ACADEMY OF THE SOCIAL SCIENCES IN AUSTRALIA

Doing social science in 2032:

Charting national research infrastructure priorities

A discussion paper inviting input into the Academy's Decadal Plan for Social Science Research Infrastructure 2023–32

July 2023

PARTNERS

Life Course Centre
ARC Centre of Excellence for Children and Families over the Life Course

cepar ARC CENTRE OF EXCELLENCE IN POPULATION AGEING RESEARCH

Australian National University

THE UNIVERSITY OF QUEENSLAND AUSTRALIA

ADM+S ARC Centre of Excellence for Automated Decision-Making and Society

About this publication

This discussion paper supports stakeholder consultation for the development of a *Decadal Plan for Social Science Research Infrastructure 2023-32*. The Decadal Plan seeks to build sector-wide consensus around the research capabilities social science researchers will need, over the next 10 years, to take advantage of an unfolding digital revolution to produce insights and outputs at the very cutting edge of global knowledge. This project is undertaken as a partnership between the Academy of the Social Sciences in Australia, the ANU Centre for Social Research and Methods (CSRM), the ARC Centre of Excellence for Automated Decision-Making and Society (ADM+S), the ARC Centre of Excellence for Children and Families over the Life Course (Life Course Centre), the ARC Centre of Excellence in Population Ageing Research (CEPAR) and the UQ Institute for Social Science Research (ISSR).

Paper prepared by Dr Isabel Ceron, Dr Chris Hatherly and Nikita Sharma.

July 2023

Academy of the Social Sciences in Australia

26 Balmain Crescent

Acton ACT 2601

Australia

www.socialsciences.org.au

Suggested citation: Academy of the Social Sciences in Australia. (2023). *Doing Social Science in 2032: Charting national research infrastructure priorities*. DOI 10.5281/zenodo.8173275

Contents

Introduction 2

1 | Producing, discovering and accessing data 8

Assets 10

Systems 14

Rules 18

Skills 20

Consultation questions 21

2 | Analysing data to generate new knowledge 24

Assets 26

Systems 27

Rules 28

Skills 29

Consultation questions 29

3 | Brokering high-value partnerships for innovation 30

Consultation questions 31

Next-level research: What will it take?

Have your say

This Discussion Paper is inviting input into the Academy's *Decadal Plan for Social Science Research Infrastructure 2023–32*.

We encourage responses to this Discussion Paper from anyone interested in improving the resources and tools available to conduct social science research. This includes stakeholders from universities, government, industry, the nonprofit sector, and the broader community.

The Discussion Paper starts by presenting a stocktake of current infrastructure (undertaken between January and April 2023) and the views of research and technical experts about both strengths and gaps of our infrastructure ecosystem.

The Discussion Paper then poses specific questions for readers, aimed at defining current and future infrastructure needs and priorities for the sector.

The consultation period will be open for four weeks, from 24 July until 18 August 2023. You can **respond to the consultation questions via email** to submissions@socialsciences.org.au.

Alternatively, we are inviting motivated researchers, technical experts, capability leads, funding partners, and anyone willing to co-draft the plan to **join the Decadal Plan Working Group**. Express your interest by contacting project lead Dr Isabel Ceron at isabel.ceron@socialsciences.org.au.

The Decadal Plan

What is a Decadal Plan? In one sense, it is exactly that: a plan for achieving a defined set of objectives over a 10-year timeframe. More importantly, it is also the product of a consensus-making process undertaken by a research community about their shared vision, needs and aspirations and how they will come together to achieve them.

Decadal Plans have been used to achieve significant infrastructure and capacity uplift in several Australian research sectors (astronomy, geoscience, nutrition science and geography).

The Decadal Plan for Social Science Research Infrastructure 2023–32 will be the first of its kind for the *Academy of the Social Sciences in Australia*; developed in partnership with five major social science research institutions and centres. The Decadal Plan for Social Science Research Infrastructure 2023–32 aims to deliver:

- **A unified vision** about the directions for research infrastructure
- **Greater sectoral coordination** for more productive and efficient research
- **Proactive responses to societal challenges,** like reducing inequality, tackling climate change, adapting to demographic change, and more
- **A pathway for significant public investment** in key infrastructure priorities.

Scope and definitions

The Decadal Plan defines its scope around four types of research infrastructures:

- **Assets:** The resources that we share and directly add value to research projects (such as data collections)
- **Systems:** The architecture enabling the production, maintenance and sharing of those assets
- **Rules:** The strategic and regulatory frameworks influencing behaviour and decision-making within the ecosystem
- **Skills:** The supports available for individuals to gain the competencies required to make best use of the assets, systems and rules. This includes supports required by the research workforce at large (users of infrastructure), as well as the mechanisms and incentives to train and retain the highly technical workforce needed to design and operate research infrastructure facilities.

An initial stocktake identified over 800 individual infrastructures. To simplify the analysis of this complex landscape, the paper is divided into three sections, each focusing on the challenges inherent to a specific aspect of the research process:

- Producing, discovering and accessing data (Section 1)
- Analysing data to generate new knowledge (Section 2)
- Brokering high-value partnerships for innovation (Section 3).

A two-part challenge

In defining a strategic pathway for the coming decade, the social sciences sector faces two distinct tasks:

- **Optimising:** Understanding the existing ecosystem, including key stakeholders, well-defined technical aspirations (e.g., FAIR and CARE principles for data reuse), regulations, available funding streams and, importantly, any specific capabilities relevant to the social sciences that already exist, with the goal of realising and optimising any achievable benefits for our sector
- **Influencing:** Critically analysing the ecosystem, to identify components that either may hinder or inadequately support the social sciences, or which are currently missing (including any uniquely relevant to our disciplines), with the goal of advocating together to fill identified gaps.

We welcome responses to this Discussion Paper that provide concrete examples of the specific circumstances or contexts affecting your team's productivity or hindering your research aspirations. This paper outlines the infrastructures that already exist, and we are now inviting you to share your experiences and insights on how these infrastructures are currently functioning for you. A detailed understanding of the practical needs of the sector is crucial to a successful Decadal Plan. ■

Infrastructure to tackle the toughest social problems

We know some of the challenges facing social science research over the coming decades, but do we have the right infrastructures to do our best research? Which capabilities could make a difference for Australian social science researchers and research teams?

INEQUALITY

Never before have there been greater differences in opportunity, health, wealth and wellbeing between prosperous Australians, and the almost one million experiencing long-term poverty. How can research infrastructures better support social scientists to understand and address inequality?

AGEING POPULATION

People are living longer lives. Most will live independently, but a significant portion will need moderate to high levels of health and aged care, leading to increased public spending. What research infrastructures can help social scientists best anticipate and inform government responses?

DIGITAL TRANSITIONS

The digital technologies reshaping our society and economy bring productivity gains, but also pose challenges such as bias in autonomous systems and workforce disruptions. What type of intelligence is required to aid decision-making, and what infrastructures would best support it?

CLIMATE CHANGE

Climate change poses threats to population well-being, safety, infrastructure resilience, and sustainable housing, requiring effective adaptation through broader and more granular datasets, and sophisticated analytical tools. Is our research infrastructure fit for the task?

DEMOCRACY

Australia's peace and stability hinge on strong governance, accountability, and social cohesion, amid a complex geopolitical landscape. With the spread of fake news and disinformation campaigns, what data and tools do researchers need to maintain a grip on these challenges?

Established capabilities to expand

Large-scale linked data assets

Australian social science enjoys vast opportunities today, thanks to a growing collection of linked data assets assembled over decades by organizations like PHRN, AURIN, and ABS. The upcoming decade offers a chance to further expand these assets, making them more accessible to a broader range of researchers.

Evidence-based housing policy

The Australian Housing Data Analytics Platform (AHDAP) exemplifies how social science research can bolster evidence-informed government decisions. The platform brings together nationally significant, harmonised housing datasets, which they use to power a suite of modelling and decision-support tools. Their innovations have proven successful with specific localities and are now ready to be scaled up to assist urban planning nationwide.

Empowering researchers to study online and social networks

Innovative researchers at ANU's VOSON Lab and University of Melbourne's MelNet have created software to efficiently collect, curate and analyse social network data. Although predominantly used to analyse social media behaviour, these tools could be adopted by social science disciplines nationally, to support the analysis of a much broader range of online networked activities and behaviours.

Capabilities under development

Data sovereignty for all Australians

Organisations like the Mayam Nayri Wingara Collective and the Improving Indigenous Research Capability project (ARDC) are actively working to operationalise the CARE principles for straightforward application in data management. Their efforts are pivotal to securing ethical use of human data in research.

Social Data and Digital Platform Observability

ADM+S is proposing an *Australian Social Data Observatory* for collecting and analysing social data using innovative new approaches such as data donations, crowdsourcing and test environments.

Real-time urban simulations through 'digital twin city' capabilities

AURIN is leading the Liveable City Digital Twin project, which will create a virtual representation of urban environments to aid in planning and managing cities for resilience. Real-time simulation models will allow planners and policymakers to assess policy impacts and improve urban liveability and climate adaptability.

The next frontier

Cracking the 'social genome': Multi-generational linked records

We have already seen the power of linked data assets like MADIP. What if we supercharged those assets with the digitised, linked historical records of past generations of Australians, all the way back to the time of colonisation? Academy Fellow Janet McCalman is spearheading this initiative with potential groundbreaking applications, including studies of inter-generational disadvantage, immigration, genetics vs environment in life outcomes, or the documentation of Indigenous genealogies.

Longitudinal ageing studies

Australia does not have a comprehensive longitudinal study of ageing that allows social scientists to understand, model and predict positive and negative correlates and outcomes of Australia's ageing population. Investment in such infrastructure would allow Australia to more clearly and confidently chart a path towards the future allowing all Australians to age well and with dignity.

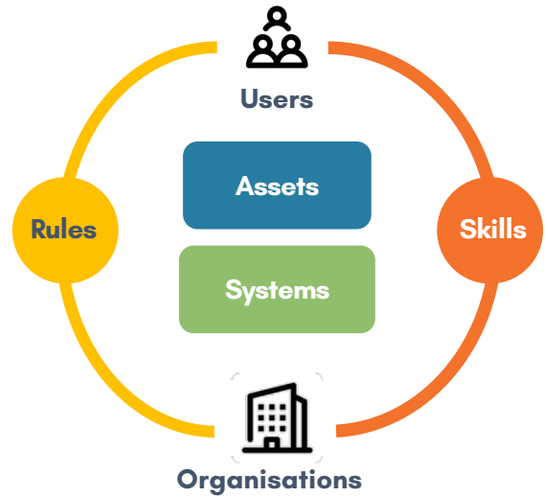
Infrastructure to empower the social sciences in a shifting technological landscape

A rapidly evolving technological landscape present both opportunities and challenges for social science research. Large language models, generative AI, internet of things, blockchain, Web3, automation, robotics, driverless cars, the metaverse, drones... Which research infrastructures can give the sector a firm grasp on these technologies?

Building blocks of research infrastructure

The research Infrastructure ecosystem

The sum of assets, systems, rules and training helping us do better, bigger, faster research.



Assets

Data, physical artefacts, software, workflows and other resources that we share, and which researchers can apply directly to produce new knowledge or technological innovations.

Physical collections
Tangible resources such as objects, specimens or documents.

Digital collections
Documents, datasets, images, videos, or other content in digital format.

Analysis tools
Software, code, workflows or similar, used in data transformation, analysis or visualisation.

Systems

The architecture underpinning or enabling the production, maintenance and sharing of those assets.

Systems that make other systems more efficient through automation and interoperability

Collection
The systems and tools used to collect data, such as survey capabilities, citizen science, digitisation, and data donations.

Curation and stewardship
The work needed to make data findable and (re) usable: licensing, preservation, discipline-specific annotation, and more.

Discovery tools
Components that facilitate the exploration and finding of resources, such as online directories, and accompanying metadata.

Storage
The systems used to store and manage physical and digital collection, including cloud-based storage.

Access management
The systems and protocols to regulate access to resources such as data encryption and user authentication.

Virtual desktop and HPC
Remote virtual environments and equipment to support research analysis, including high-performance computing (or HPC, for large-scale data analysis).

Technical tools and standards
Solutions that ensure consistency and interoperability across capabilities, encompassing the application of metadata or vocabularies, among other elements.

API support tools
Platforms, libraries, frameworks and other tools to support the development, deployment and management of APIs.

PID generation
Tools and frameworks to create and assign Persistent Identifiers (PIDs) to research outputs, such as datasets, publications, or software.

Rules

The strategic and regulatory frameworks and actions influencing stakeholder behaviour within the ecosystem, from national legislation down to best practice standards.

Strategic and regulatory policy
The principles and rules that govern decision-making, resource allocation, and the sustainable operation of research infrastructure.

Funding streams
Government grants, institutional funding, industry partnerships, and other sources of funding to establish, operate, and maintain research infrastructure.

Leadership capabilities
Expert groups, committees and similar who collaborate to develop standards of practice, or advocate for improvements, on behalf of the sector or specific communities.

Skills

The supports available for individuals to upskill. Includes support for the broader research workforce (users of infrastructure) and incentives to train and retain the highly-skilled technical personnel needed to design and operate research infrastructure facilities.

Data archival and management
The range of specialist skills required by data custodian organisations, to preserve and manage physical or digital collections.

IT development
The range of skills required to create and sustain the diverse tools and platforms offered as research infrastructures and facilitate their seamless integration.

Data science
Including skills in machine learning and AI, which are becoming essential for handling the rising volume and complexity of data. Crucial for pattern discovery and predictive modelling.

Discipline-specific research skills
To ensure data is collected, managed and analysed in alignment with the unique requirements and methodologies of each discipline; fundamental for designing usable and effective infrastructures.

1 | Producing, discovering and accessing data

This section examines the current state of infrastructure supporting data production, discovery and access (Figure 1), on the assumption that shared infrastructure can significantly increase research impact, quality and productivity.

Collective benefits

Research impact

- Increase availability and diversity of data we collect, access and share, to respond to Australia's most pressing challenges
- Ensure Indigenous researchers and communities have access to the data they need to rebuild the nation, and appropriate control over access and use
- Secure long-term preservation of data of significant heritage, historical, or longitudinal value
- Facilitate data access (where possible) to the broader community.

Research quality

- Bring currently dispersed high-value data under collective stewardship, for increased access, reuse and interoperability
- Collectively broker access to data held by non-academic institutions, such as government or the private sector
- Facilitate access to international data, for global or comparative research.

Research productivity

- Implement systems to efficiently identify relevant data across disciplines, sources and domains
- Formulate nation-wide, streamlined processes and criteria to ethically access sensitive data for research purposes. ■

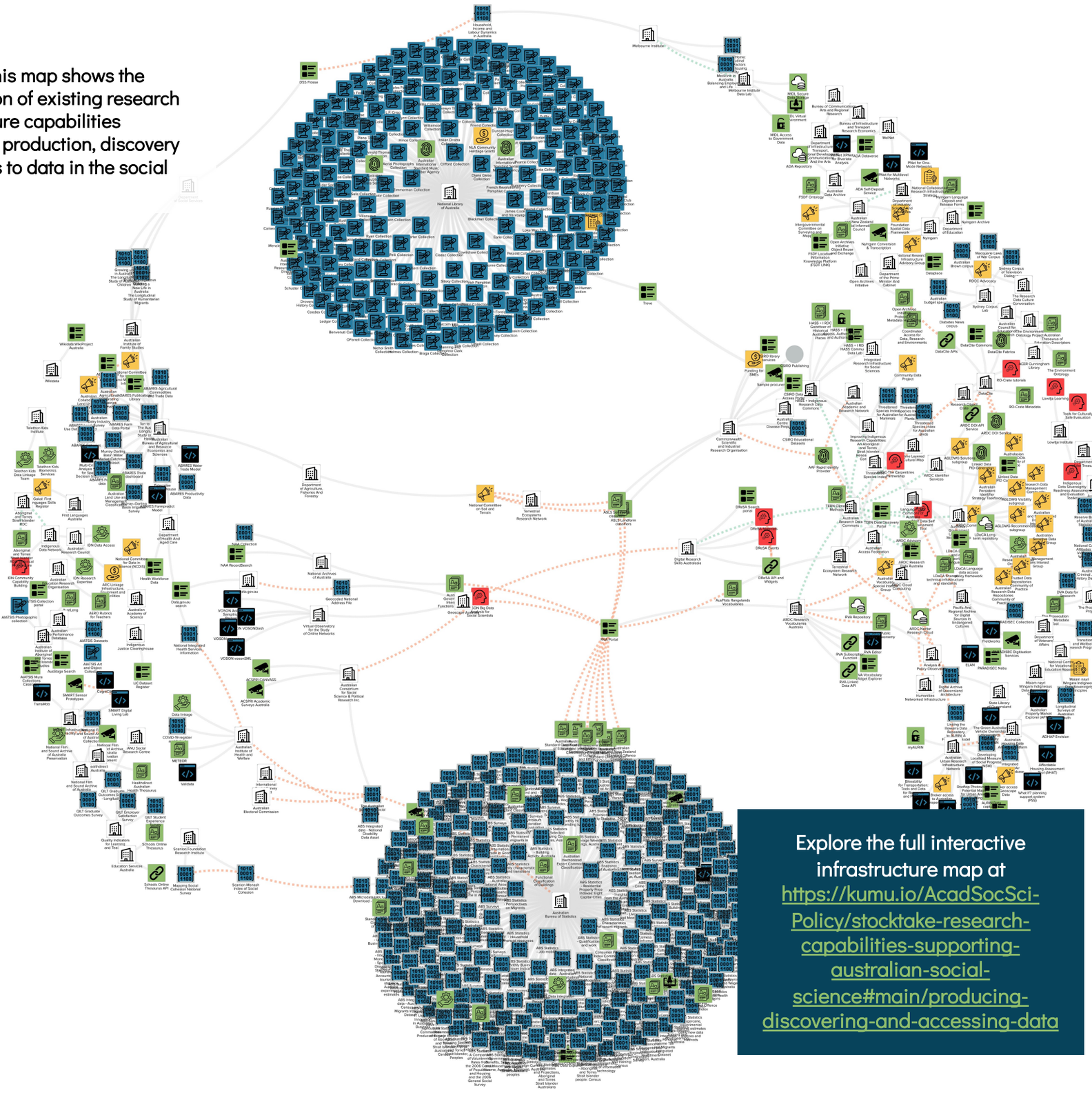


Figure 1. This map shows the constellation of existing research infrastructure capabilities supporting production, discovery and access to data in the social sciences.

Explore the full interactive infrastructure map at https://kumu.io/AcadSocSci-Policy/stocktake-research-capabilities-supporting-australian-social-science#main/producing-discovering-and-accessing-data

Assets

An initial stocktake identified 513 curated collections with data assets of interest to the social sciences and available for research use. These included digital and physical assets, maintained by 27 custodian organisations. Most assets (around 90 per cent) are held by two institutions (Australian Bureau of Statistics, National Library of Australia), and the remainder spread at a median of two per organisation.

Availability

An enormous amount of data critical to social science research is yet to be collected or placed under stewardship and, therefore, at risk of being irreversibly damaged or lost.

Also, much of what is already collected has not been made broadly accessible for reuse. Only a small proportion has been appropriately curated, catalogued, digitised (where relevant) and made available through national repositories. Critical out-of-reach data assets include:

- Data produced by the research sector and stored in institutional repositories or on personal devices
- Administrative and archival data held by government agencies at all levels (local, state/territory and federal)
- Private-sector data held for commercial purposes but of enormous value in addressing key research questions.

Figure 2 (next page) provides an indication of the diversity of organisations holding data assets of potential interest to the social sciences.

Linked assets

Datasets that combine data about the same individual across multiple sources (e.g., education, income, health) are an emergent, critical capability in social science research.

In Australia, linked data assets were first pioneered by researchers in the health and medical fields (back with the Western Australian model in 1960s). Nowadays, we have a growing base of linked assets through the Population Health Research Network (PHRN), the Australian Urban Research Infrastructure Network (AURIN) and the Australian Bureau of Statistics (ABS).

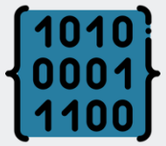
Linked assets can revolutionise social science research, similarly to how human genome data transformed research in biology and healthcare. Unlike human genome data, however, assembling an individual's 'social DNA' requires merging datasets from across multiple custodian organisations, and jumping through enormous hurdles in terms of obtaining approvals for the use of sensitive (human) data.

In order to boldly grasp the opportunities opening for social science research in a digitalised world, the social sciences sector needs to grapple with the problem of efficiently (and safely) linking sensitive data assets.

Indigenous Data

There is an urgent need for improved training of non-Indigenous researchers on appropriate protocols for prioritising and designing research with Aboriginal and Torres Strait Islander communities, and on collecting and using Indigenous data.

In addition, much existing data on Indigenous people and communities is yet to be appropriately transferred to Indigenous ownership



288
Digital collections

225
Physical collections

> The first step is to find out and catalogue what data is available out there. There is much to be unearthed and discovered. For many datasets, surface level listing can be enough, but it is important that at the very least everything is catalogued. That should be the bottom line.
>
> **Prof Marcia Langton AO FASSA FTSE**

> There is significant reuse and strategic research value in the data being generated and maintained in higher education repositories. If we don't inventory it, prioritise it, categorise it, and build the necessary supports around it... we are selling ourselves short.
>
> **Ingrid Mason Consultant**

> At present, we have all these researchers writing endless proposals and data management plans... too many plans. We need processes that are more efficient, and really get a handle on which researcher-generated data could have national significance or should otherwise be deposited into a collection.
>
> **A/Prof Nick Thieberger FAHA**

and control. Such rematriation of Indigenous data is a critical step in the reconciliation journey of the Australian social sciences.

Identifying asset gaps

Stakeholders consulted thus far have suggested two actions to identify data asset gaps:

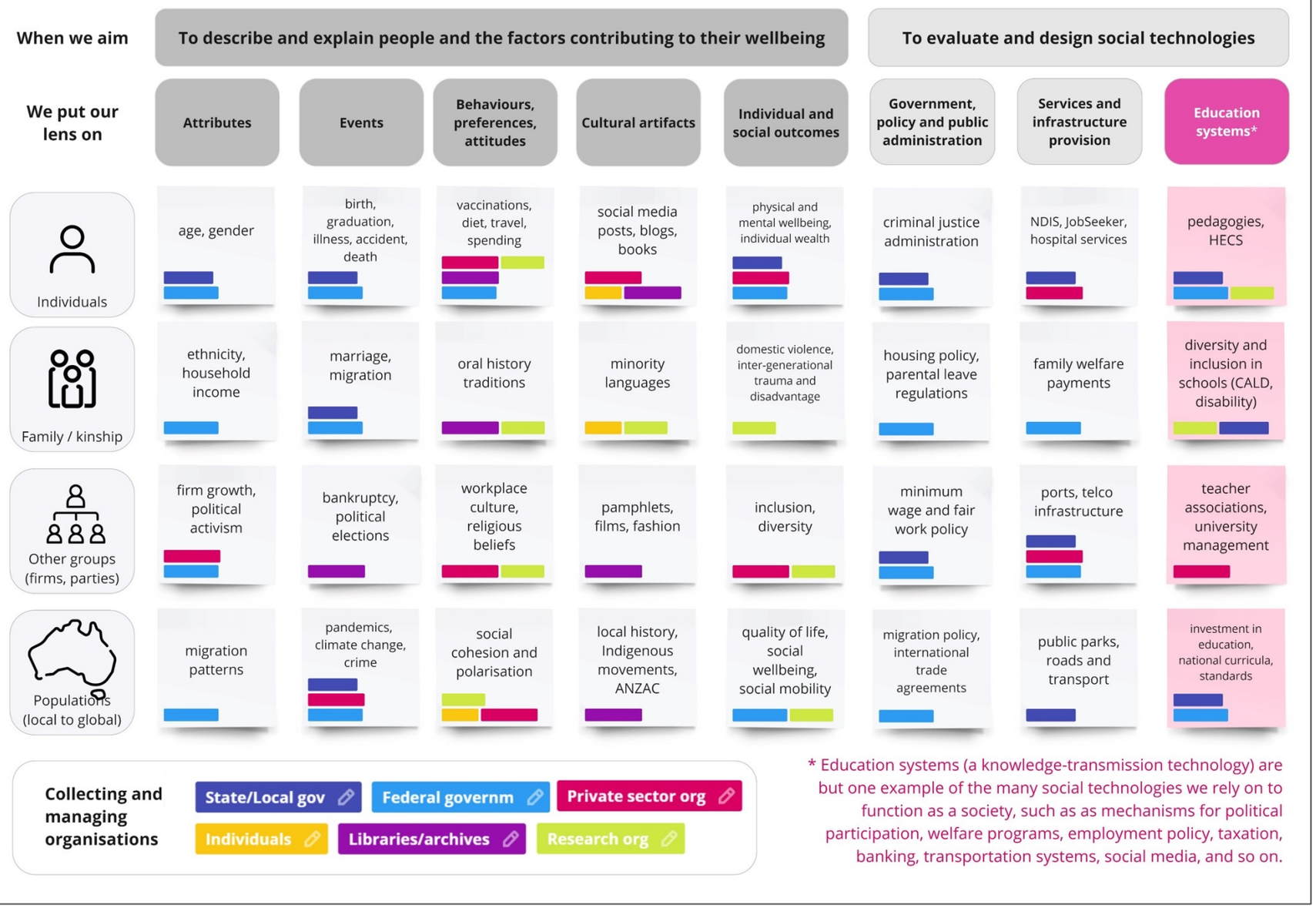
- **Bottom up.** For universities and government agencies (initially) to participate in a national *survey of data assets*. Such a catalogue would provide a basis to identify orphaned assets that should be preserved and shared, including any assets of interest to Aboriginal and Torres Strait Islanders researchers and communities
- **Top down**. For disciplinary societies and associations to develop *collection policies*, or explicit statements about what data are needed to drive research in each discipline over the next decades, and why that research is important for Australia. Such policies become the basis to guide decision-making around collections and the prioritisation of investments.

Any actions would need to consider changing or emerging social phenomena (e.g., the many facets of life in a digital age) and data sources (e.g., sensors, citizen science platforms, data donations).

Ultimately, the social sciences need to collectively trace a path to ensure Australian researchers have the best possible data about Australia (its peoples, history, institutions), as well as access to comparable datasets in other countries. ■

Figure 2. The social sciences: what our data looks like, and where it's found

A sampler of the variety of data of interest to the social sciences, and its collecting and managing organisations



Systems

Collection

The stocktake exercise identified 10 data collection capabilities available to researchers nationally. These include survey infrastructure, preservation and digitisation equipment, web crawling software, sensor prototyping labs, and manuscript conversion and transcription.

Our sector is in the early stages of adopting advanced data acquisition methods, such as Internet of Things (distributed sensors) or monitoring of social media and other online behaviour. The current situation presents the sector with a significant opportunity to actively establish collaborative infrastructure that effectively and efficiently addresses these emerging gaps.

Storage

The identified 5 storage capabilities, comprising cloud storage for specific purposes and applications (e.g., ARDC Nectar Research Cloud).

At present, there is no national, common deposit and storage infrastructure to share data assets produced by the academic sector. Assets sitting in institutional repositories are out of reach to users outside the hosting institution; and the few leaders who set up online platforms to share assets more broadly mostly operate under insecure funding agreements that don't guarantee secure long-term storage.

There's significant appetite for national deposit and storage infrastructure, to share assets produced by the research sector, which guarantees the safety and longevity of nationally significant assets.

Curation and stewardship

Curation and stewardship play a pivotal role in the long-term preservation of data assets, in ways that optimise discoverability, access, and reuse.

The stocktake revealed over 500+ research collections, which owe their discoverability to the active curation and stewardship efforts of 27 organisations, like the Australian Data Archive (ADA), AURIN, the National Library of Australia (NLA), the Pacific and Regional Archive for Digital Sources in Endangered Cultures (PARADISEC) and the Melbourne Institute: Applied Economic & Social Research (MI). These institutions consistently allocate significant resources to effectively preserve and provide access to this valuable assets.

Stakeholders have emphasised the crucial distinction between mere storage and comprehensive curation capabilities. Historically, more attention and funding have been directed toward storage facilities, leading to the accumulation of underutilised data.

To unlock the research value of the vast amount of dormant assets currently in storage, the sector must align its goals with adequate funding and support for curation infrastructure. Only through this concerted effort can we fully leverage the potential of these assets for collaborative, impactful research.

Discovery

The stocktake identified 32 discovery tools, such as online data catalogues or directories.

This abundance of discovery tools implemented by custodian organisations reflects their commitment to facilitating the data discovery process. However, with such vast and diverse custodian agency landscape, finding relevant information is known to be an arduous and time-intensive task. Researchers must frequently rely on their personal networks to identify appropriate sources.

Stakeholders consistently emphasise the need for a centralised discovery platform, that aggregates all data sources of research value, across all disciplines, to maximise asset visibility (and reuse), and facilitates interdisciplinary collaborations.

To address this need, the *Australian Research Data Commons (ARDC)* has established *Research Data Australia*, as a dedicated facility. Substantial efforts are required to ensure seamless integration and collaboration between social science data asset custodians and this platform.

Access management

The stocktake identified 3 access management capabilities, servicing specific platforms and applications.

The *Australian Access Federation (AAF)* was established to streamline access requests within the university sector and could play a major role in streamlining access to shared data assets in the future.

Interoperability

A total of 54 technical standards and tools were identified, including vocabularies, and metadata, PID and API generation support tools. Despite the seemingly large number of components identified, interoperability is still in early stages. Most of the standards and vocabularies identified were created by individual organisations to support specific assets, as opposed to groups of disciplinary experts to link together a *variety* of assets.

The ARDC is specifically dedicated to support the development of sharing acceleration infrastructures, including a suite of projects specifically in the HASS space, or *HASS and Indigenous Research Data Commons (HASS + I RDC)*. Among them, the project *Integrated Research Infrastructure for the Social Sciences (IRISS)* is specifically concerned with developing systems to support data harmonisation and interoperability standards across social science domains.

The Australian research sector has a phenomenal advantage in having this national entity leading, coordinating, advising on, and co-investing in the development of sharing acceleration infrastructures. Any sector-wide aspirations defined by the social sciences sector over the course of the Decadal Plan would benefit from development in consultation, coordination or partnership with the ARDC.

A federated future

In the current systems landscape, many individual organisations operate their own platforms to manage and share data assets. This approach is beneficial in terms of expediency and local control, but can lead to duplicated efforts and high costs for custodian organisations. Meanwhile, researchers seeking to access distributed data face the daunting task of searching for assets across multiple platforms, navigating diverse processes, and transforming heterogeneous datasets to ensure compatibility during aggregation.

A more efficient, federated structure (see Figure 3) could hold the key to overcome these limitations and driving progress in the sector. Under a federated structure, custodian organisations could enable significant new opportunities, including:

- **Digital preservation**: Preserving digital collections is often more expensive than preserving physical ones, contrary to popular belief. The expenses associated with data management security and preservation often exceed the capacities of individual organisations. Under a federated ecosystem, organisations could share these operational costs
- **Staff retention**: Research data management requires highly skilled individuals in specialist roles such as archiving, preservation, and platform development and maintenance
- **Continuous improvement**: Keeping pace with software innovation, evolving standards, and establishing and maintaining global partnerships.

Another challenge stemming from the current fragmented structure is the platform-specific formatting of data assets.

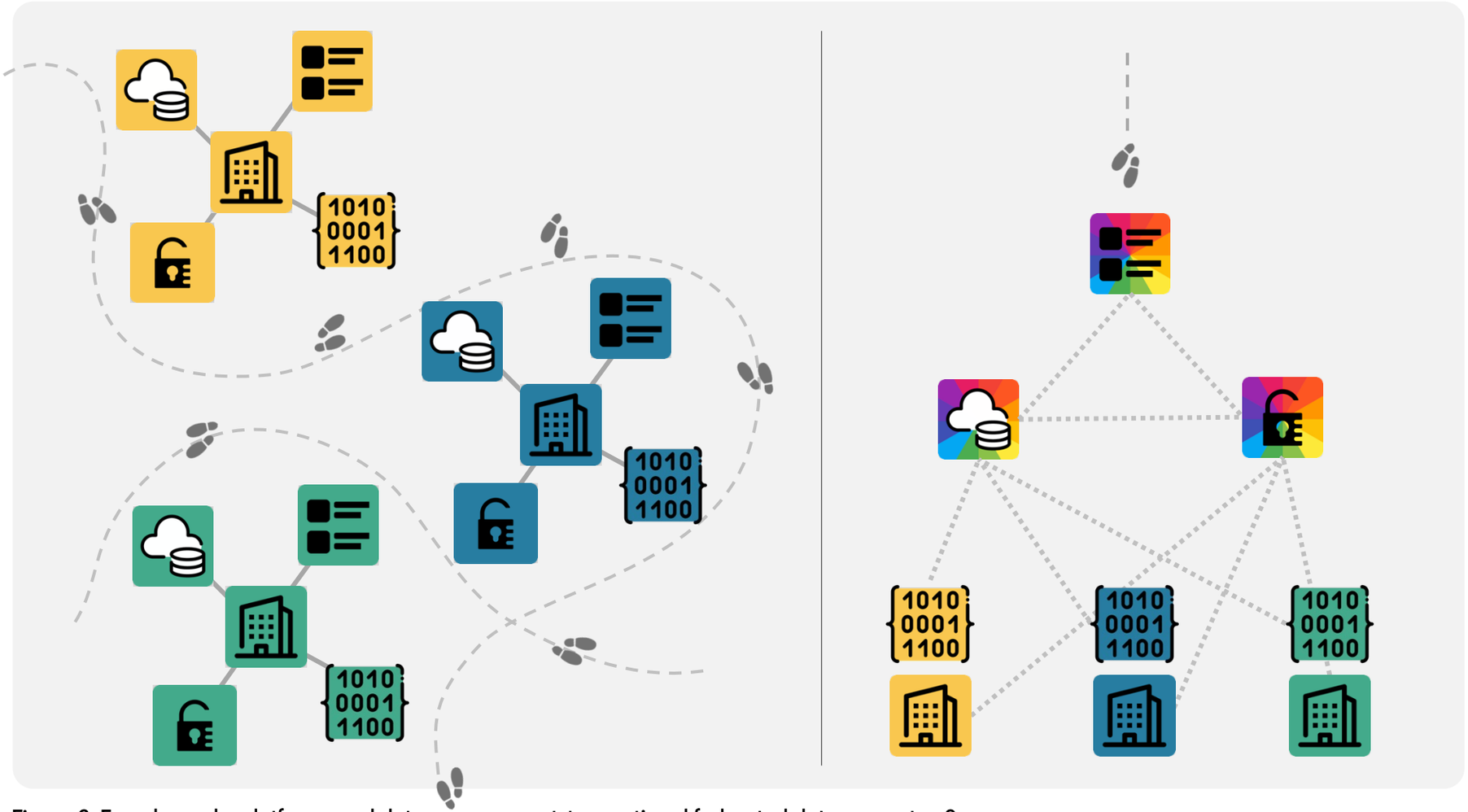


Figure 3. From bespoke platforms and data management, to a national federated data ecosystem?

This poses a significant hurdle in salvaging valuable data when platforms become obsolete. In tandem with transitioning to a federated ecosystem, the sector could prioritise the packaging of data in formats that are independent of specific platforms or technologies. This approach could ensure the resilience of data assets amid the constantly changing landscape of platforms and technologies. ■

Rules

4
Strategic and regulatory policy

7
Funding streams

25
Leadership capabilities

Strategic policy

Multiple sources of strategic policy influence decision-making in relation to data infrastructures:

- **National science and research priorities,** set by the Department of Industry, Science, and Resources, and applied to funding allocations via the National Collaborative Research Infrastructure Strategy (NCRIS) and the Australian Research Council (ARC)
- **Vision and priorities set by the technical community**, through organisations such as the Research Data Alliance or CODATA
- **Community-defined standards**, such as the National Statement on Ethical Conduct in Human Research, the Indigenous Data Sovereignty principles or any priorities defined by individual disciplines.

Effective sectoral leadership will involve defining sector-specific strategic priorities, both through initiatives like this Decadal Plan and future plans tailored to individual disciplines; and advocating for their inclusion in the higher-level strategic forums identified above, so that infrastructure-shaping policies align with the sector's interests and needs.

Funding

We have identified three main funding streams for data-related capabilities:

- **NCRIS**. Federal government funding specifically for research infrastructure, non-competitive, and uncapped for individual projects. This stream supports two social science capabilities, AURIN and PHRN, as well as the ARDC
- **ARDC.** NCRIS-funded, it itself acts as a funding body, co-investing into (and incubating) infrastructure-building projects. In the latest NCRIS round (2021-23), ARDC co-invested 8.9 million into eight HASS-specific projects
- **Research councils (ARC, NHMRC)**. These agencies support infrastructure development directly through short-term grants (i.e., LIEF) and, indirectly, through regular research grants, where capabilities emerge to support, or as spillovers of research projects.

While concrete data on funding arrangements for mapped capabilities is still pending, previous discussions indicate that most infrastructures operate through institutional partnerships, established for mutual benefit and a desire to share resources with the research community (usually free of charge). But these partnerships have downsides. They tend to be short-term, susceptible to changes in contributors' capacities, and the limited financial resources constrain their potential to grow and innovate.

While federal NCRIS funding remains primarily focused on a limited number of larger-scale facilities, it is crucial to devise strategies that promote security and stimulate innovation for the multitude of smaller infrastructures that diversify and enhance our data ecosystem.

Regulatory policy

The regulations impacting data include:

- **Technical.** Those dictating best practice in relation to the archiving, annotation and preservation of assets. These are still emerging (e.g., the regulations to operationalise the FAIR principles), but will be key to maximising productivity and interoperability
- **Data sharing.** Those encouraging producers and custodians to share data with the research sector. ARC grants, for example, require grantees to place data in repositories with open access
- **Privacy and sovereignty.** Those dictating the conditions for safe and ethical ownership and reuse of *human* data in research.

Currently, there is no single data practice guideline unifying all these requirements. Instead, custodian organisations and research teams must individually navigate each layer and devise their own data management plans.

The resulting diversity of practices poses challenges for interoperability. It also negatively impacts accessibility to sensitive human data, because in absence of a standard of practice, custodians often respond by adopting risk-averse and highly discretionary processes, costly both to custodian agencies and researchers.

Active participation in the development of any emerging standards is critical to ensure they'll meet our specific needs. ■

> "At present, access to MADIP data is approved on a project-by-project basis. This requires data custodians [primarily Commonwealth departments] to assess each request for access to their data. This process places a heavy and increasing burden on data custodians, particularly as the number of MADIP projects is growing at about 30% year-on-year.
>
> **Marcel van Kints**
>
> **Data Strategy and Services, Australian Bureau of Statistics (ABS)**

> "Accessing sensitive health data can be a complicated process for newcomers. Research teams need an ethics approval(s) plus additional approvals from the relevant data custodian organisation(s) before they can access the data. Preparing applications that navigate efficiently through these processes takes skill.
>
> **Dr Merran Smith**
>
> **Population Health Research Network (PHRN)**

10
Training

Skills

An initial stocktake identified 10 skills & training capabilities related to data, specifically, training in archival, annotation, preservation and data management skills. Identified capabilities varied in type, from individual training materials and self-assessment tools, to training directories, to partnerships and other sectoral initiatives to enhance skilling opportunities.

Consultations undertaken by the Academy in 2022 showed social science researchers are keen to bridge the growing technical skills gap but find the emerging skills landscape too complex and hard to navigate. The ARDC is already developing a skills framework to address this issue, but individual disciplines are yet to contribute to the definition of skill development pathways for specific fields.

In terms of the training programs available to researchers to develop data-related skills, a sparsely populated stocktake suggests archival, annotation, preservation and data management skill gaps are being met largely through institutional or individual means, with limited work at the national scale.

General and specialist skill requirements

In responding to skill gaps, the sector must tailor its approaches to cater for at least three distinct needs:

- **Professional staff in data managing organisations**. These organisations employ staff with highly technical archival, preservation and IT development skills; to oversee and manage research infrastructures and collections, ensure best practice, and designing systems and procedures with which researchers can easily comply
- **Data science experts with discipline-specific knowledge**. Usually employed by universities to assist multiple research teams. They are highly adept at handling large-scale datasets (can create code, algorithms and workflows to process and analyse data) and understand the conceptual and methodological intricacies of specific disciplines. They are excellently placed to advocate on behalf of disciplines during the design of national standards or infrastructure
- **Data management skills for researchers across the board**. The skills every researcher should have, to make best use of existing data assets and tools, maximise their productivity and contribute positively to the ecosystem. The current absence of a single, integrated corpus defining best practice makes it very hard for average researchers to navigate what should be a simple set of rules and practices.

Considering about 70-90% of research time is devoted to data management (according to stakeholders and varying across fields), better training infrastructures can be one of the most powerful catalysts to accelerating national research productivity. ■

> “A parallel issue is the shortage of appropriately skilled workforce, specifically, of developers who can support the implementation of the interoperability infrastructure, and of curators and archivists to work with them. The lack of a career pathway for such technically-oriented, data-skilled professionals is possibly behind the insufficient supply of such critical skilled workforce.
>
> **A/Prof Steven McEachern**
> **Australian Data Archive,**
> **Australian National University**

Consultation questions

This Discussion Paper is inviting community input to define the sector’s needs in relation to infrastructure to produce, discover and access data.

Current state

Q1. How would you modify or augment our description of the current state of assets, systems, rules and skills and training?

Your needs

Q2. Can you provide specific examples of data-related challenges your research team faces, where shared infrastructure could significantly boost productivity or support your research aspirations?

Delivering solutions

Q3. Which needs can be met through improvements to existing assets, systems, rules or skills and training? Briefly describe the improvements required.

Q4. Which needs require that the sector advocates for new assets, systems, rules or training? Briefly describe any new infrastructures you think are required, including where possible examples and any requirements for successful implementation (e.g., incentives, funding, partnerships).

The case for prioritising sensitive data and ways forward

Good social science relies on access to high-resolution data about people, institutions, communities and firms. However, in many cases, the data that provides the most relevant insights is highly sensitive; containing personal or identifying information that needs to be protected for ethical and legal reasons.

Professional data management agencies, such as the ABS or PHRN, have developed facilities that can provide secure access to highly-sensitive data, for approved researchers, under strict conditions. But the approval processes remain costly and time-consuming for custodian agencies (which are currently under resourced to meet growing demand from the research sector) as well as for research teams. Approval processes are particularly problematic when dealing with linked data assets, which require multiple approvals from individual custodian agencies.

As a result of these complexities, research on many important social issues is being hindered or delayed. A question for the sector is, therefore, how to better balance the need to protect individual privacy with the need for timely and cost-effective research into, and solutions to, pressing social challenges.

Specifically, what could be realised over the coming years by way of more efficient, standardised mechanisms to procure and link high-resolution data? And how could the sector make these capabilities available nation-wide for social science? Figure 4 presents potential ways forward.

A Decadal Plan for Social Science Research Infrastructure is an opportunity for the sector to reinvigorate the national conversation on sensitive data; a conversation the social sciences have both high stakes in and the right expertise to lead. ■

Figure 4. Streamlining researcher access to sensitive human data: key considerations for progress

Make Indigenous Data Sovereignty the gold standard for handling all human data.

Eliminate double standards while setting the bar high for everyone.

Sort licensing at the time of collection or deposit. Many assets remain unutilised because licensing wasn't set up at the time of collection and obtaining the required permits is impractical.

Trust professional, certified data agencies to eliminate or minimise the risk of reidentification, manageable with current technology, under the right hands (e.g., ABS, PHRN, MIDL).

Assist custodians in determining lawful uses of human data. Provide clear guidelines to evaluate whether research questions and methodologies align with agency mandates and community license.

Review privacy regulations in light of evolving social license aspirations and new technologies.

Where do we need increased stringency or flexibility?

Develop responsible AI guidelines and regulations. Which uses of artificial intelligence over large human datasets can be lawful/unlawful? A question for experts and the entire nation.

2 | Analysing data to generate new knowledge

This section examines the current state of infrastructure supporting the analysis of data (Figure 5); on the assumption that shared infrastructure can significantly increase research impact, quality and productivity.

Collective benefits

Research impact

- Access to state-of-the-art data analysis capabilities, so the sector can maintain and expand their position as global leaders in research excellence, and provide high quality evidence to inform policy, practice and services
- Social science researchers are supported to develop and access innovative research software to meet their evolving needs.

Research quality

- Access to high-performance computing (HPC) to perform complex analyses over large volumes of data
- Improve skills and competencies in the use of digital research tools for data analysis at scale and with increased productivity (e.g., scripting routine tasks, data wrangling, machine learning).

Research productivity

- Fully automate or incorporate computer assistance to routine data transformation, analysis or visualisation tasks, to efficiently scale up the sector's analytical capabilities. ■

Explore the full interactive infrastructure map at https://kumu.io/AcadSocSci-Policy/stocktake-research-capabilities-supporting-australian-social-science#main/analysing-data-to-generate-new-knowledge

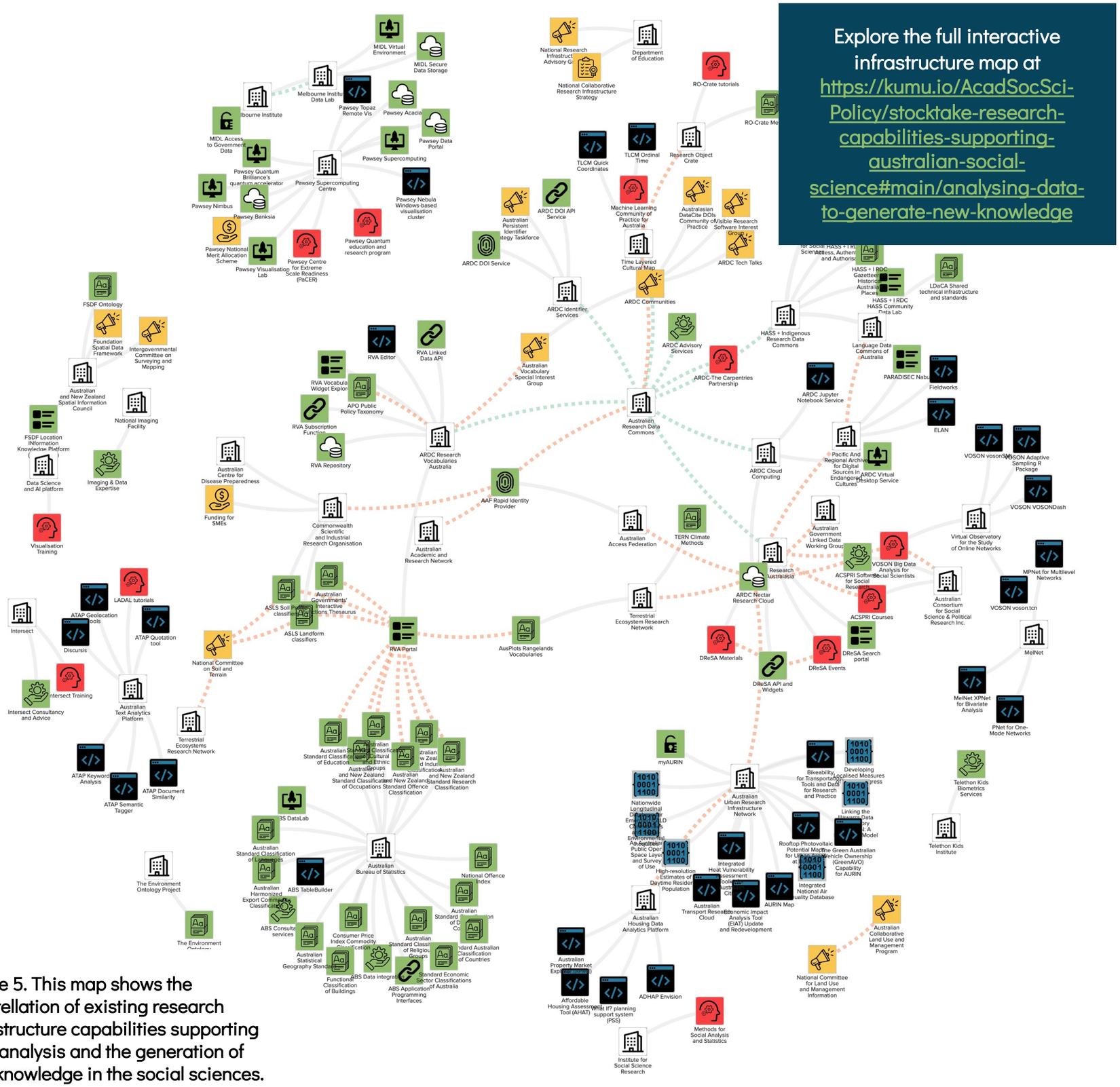


Figure 5. This map shows the constellation of existing research infrastructure capabilities supporting data analysis and the generation of new knowledge in the social sciences.

42
Analysis tools

Assets

An initial stocktake identified 42 capabilities for analysing, transforming and visualising data, from 14 organisations, and comprising a mix of standalone software (for download) and tools accessible via online platforms. It also included at least one asset for the sharing of code and workflows (ARDC's Jupyter Notebook Service).

Identifying asset gaps

An assessment of the robustness of current stock, in terms of how well it supports contemporary and emerging disciplinary and societal needs, is pending. This Discussion Paper welcomes input from stakeholders to better understand any critical gaps.

Looking ahead, the sector would benefit from **identifying existing but currently underutilised assets** which, more effectively deployed, could elevate national capability and productivity, for example:

- **Discipline-specific modelling applications,** such as climate, economic impact, or walkability models; built by specialists in a given field (e.g., climate), but which can support a myriad of multidisciplinary inquiries
- **Tools that support complexity in research,** such as the use of geospatial, network analysis, and machine learning methodologies
- **Lesser-known, smarter ways to work.** A good are example are the many tools developed over decades by technically-skilled linguists, to handle and interrogate large corpuses of text, and which could, nowadays, elevate the technical efficiency of any disciplines working with text.

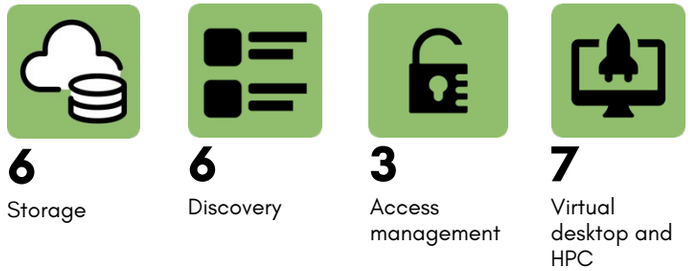
Qualitative, at scale

Stakeholders have pointed the social sciences are at a turning point, where qualitative research is starting to truly realise the benefits of computer-assistance and automation, traditionally enjoyed by the more natively-quantitative fields (e.g., economics, statistics). Qualitative-kind tasks, such as transcriptions, descriptions or tagging are increasingly being facilitated by an influx of innovations in image and audio recognition, semantic analysis, machine learning and others.

By swiftly harnessing these technologies, the sector can not only stay competitive (how else will we handle the massive wave of digital-life data?) but achieve unprecedented feats in terms of the scale and sophistication of qualitative-oriented research applications.

Next frontier tools

One of our leading NCRIS capabilities, AURIN, has set a goal to build truly functional twin city capabilities for Australia over the next decade, enabling visualisations and predictive modelling of energy, climate and demographic trends. What other visionary analysis capabilities should the social sciences be planning for? ■



6
Storage

6
Discovery

3
Access management

7
Virtual desktop and HPC

Systems

The stocktake identified six storage, six discovery, and three access management capabilities that support the deployment and sharing of analytical tools. Additionally, there are seven remote or virtual desktop and High-Performance Computing (HPC) capabilities available to researchers nationwide.

Visibility

Like the data asset landscape, the tool landscape is also largely fragmented (tools available through different organisations, mostly disconnected from one another). This could potentially hinder tool discovery by researchers, due to low awareness of their availability, leading to unrealised productivity gains and demand for these tools, and ultimately discouraging greater investment in their continued development. A centralised discovery point for social science analysis tools is currently lacking.

Interoperability

Building on the previous section (*Producing, discovering and accessing data*), progress towards standardisation of data practices and interoperability will ultimately increase the usability of available tools (provided those tools evolve alongside standard practices). In other words, researchers could have a future where they can easily test and move data across tools, thanks to high levels of standardisation and interoperability.

High-performance compute

Out of the seven virtual desktop and HPC capabilities identified, four are specifically HPC. Some stakeholders have expressed concerns that access to HPC infrastructures is limited to a number of projects each year, which could impact on access to social scientists. Anecdotal evidence also suggests research institutions are satisfying increasing demand for these infrastructures through commercial services, such as Amazon Web Services (AWS), as a more cost-effective alternative to building those capacities within each institution.

This Discussion Paper welcomes input from the research and technical infrastructure communities that helps understand the suitability of the systems supporting the use and circulation of analytical assets. ■

4
Strategic and regulatory policy

7
Funding streams

25
Leadership capabilities

Rules

Strategic policy

At present, in Australia, the strategic policy framework underpinning federal decision-making and funding allocations for research infrastructure are the same for *data* and *analysis* infrastructures.

These two types of capabilities have **markedly different funding requirements** (data infrastructure relies more heavily on operational expenses; while analytical infrastructure tilts towards capital) **and lifespans** (in the order of 50–200 years for data assets; compared to 5–20 years for analysis capabilities).

It is unknown whether the lack of differentiated strategic policies could be negatively impacting the sector.

On the *data infrastructure* side, stakeholders have expressed concerns that fundamental data-supporting infrastructure may not be perceived as worthy of investment as analytical assets (e.g., high-performance computing).

On the *analysis infrastructure* side, there seems to be appetite for seed funding and incubation support to get innovative, early-stage capabilities off the ground. Such incubation support might include assistance towards the brokering of alliances and partnerships that see facilities develop into sustainable enterprises (e.g., industry partnerships, institutional consortia).

Turning spillovers into advantages

Similar to data infrastructure, many analytical capabilities are indirectly funded through research work and grants. It is a common trajectory for new analysis capabilities to emerge as spillovers from talent clusters in research centres, which are later sustained and amplified through institutional funding or partnerships.

Tools developed in this manner and made freely accessible to researchers have the potential to contribute enormous value to the research ecosystem. However, ensuring their sustained quality, level of service, and longevity depends on ongoing financial support.

The sector needs a systematic approach to identify and appropriately support any existing and emerging high-value analytical capabilities.

Regulating AI

Lastly, the sector must proactively develop a framework to regulate analysis capabilities that apply machine learning and artificial intelligence methodologies to human data, or that are used to assist in decision-making processes impacting individuals and societies.

Social science researchers could also play a critical role working with and advising government, civil society and industry on responsible AI. ■

12
Training

Skills

An initial stocktake identified 12 skills and training capabilities specifically to support research analysis, and ranging from training organisations, to communities of practice, to summer schools and similar events.

Training gaps

Consultations undertaken by the Academy in 2022 showed social science researchers are keen to acquire skills in emerging methodologies, such as **machine learning**, but find the skills landscape hard to navigate.

In addition to contributing to discipline-specific skills frameworks and learning pathways (discussed in previous section), the social sciences are well-positioned to critically contribute to the definition of any national approaches to workforce skill development.

Amidst a growing supply of online training options, for example, the university-consortium national non-profit *Intersect*, a leader in the eResearch training space, advocates for training delivered on site, in the context of real research projects and supported by person-to-person mentoring, as the best way to build research-level technical acumen.

The sector has an interest and expertise to progress this national conversation. ■

Consultation questions

This Discussion Paper is inviting community input to define the sector's needs in relation to infrastructure to analyse data and generate new knowledge.

Current state

Q5. How would you modify or augment our description of the current state of assets, systems, rules and skills?

Your needs

Q6. Can you provide specific examples of data-related challenges your research team faces, where shared infrastructure could significantly boost productivity or support your research aspirations?

Delivering solutions

Q7. Which needs can be met through **improvements to existing** assets, systems, rules or training? Briefly describe the improvements required.

Q8. Which needs require that the sector **advocates for new** assets, systems, rules or training? Briefly describe the required new infrastructures, including where possible, any requirements for successful implementation (e.g., incentives, funding, partnerships).

3 | Brokering high-value partnerships for innovation

This final section explores the question of what shared infrastructures could support social science research innovations and impact at a broad scale.

University researchers dedicate a portion of their time to engagement and impact activities, sometimes with institutional supports, such as science communication training. Yet, the effectiveness of these efforts is highly dependent on whether the researcher can deliver his/her pitch to the right partner, be it a government agency, news media outlet, community or industry organisation. The same difficulty applies to potential partners needing to locate the right researchers, teams or projects.

The brokering of high-value partnerships, in the sense of finding and reaching out to the right organisation, with the right challenge or solution, at the right time, can exceed the means and abilities of individual research teams or institutions and, in turn, make a good case for nationwide collaborative infrastructure. Some of the potential collective benefits are described next.

Collective benefits

Research impact

- Government, industry and community can easily locate appropriate research partners
- Social scientists can target specific government, industry and community agencies where their knowledge could make greatest positive impact.

Innovation quality

- Researchers have access to social-science specific R&D support, including patent, product or platform development, business incubation, and the brokerage of for-purpose or for-profit partnerships.

Research productivity

- Government, industry and community can quickly/easily access well synthesised social science knowledge, appropriate for their needs.

Existing and emerging capabilities

The stocktake identified a few existing and emerging capabilities in this space:

- **Research-Industry partnerships to deliver data or analysis infrastructure,** such as the various flagship projects by the *Australian Urban Infrastructure Network (AURIN)* or the *Australian Housing Data Analytics Platform (AHDAP)* –all developed as strategic partnerships with government or industry. ARDC's *Bushfire Data Challenges* is another good example. These infrastructures initiate as focused collaborations with the right industry or government partners, and can later be scaled-up for national impact
- **Researcher directories**. *Research Link Australia*, under development by the ARDC, will operate as an industry-oriented directory of research talent
- **Research marketplaces**. Some of the innovations produced by CSIRO's research teams (for example, a new data analysis model with potential industry applications) are advertised through *AWS Marketplace*. Beyond the goal of commercialising a specific innovation, publicly displaying industry-ready innovations has become an avenue to attract industry partners in highly-relevant niches
- **Knowledge synthesis capabilities**. Prospective partners outside the research sector face challenges in accessing scientific knowledge directly relevant to their needs. Research publisher paywalls, the overwhelming volume of published research, and highly technical language are among the barriers they encounter. Could emerging technologies like machine learning and large language models (e.g., ChatGPT) be harnessed to facilitate these knowledge synthesis and discovery tasks? (e.g., sift through vast amounts of literature, identify crucial insights, and present them in a more accessible manner).

Specialised support

An additional issue to consider is the availability of dedicated R&D, patent development, IT platform development and enterprise incubation **specialised in social science applications**.

During the consultations for the State of the Social Sciences 2021, stakeholders described limitations in this area, such as a shortage of university R&D support staff specialised in social innovations (see *Innovation our way*, overleaf); and the fact that limited institutional resourcing often meant innovation support went to projects with the highest immediate commercial value (usually STEM and Medical).

This Discussion Paper welcomes input from the research community, government, think tanks and media, industry and non-profit sectors, that informs the kinds of infrastructural improvements that could propel social science innovations in the next decade. ■

Consultation questions

This Discussion Paper is inviting community input to define the sector's needs in relation to infrastructure to broker high-value partnerships for innovation.

Current state

Q9. How would you modify or augment our description of the current state of innovation supporting infrastructures?

Your needs

Q10. Can you provide specific examples of innovation-related challenges your research team faces, where shared infrastructure could significantly boost productivity or support your research aspirations?

Delivering solutions

Q11. What national approaches could facilitate the brokering of high-value partnerships **with government**?

Q12. What national approaches could facilitate the brokering of high-value partnerships **with industry?**

Q13. What national approaches could facilitate the brokering of high-value partnerships with **the community sector?**

Innovation our way: Social technologies in everyday life

Social technologies are the backbone of our societies. What national-level supports could pave the way for stronger collaborations between social science researchers, industry, communities, and government?

While technological innovations are commonly associated with STEM disciplines, the technology-oriented disciplines of the social sciences have boldly spearheaded societal progress through a multitude of system-defining (though often invisible) innovations (see Figure 6).

Researchers in technology-oriented disciplines leverage knowledge from *primary* research areas like economics, psychology, political science, or sociology, and apply them to create *new* or *improved* ways to educate, legislate, plan, incentivise, trade, pay, assist, distribute, and the many other mechanisms through which social innovation takes place. Some social innovations generate tangible monetary gains, while others are best measured through increased social wellbeing, efficiency, effectiveness, and equity.

As we explore the sector's need for infrastructure over the next decade, we invite readers to reflect on the unique needs and opportunities in their respective disciplines. How can collaboration between researchers, industry, communities, and government be further strengthened to unlock the full potential of social technologies?

The *Decadal Plan for Social Science Research Infrastructure 2023–32* is an opportunity to chart a course toward a more interconnected, sustainable, and thriving Australia. ■

Figure 6. Social technologies and related innovations

- Education
- Law and justice
- Policy and public administration
- Management
- Marketing
- Tourism
- Normative economics
- Public Health
- Urban and Regional Planning
- Social Work
- Media and communications.


ACADEMY OF
THE SOCIAL SCIENCES
IN AUSTRALIA