian federation have been addressed through formal Memoranda of Understanding (MoUs) and EU incentives (fiscal and regulatory). Significant effort has been invested in EU legislation to address data privacy and data rights, which could usefully inform similar considerations in Australia, for example, the broad General Data Protection Regulation (GDPR).

Within our region, the New Zealand Integrated Data Infrastructure (IDI) database is a world-leading exemplar of a research infrastructure that value-adds government data for research. In the context of the current Australian activities of the Office of the National Data Commissioner and the recommendations of the Productivity Commissioner cited above, the NZ IDI appears relevant.

### 3.3.4. Transformative digital capability for humanities

In addition to the broad-based impact across the HASS sector of systemic infrastructure improvement activities outlined above, the international mapping identifies the emergence of new and unique HASS research infrastructures that are heavily data, information and computer science enabled and provide different arrangements of data custody, new data and large scales of data that lead to radically new research techniques.

The examples identified are mostly at the national scale with a focus on long timelines and unlocking large scale cultural heritage datasets (e.g. UK, Netherlands, Denmark, Sweden). Although there are Australian research exemplars that demonstrate this capability (e.g. Virtual Archaeology Unravels Historic Shipwreck Mystery (Pawsey, 2018)), there is no national domain focussed capability to support computationally intensive HASS research of this transformative nature.

### 3.3.5. Regional approach

National strategies and roadmaps for research infrastructure are primarily focussed on providing national outcomes. Significantly, the mapping for this project reveals that these national strategies also recognise the critical importance of participating in, and contributing to, international and regional research infrastructures.

As noted above the regional consolidation and leadership role the EU plays is crucial in bringing together the large community of disparate stakeholder countries and institutions from across Europe to co-fund capabilities that deliver outcomes for all of Europe.

There are lessons here for Australia at a national scale where the Federal government has the capacity to incentivise and catalyse a national federated approach to HASS research infrastructure across the broad stakeholder community. Proven collaborative mechanisms such as the National Collaborative Research Infrastructure Strategy (NCRIS) are well suited to such an opportunity.

At a larger scale, an opportunity exists for Australia to take a significant regional leadership or partnership role in a potential Indo-Pacific wide approach to research infrastructure. This might include, as it does in Europe, expertise and resources from larger nations enabling smaller nations to gain access to research infrastructures that they otherwise might not have been able to sustain themselves, as well as building stronger collaborative ties with HASS research communities within larger nations. Australia's close research links to New Zealand, and New Zealand's success with the IDI (for example), suggest such a program could be readily initiated.

The mapping identifies an international leadership opportunity gap that could be addressed through an investment focussed on Indigenous health, wellbeing and culture research, with several other countries having a related interest in indigenous research (New Zealand, Norway, Canada). Such an approach is also consistent with significant portions of Australia's HASS research interests as well as aligning strongly with Australia's reconciliation agenda, with Australia's Pacific Step-Up and Australia's broader geo-political interests as a 'soft power' (Department of Foreign Affairs and Trade).

## 4 Potential Priorities

---

Potential priority areas for Australian HASS National Research Infrastructure (NRI) development and international collaboration/partnership, are intended to address both the gaps identified and the unique characteristics of the Australian NRI landscape.

The approach recognises that this work will include initiatives that are unique to the HASS community and hence for which the HASS community needs to be primarily responsible, and initiatives that are not unique to HASS, but which need to be guided by HASS and be HASS-relevant.

In summary, Table 9 shows alignment between the international research infrastructures mapped and the proposed programs.

**Table 9: International research infrastructures mapped to Australian priorities**

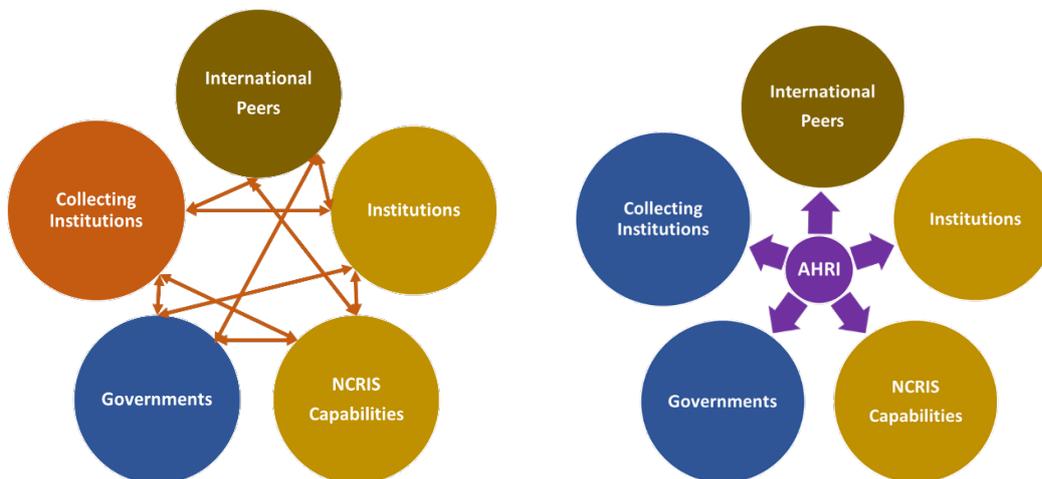
	Coordinating	Data Hub	Digitisation	Peak	Indigenous
<b>ESS</b>	✓	✓			
<b>SHARE</b>	✓	✓			
<b>CLARIN</b>	✓	✓	✓	✓	
<b>DARIAH</b>	✓		✓	✓	
<b>CESSDA</b>	✓	✓			
<b>SAIL</b>	✓	✓			
<b>IMPACT</b>	✓		✓		
<b>EUROPEANA</b>			✓	✓	
<b>HathiTrust</b>			✓	✓	
<b>FNIGC</b>	✓				✓
<b>CLOSER</b>	✓	✓			
<b>MDSN</b>	✓				✓
<b>COMPUTE CA</b>	✓				
<b>R-RIHS</b>			✓	✓	

### 4.1 Australian HASS research infrastructure entity

Based on the lessons from the international mapping survey, one of the highest priorities for Australian HASS national research infrastructure development and investment is the establishment of an entity to create focus, clarify responsibility, maximise collaboration and reduce complexity regarding HASS research infrastructures. This recommendation is a direct response to the findings from the mapping exercise with regards the strong consolidation trend across HASS infrastructures in Europe outlined above and is a means to avoid fragmentation and duplication in the Australian system.

Such an entity would lead requirements gathering and roadmap development, and coordinate, facilitate and partner to deliver tools and systems to support HASS research.

This ‘focal point’ organisation, Australian HASS Research Infrastructure (AHRI; a working title), will simplify the existing mesh of interactions across the national HASS research stakeholder landscape and provide a national point of contact for international collaboration, as indicated below.



**FIG.2 A single entity, Australian HASS Research Infrastructure (AHRI) will create focus, clarify responsibility and reduce complexity**

AHRI is envisaged as having a mission and structure analogous to existing NCRIS capabilities such as Astronomy Australia Limited (AAL) and BioPlatforms Australia (BPA). A membership model that included both research and collecting institutions would be one mechanism to support a co-investment partnership with the Commonwealth.

In addition to the leadership and coordination role AHRI would undertake programs to foster the capacity and capabilities of the Australian HASS sector to boost its contribution to Australia’s health, environmental, economic and social wellbeing. The specific goals of these programs would be to:

- > Bridge the systemic HASS research infrastructure capacity gap through an awareness and upskilling initiative.
- > Liaise with Federal and State government departments to ensure HASS-relevant government datasets are made readily accessible to researchers.
- > Improve the consistency and coherence of access of digital data and physical objects held by collecting institutions by formally engaging the collecting institutions as partners in Australia’s national HASS research infrastructure. Collections will continue to be maintained by the collecting institutions, but it will be the responsibility of AHRI to deliver a layer of access and linkage capability across all these collections to best serve the HASS community’s research needs by addressing issues of HASS data sensitivity and access to government data.
- > Initiate work to identify humanities disciplines that would benefit from a peak HASS ‘compute+data’ capability to support transformative research methods (perhaps related to Indigenous research).

The following five programs are proposed to achieve these goals (described in more detail below):

- > HASS Research Data Commons (HRDC)
- > Social Science Data Hub (SSDH)
- > Indigenous Data Framework (IDF)
- > National Digitisation Capability (NDC)
- > Digital HASS Peak Capability (DHPC).

## 4.2 HASS Research Data Commons

Deliver a HASS Research Data Commons (HRDC) leveraging the experience, expertise and resources of relevant Australian programs (e.g. ARDC, Population Health Research Network (PHRN), and AURIN), relevant HASS stakeholder institutions (e.g. collecting institutions, institutional programs like the Australian Data Archive) and informed by and aligned with international programs. This would be a ‘catch up exercise’ focussed on improving awareness and expertise around some ‘minimally viable’ or ‘common denominator’ data commons tools for researchers and related professional staff (e.g. librarians).

The critical success factor would be to do this quickly and deliver something for as many HASS researchers as possible. The key elements might include the most common categories and capabilities of the programs included in the mapping (Table 10).

**Table 10: Most common categories and capabilities on research infrastructure identified in mapping**

Infrastructure Categories	Infrastructure Capabilities
Capability Building	Data Curation
Community Building	Data Reuse
Information Services	Domain Expertise
Discovery Platform	
Research Practises and Methods	
Data Service	
Research Tools and Platforms	

## 4.3 Social Sciences and Languages Data Hub

Establish a social science and languages data hub (SSDH) to provide a national focus for identifying and facilitating access to government datasets for research and the necessary tools to support research with them. This will include addressing cross-jurisdictional inconsistencies regarding data formats, access, open data policy, and working with ARDC’s sensitive data program on issues related to Indigenous data sovereignty, providence, privacy, etc. in partnership with, for example, PHRN and AURIN.

Social science research infrastructure, specifically longitudinal survey data, has been operational for a longer period than for humanities and arts in Europe and in the UK. Consequently, the global networks around this data infrastructure and associated tools and skills, are more evolved. Three of the five Social and Cultural Innovation (SCI, or HASS) landmark research infrastructures identified in the European Strategy Forum on Research Infrastructures (ESFRI) Roadmap (2018) are social science based – Consortium of European Social Science Data Archives (CESSDA), European Social Survey (ESS) and Survey of

Health, Ageing and Retirement in Europe (SHARE). Given the large number of countries participating in these programs, some from outside Europe, the Australia SSDH should strongly consider joining these programs at the European level, or some of the national programs (e.g. UK's Cohort and Longitudinal Studies Enhancement Resources (CLOSER)).

The HASS community needs to lead the SSDH program in partnership with others.

#### 4.4 Indigenous Data Framework

The 2016 Roadmap identifies a number of existing platforms that support research into Indigenous health, social well-being, culture, language and history, noting that “creating a cohesive platform that harvests information [data], that is interoperable and that provides appropriate levels of accessibility for communities and researchers alike is required” (p.36).

The value of such a platform is well articulated in the Indigenous Data Network (IDN) roadmap (2019):

Addressing the complex issues of disadvantage for Aboriginal and Torres Strait Islander people requires relevant high-quality data. Data provides the evidence-base for the development, implementation and evaluation of effective policy solutions at local, state/territory and national levels. As the Prime Minister's 2019 Closing the Gap Report emphasises, building an evidence base is key to success. However, the lack of reliable and consistent data for Indigenous Australians results in a paucity of evidence-based Indigenous policy-making (p.3).

The Indigenous Data Framework (IDF) program proposed would address barriers to realising this opportunity, by aligning and integrating the challenges of Indigenous data sovereignty, access and repatriation with related efforts to improve and facilitate access to sensitive research data being undertaken by other national and international bodies (including in the fields of health and medical research).

The HASS community needs to be a partner in the IDF program.

#### 4.5 National Digitisation Capability

Digitisation as a national capability is not an explicit goal of most of the programs included in the mapping but is an intrinsic requirement for many programs as it enables access to relevant data sources including historic government records, heritage collections and clinical records. It is recognised as a core capability for many of Australia's collecting institutions including the NFSA (Digitisation Strategy 2018-2015) and NLA (Digitisation, including through Trove). Integrating and aligning these institutional programs to create a national HASS research infrastructure capability is highlighted in the 2016 Roadmap (and 2011 Research Infrastructure Roadmap).

This National Digitisation Capacity (NDC) would be well suited to the ‘distributed node’ model of research infrastructure, and one that can draw on and leverage international programs, notably European Research Infrastructure for Heritage Science (E-RIHS).

The HASS community needs to be a lead in the NDC program.

## 4.6 Digital HASS Peak Capability

The research techniques enabled by the integration of modern computation technologies such as cloud computing and High-Performance Computing (HPC) with existing and newly generated and created big data are already transforming research techniques across many HASS disciplines.

This program would seek to extend and deepen the accessibility, usability and adoption of computationally intense methods across the broader HASS research community in area such as machine learning, artificial intelligence and text mining and analysis. The Digital HASS Peak Capability (DHPC) would necessarily leverage big data and HPC facilitates and other technology platform programs including NCI, Pawsey, and ARDC informed by similar international programs such as those identified in the mapping.

Unlike the other potential programs, which seek to achieve international parity with other national research infrastructures, one of the desired goals of the HASS Peak capability program would be to identify a field of research of significance to Australia where these techniques could be applied in a world-leading capacity. Further analysis is required to identify the specific fields of research in Australian histories and cultures, or Indigenous studies.

The HASS community needs to lead the Digital HASS Peak program.

## 4.7 Implementation

The table overleaf (Table 11) provides an assessment of the overall impact of the potential priorities.

### **EXPLANATORY NOTES:**

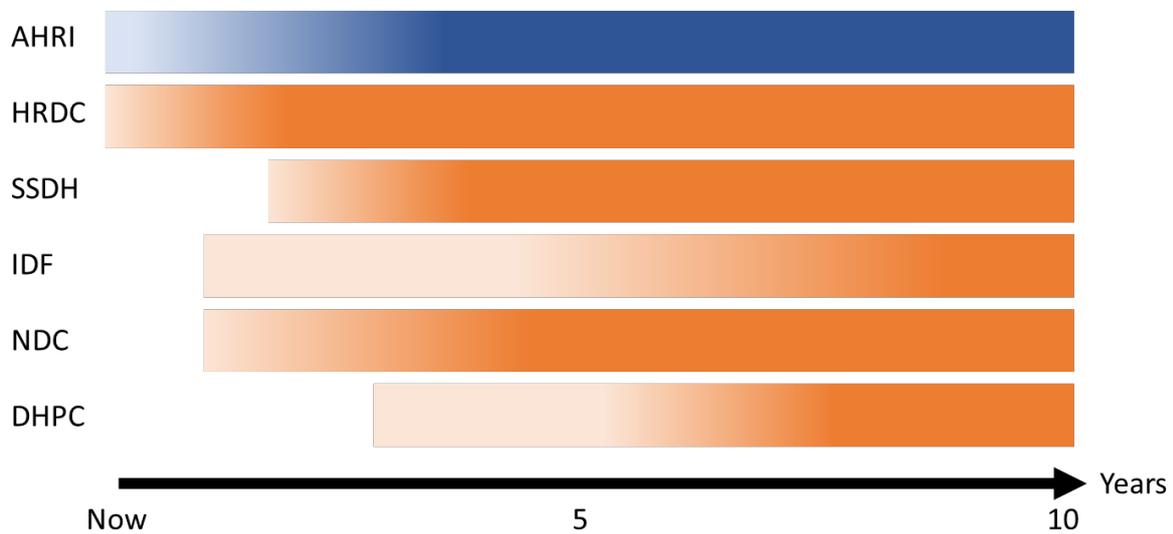
Complexity – Is the solution/approach well understood and can be well defined? Are there exemplars at national or supra-national scale internationally? Is there an existing body of expertise and experience in Australia that can be leveraged?  
Note – timeline is addressed below.

Impact – ‘Overall’ indicates scope of impact across all stakeholders – is a proxy for the estimated size of the research community that will directly benefit; ‘Researchers’ implicitly includes impact for research institutions; ‘Collecting Institutions’ includes public funded galleries, libraries, archives and museums; ‘Governments’ include state and Federal governments (policy) and agencies (service delivery).

**Table 11: Program assessment: cost, complexity, impact**

		Impact			
Program	Complexity	OVERALL	Researchers/Research Institutions	Collecting Institutions	Governments
Australian HASS Research Infrastructure	Low (Small Entity)	Very Broad	Consistent best practise research across sector. Avoid duplication of effort, and drive economies of scale (expertise/infrastructure).	Recognition of unique role in HASS research infrastructure.  A framework to guide engagement with research institutions.	Improved focus, clearer responsibility and less complex HASS engagement.  Clear Alignment between HASS RI investments and policy agenda.
HASS Research Data Commons	Low (Leverage Existing Programs)	Very Broad	Increased awareness and adoption of digital tools and resources;  Builds a common framework for data initiatives and services	Improved collection curation.  Platform to contribute data augmentation (interdependent value)	Aligns with broader data sharing and release and digital transformation policy agendas. Maximises sharing and interoperability of data across sectors.
Social Science Data Hub	Medium (Leverage Existing, Government Engagement)	Broad	Better access to relevant datasets; links to data networks within Australia (e.g. PHRN) and international (e.g. CESSDA); standardises safe, secure environments and outputs.	Longitudinal/linking of historic datasets (census, criminal trial records etc) for research and community use.	Improved understanding of the digital transformation of society.  Better informed evidence-based policy.
Indigenous Data Framework	High (Complex Multi-shareholder issues)	Focussed	Facilitate access to Indigenous data with strong emphasis on community protocols.	Better sharing and linking of knowledge and collections, including vast community-based collections	Better informed evidence-based Indigenous policy, community-led.
National Digitisation Capability	Medium (Existing Institutional Capacity to be Coordinated)	Broad	Improved access to collections underpinned by data as enabler	Digitally preserved collections, improved curation;  Formally engaged as research infrastructure	Leverages investments made in collecting institutions.  Avoid duplication of effort, and drive economies of scale in expertise and infrastructure.
Digital HASS Peak Capability	High (Scoping, Mapping, Analysis)	Focussed	New research techniques enabled by large scale infrastructures	New techniques for engaging public	Social, economic and educational resource. Soft power asset.

The proposed programs are expected to commence in a staggered manner over the next five years with the objectives of delivering aggregate research impact over a decade within available resources. An indicative view of this staged approach is shown in the figure below (Fig.4). Note that the shading is suggestive of when greater intensity and progress may be required, or where existing programs can be leveraged, e.g. the ARDC already has a program of work underway regarding a HASS Research Data Commons.



**FIG.4 Proposed timelines for staged investment**

There is some overlap and mutually supportive interaction between these programs, as highlighted in the table below (Table 12). Darker shading indicates a program directly addresses an AHRI goal; lighter shading indicates the program supports a goal.

**Table 12: Goals and Priorities**

Program (Priority)	Goals			
	Systemic HASS Capacity	Government Data Access for HASS	Integrating Collecting Institutions	Peak Digital HASS
HASS Research Data Commons	Light Blue	Light Blue	Light Blue	
Social Science and Languages Data Hub	Light Blue	Light Blue		
Indigenous Data Framework	Light Blue	Light Blue	Dark Blue	Light Purple
National Digitisation Capability			Dark Blue	Light Purple
Digital HASS Peak Capability				Dark Purple

## Appendix A – Research phases

---

The research undertaken for this project was essentially desktop in design. The project did not have a mandate to undertake consultation with the HASS sector in Australia.

An Advisory Group was established to provide advice on international models and to guide the gap analysis. The breadth of expertise of the group spans key stakeholder communities – HASS research, Indigenous research, the cultural and collecting sector, and universities. The Advisory Group provided feedback at each stage of the research and on the final report.

Based on advice from the Advisory Group the project focused on select international models or exemplars. The analysis in this report is primarily informed by a detailed mapping of research infrastructures in Europe, the UK, the Netherlands, USA, Canada and New Zealand.

The key phases of research and focus questions were as follows:

---

### 1. MAPPING

Identify and map key international infrastructures in the HASS domain at national and pan national scale.

This phase of work addressed the following questions:

1. What international examples can Australia learn from?
2. How do they operate?
3. Are there international best practice models?

The mapping produced a database of international and pan national HASS infrastructure categorised according to type, scale, funding, costs, organisational/governance structure, communities of users. Where possible the project was able to draw out some data on levels of internationalisation, outputs, impact and sustainability.

---

### 2. GAP ANALYSIS

Identify gaps in Australian HASS infrastructure as informed by the analysis of the mapping of international models. This phase of work addressed the following questions:

4. Are there international exemplars which could apply or be adapted to Australian HASS NRI development?
5. What are the conditions required for Australia to operate similar models to those international exemplars?

The gap analysis drew on the project's Advisory Group for its:

- > Evaluation framework for assessing international HASS infrastructures and use-value for Australia.
  - > Analysis of the gaps in RI in Australia vs international.
  - > Identification of potential priorities and areas of focus, such as digitisation, international interoperability and access and data management infrastructures.
- 

### 3. OPPORTUNITIES

The final report to inform consideration of research infrastructure investment and delivery for HASS NRI in Australia, includes an analysis of lessons and opportunities from international HASS research infrastructure investments, a list of potential areas of cooperation/collaboration and options for HASS NRI into the future, and identifies where further work needs to be undertaken.

---

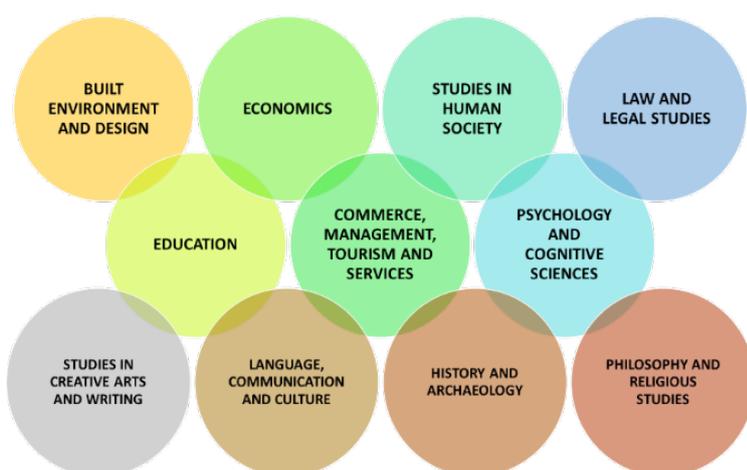
## Appendix B: HASS Research

---

Humanities, Arts and Social Science (HASS) research enhances understanding about Australia, the Australian people, and Australia's place in the world.

More broadly, HASS research provides new frameworks for the analysis of people in the world and the products and outcomes of human activity. The disciplines that make up the HASS sector are fundamental to the development of a knowledge-based economy and in developing inter-disciplinary solutions to a broad range of complex problems and issues such as climate change, social cohesion, the impact and relevance of technological development, future workforce solutions, resource management, health and welfare.

The humanities investigate how people experience, understand and describe the world and their place in it. The humanities examine human cultures, values and beliefs (Australian Academy of the Humanities, 2019). The social sciences work on the systematic development of logic and evidence to understand human behaviour in its social setting, including the nature of economic, political, and community activity and institutions (Academy of the Social Sciences in Australia, 2019).



HASS comprises more than 50 disciplines at the granular four-digit field of research level (Fig.1) (Appendix A).

Australian HASS researchers comprise 41 per cent (16,488 FTE) of the university-based system, based on the latest Excellence in Research Australia (ERA) audit (Australian Research Council, 2019).

**FIG.1 HASS FIELDS OF RESEARCH**

*SOURCE:* Australian Bureau of Statistics (2008), ANZSRC Fields of Research.

HASS research is undertaken at every university in Australia, and HASS is responsible for teaching 61 percent (606,721 students) of the university population in Australia (derived from Department of Education higher education statistics collection, where humanities, arts and social sciences (HASS) comprises, at the broad Field of Education level 12 to 22).

## Humanities, Arts and Social Sciences (HASS) fields of research, two- and four-digit level

Social, Behavioural and Economic Sciences (SBE)	Humanities and Creative Arts (HCA)
13 EDUCATION	12 BUILT ENVIRONMENT AND DESIGN
1301 Education Systems	1201 Architecture
1302 Curriculum and Pedagogy	1202 Building
1303 Specialist Studies in Education	1203 Design Practice and Management
1399 Other Education	1204 Engineering Design
14 ECONOMICS	1205 Urban and Regional Planning
1401 Economic Theory	1299 Other Built Environment and Design
1402 Applied Economics	18 LAW AND LEGAL STUDIES
1403 Econometrics	1801 Law
1499 Other Economics	1802 Maori Law
15 COMMERCE, MANAGEMENT, TOURISM AND SERVICES	1899 Other Law and Legal Studies
1501 Accounting, Auditing and Accountability	19 STUDIES IN CREATIVE ARTS AND WRITING
1502 Banking, Finance and Investment	1901 Art Theory and Criticism
1503 Business and Management	1902 Film, Television and Digital Media
1504 Commercial Services	1903 Journalism and Professional Writing
1505 Marketing	1904 Performing Arts and Creative Writing
1506 Tourism	1905 Visual Arts and Crafts
1507 Transportation and Freight Services	1999 Other Studies in Creative Arts and Writing
1599 Other Commerce, Management, Tourism and Services	20 LANGUAGE, COMMUNICATION AND CULTURE
16 STUDIES IN HUMAN SOCIETY	2001 Communication and Media Studies
1601 Anthropology	2002 Cultural Studies
1602 Criminology	2003 Language Studies
1603 Demography	2004 Linguistics
1604 Human Geography	2005 Literary Studies
1605 Policy and Administration	2099 Other Language, Communication and Culture
1606 Political Science	21 HISTORY AND ARCHAEOLOGY
1607 Social Work	2101 Archaeology
1608 Sociology	2102 Curatorial and Related Studies
1699 Other Studies in Human Society	2103 Historical Studies
17 PSYCHOLOGY AND COGNITIVE SCIENCES	2199 Other History and Archaeology
1701 Psychology	22 PHILOSOPHY AND RELIGIOUS STUDIES
1702 Cognitive Science	2201 Applied Ethics
1799 Other Psychology and Cognitive Sciences	2202 History and Philosophy of Specific Fields
	2203 Philosophy
	2204 Religion and Religious Studies
	2299 Other Philosophy and Religious Studies

Source: Australian Bureau of Statistics (2008) 'Australian and New Zealand Standard Research Classification (ANZSRC)', cat. no. 297.0. Available from

<http://www.abs.gov.au/Ausstats/abs@.nsf/Latestproducts/4AE1B46AE2048A28CA25741800044242?opendocument>

## Appendix C: HASS Data

---

### HASS Research Infrastructures

Research infrastructures in HASS comprise:

- > Physical collections (including artefacts and larger physical structures such as archaeological sites) and their storage facilities
- > Galleries, Libraries, Archives and Museums (including both physical objects and digital artefacts)
- > Digital infrastructure (including digital record creation, digital data storage and tool sets)
- > Laboratory facilities for heritage science and archaeology
- > Infrastructures for data collection, services, linkage and analysis, such as public statistics and longitudinal surveys. Data collection may span topic areas such as election studies, involve long-term multi-generational population studies and surveys, and more recently social media data

The assemblages of data – data files, datasets, databases and data streams – collected, generated and curated by these infrastructures have all the characteristics of ‘big data’ in that their volume, their variety and the velocity of their creation pose severe challenges for many conventional analytical and computational methods. As a result, HASS research infrastructures, both internationally and in Australia, are increasingly focused on equipping HASS researchers with the tools and techniques to take on these challenges.

### Government data

The Australian HASS community, like other research communities, makes extensive use of data held by federal and state agencies, including Departments of Health, Departments of Education, Departments of Social Services, Departments of Innovation and Industry, the Australian Bureau of Statistics (ABS), the Australian Taxation Office (ATO) and many others. There are significant challenges in achieving effective and consistent access to data across all jurisdictions of Australia’s federation, that are not unique to the HASS sector, as identified by the Productivity Commission’s report into *Data Availability and Use* (2017):

Governments across Australia hold enormous amounts of data, but mostly lag behind other comparable economies in beneficially using data beyond the purposes for which it was initially collected, or allowing others access to do so (p.24).

The Government’s response proposed reforms to:

... empower Australian citizens, governments, industries and researchers to use and share data, while maintaining the strict privacy, security and transparency safeguards essential to maintain trust in the system.

With the following goal:

These advances to Australia's data system will mean we can harness the power of data to drive innovation and opportunity for the Australian economy (p.1).

The HASS research sector is fundamental to realising this goal through the development of a social license to use data consistent with current activities with the Office of the National Data Commissioner and the Data Sharing and Release reforms. Using mechanisms like the EU Responsible Research and Innovation Toolkit (which helps to “align ... research and innovation processes to societal needs and challenges” (Responsible Research and Innovation Project, n.d)) alongside the Five-Safes Framework (“a multi-dimensional approach to managing disclosure risk” (Australian Bureau of Statistics, 2017)) will provide a way to secure and maintain the social license, which will be fundamental to using integrated data for research.

### Indigenous data

Improving the lived experience of Australia's first nations peoples requires that data is made accessible for research. Advancing cultural connection and social cohesion of Australians and the wider Indo-Pacific region and globally requires data access and sharing, and services and processing tools that are co-located. Improving social and economic outcomes for Australians will be underwritten by the quality of data and data linkage that will bring together longitudinal social survey data with health and environmental data.

Research related to Australia's Aboriginal and Torres Strait Islander people (Indigenous research) faces challenges relating to data sovereignty and access.

Indigenous data needs a variety of community consent and access controls to allow Indigenous communities to access and selectively and safely release their data to individual researchers. A similar process also allows digitised materials to be repatriated to Indigenous communities.

Responses to the 2016 Roadmap, from a range of stakeholders, emphasised the need for Indigenous leaders and organisations to be at the heart of determining capabilities required to support Indigenous aspirations and needs. Indigenous co-participation will be essential in any project involving Indigenous communities directly and consideration given to cultural sensitivities and associated rights.

Several organisations in Australia are working on Indigenous data governance and access policies and strategies, including the Australian Institute for Aboriginal and Torres Strait Islander Studies (AIATSIS), the Indigenous Data Network (based at the University of Melbourne), and the National Aboriginal and Torres Strait Islander Consortium.

### Collecting institutions

Galleries, Libraries, Archives and Museums (GLAMs), the collecting institutions, are therefore important stakeholders in Australian national HASS research infrastructure, particularly those that are federally funded with an explicit national collection mandate, e.g. National Library of Australia (NLA), National Archives of Australia (NAA), National Museum of Australia (NMA), and the National Film and Sound Archive (NFSA).

HASS researchers rely heavily and opportunistically upon the digitisation and digital access agendas of these public sector institutions. The value of these digital services is highlighted

by, for example, the frequent reference by individual researchers and bodies within the HASS sector to the usefulness and importance of the Trove service provided by the NLA.

Although nationally funded collecting institutions typically have a national mandate to collect, it is *not* for the sole purpose of supporting research, as the ‘mission to share’ is only a portion of their typical Collect-Preserve-Share charter. Consequently, a key aspect of the relationship between the HASS research community and the collecting and cultural institutions is that there is a high level of interdependent value. Research infrastructures that improve access to objects within collections facilitates research that creates new knowledge and data that can be attached to the objects. This process improves the collection, enables the linking of data and objects between collections and improves the value of the collection to the public.

Existing programs that operate across many collecting institutions in support of their activities enable and facilitate access to collections for researchers, and hence can also be considered components of national research infrastructure, e.g. the National eDeposit system is a national system operating across nine jurisdictions (and related legislation, accounting standards, Crown Solicitors, collecting and description policies).

## References

---

- Academy of Social Sciences in Australia (2019), 'About the Academy', <https://www.assa.edu.au/about-the-academy/>
- Alan Turing Institute (2019), 'Living With Machines', <https://www.turing.ac.uk/research/research-projects/living-machines>
- Arts and Humanities Research Council (AHRC) (2019), AHRC Network for a Global Dataset of Digitised Texts, <https://gddnetwork.arts.gla.ac.uk/>
- Astronomy Australia Limited (n.d), <http://www.astronomyaustralia.org.au/>
- Australian Academy of the Humanities (2019), 'What are the Humanities?', <https://www.humanities.org.au/about-us/what-are-the-humanities/>
- Australian Bureau of Statistics (2017) 'Managing the Risks of Disclosure: the Five Safes Framework', 1160.0 - ABS Confidentiality Series, August 2017, <https://www.abs.gov.au/ausstats/abs@.nsf/Latestproducts/1160.0Main%20Features4Aug%202017>
- Australian Bureau of Statistics (2008) 'Australian and New Zealand Standard Research Classification (ANZSRC)', cat. no. 297.0, <http://www.abs.gov.au/Ausstats/abs@.nsf/Latestproducts/4AE1B46AE2048A28CA25741800044242?opendocument>
- Australian Government (2018), *2016 National Research Infrastructure Roadmap* [https://docs.education.gov.au/system/files/doc/other/ed16-0269\\_national\\_research\\_infrastructure\\_roadmap\\_report\\_internals\\_acc.pdf](https://docs.education.gov.au/system/files/doc/other/ed16-0269_national_research_infrastructure_roadmap_report_internals_acc.pdf)
- Australian Government (2011), *2011 Strategic Roadmap for Australian Research Infrastructure*, [https://docs.education.gov.au/system/files/doc/other/national\\_collaborative\\_research\\_infrastructure\\_strategic\\_roadmap\\_2011.pdf](https://docs.education.gov.au/system/files/doc/other/national_collaborative_research_infrastructure_strategic_roadmap_2011.pdf)
- Australian Government (2008), *Strategic Roadmap for Australian Research Infrastructure* [https://docs.education.gov.au/system/files/doc/other/national\\_collaborative\\_research\\_infrastructure\\_strategic\\_roadmap\\_2008.pdf](https://docs.education.gov.au/system/files/doc/other/national_collaborative_research_infrastructure_strategic_roadmap_2008.pdf)
- AustLit (n.d), <https://www.austlit.edu.au/>
- Analysis and Policy Observatory (n.d), <https://apo.org.au/>
- Australian Research Council (2019a), *ERA National Report: State of University Research 2018-19*, <https://dataportal.arc.gov.au/ERA/NationalReport/2018/>
- Australian Research Council (2019b), *Engagement and Impact Assessment 2018*, <https://dataportal.arc.gov.au/EI/NationalReport/2018/>

- Australian Research Council (2019c) Impact Study: ‘Young and Resilient: Transforming research, policy and practice to maximise the benefits of digital technologies for young people’, <https://dataportal.arc.gov.au/EI/Web/Impact/ImpactStudy/800>
- Australian Research Council (2019d) Impact Study: ‘The Household, Income and Labour Dynamics in Australia (HILDA) Survey: Changing the Way We Think About Life in Australia’, <https://dataportal.arc.gov.au/EI/Web/Impact/ImpactStudy/872>
- Australian Research Council (2019e), Media Release: Centre of Excellence for Automated Decision-making and Society, 9 October, <https://www.arc.gov.au/news-publications/media/media-releases/improving-automated-decision-making>
- Australian Research Council (2019f), Media Release: Centre of Excellence for the Digital Child, 27 September, <https://www.arc.gov.au/news-publications/media/media-releases/supporting-children-digital-age>
- BioPlatforms Australia (n.d), <https://www.bioplatforms.com/what-we-do/>
- CLARIN (n.d) Information for Potential Members’, <https://www.clarin.eu/content/information-potential-new-members>
- Commonwealth of Australia, Department of the Prime Minister and Cabinet, (2018), *The Australian Government’s response to the Productivity Commission Data Availability and Use Inquiry*, <https://www.pc.gov.au/inquiries/completed/data-access/data-availability-use-government-response.pdf>
- Department of Foreign Affairs and Trade (n.d) ‘Soft Power Review’, <https://dfat.gov.au/people-to-people/soft-power-review/Pages/soft-power-review.aspx>
- European Commission (2009) *Community Legal Framework for a European Research Infrastructure Consortium (ERIC)* [https://ec.europa.eu/research/infrastructures/pdf/council\\_regulation\\_eric.pdf](https://ec.europa.eu/research/infrastructures/pdf/council_regulation_eric.pdf)
- European Commission (n.d), ‘What Are Research Infrastructures?’, [https://ec.europa.eu/info/research-and-innovation/strategy/european-research-infrastructures\\_en](https://ec.europa.eu/info/research-and-innovation/strategy/european-research-infrastructures_en)
- European Commission (n.d), ‘What is an ERIC?’, [https://ec.europa.eu/info/research-and-innovation/strategy/european-research-infrastructures/eric\\_en](https://ec.europa.eu/info/research-and-innovation/strategy/european-research-infrastructures/eric_en)
- European Strategy Forum on Research Infrastructures (2018), *Strategy Report on Research Infrastructures Roadmap Part 2: Landscape Analysis* <http://roadmap2018.esfri.eu/media/1050/roadmap18-part2.pdf>
- European Strategy Forum on Research Infrastructures (n.d), ESFRI Glossary <https://www.esfri.eu/glossary>
- European Union (n.d) General Data Protection Legislation, <https://eugdpr.org/>
- Indigenous Data Network (n.d) <https://mspgh.unimelb.edu.au/centres-institutes/centre-for-health-equity/research-group/indigenous-data-network>
- Indigenous Data Network (2019) *Indigenous Data Network Roadmap*, [https://mspgh.unimelb.edu.au/\\_data/assets/pdf\\_file/0011/3070487/NIDN-2019-03-08-Indigenous-Data-Network-Roadmap-190301A.pdf](https://mspgh.unimelb.edu.au/_data/assets/pdf_file/0011/3070487/NIDN-2019-03-08-Indigenous-Data-Network-Roadmap-190301A.pdf)

- National Collaborative Research Infrastructure Strategy (NCRIS) Network (2017), 'Facilitating Real Research Impact: NCRIS Supports Prestigious ARC Centres of Excellence', <https://www.ncris-network.org.au/single-post/2017/11/06/Facilitating-real-research-impact>
- National Film and Sound Archive (2018), *Digitisation Strategy 2018-2025*, <https://www.nfsa.gov.au/footer/corporate-information/publications/digitisation-strategy>
- National and State Libraries Australasia (2019), National eDeposit, <https://ned.gov.au/portal/>
- National Library of Australia (n.d), Trove, <https://trove.nla.gov.au/>
- Office of the National Data Commissioner (2019), Data Sharing Principles, <https://www.datacommissioner.gov.au/safeguards/sharing-principles>
- Responsible Research and Innovation Project (n.d), 'The PRISMA Responsible, Research & Innovation Toolkit', <https://www.rrri-tools.eu/-/the-prisma-responsible-research-innovation-toolkit>
- Pawsey Supercomputing Centre (2016), 'Virtual Archaeology Unravels Historic Shipwreck Mystery', [https://pawsey.org.au/wp-content/uploads/2018/10/AndrewWoods\\_HMASSydneyJuly16web.pdf](https://pawsey.org.au/wp-content/uploads/2018/10/AndrewWoods_HMASSydneyJuly16web.pdf)
- Productivity Commission (2017), *Data Availability and Use*, <https://www.pc.gov.au/inquiries/completed/data-access/report/data-access.pdf>
- Securities Industry Research Centre of Asia-Pacific (n.d), <https://www.sirca.org.au/about-sirca/>
- Statistics New Zealand (2019), Integrated Data Infrastructure, [http://archive.stats.govt.nz/browse\\_for\\_stats/snapshots-of-nz/integrated-data-infrastructure.aspx](http://archive.stats.govt.nz/browse_for_stats/snapshots-of-nz/integrated-data-infrastructure.aspx)
- Statistics New Zealand (n.d), Integrated Data Tools, <https://www.digital.govt.nz/showcase/integrated-data-tools/>
- UK Research and Innovation (2019), UKRI Infrastructure Roadmap Progress Report, <https://www.ukri.org/files/infrastructure/progress-report-final-march-2019-low-res-pdf/>