ACADEMY OF THE SOCIAL SCIENCES IN AUSTRALIA

AUSTRALIA'S DATA-ENABLED RESEARCH FUTURE: THE SOCIAL SCIENCES

A collaboration between ARDC, ACOLA and Australia's Five Learned Academies

Sponsored by

ARDC

Australian Research Data Commons

About this report

This project is the result of a partnership between the ARDC, Australia's five Learned Academies and ACOLA to ensure Australia can undertake excellent data-enabled research across all fields of research. Notably, the project sought to help build a more coherent data policy and strategic data planning environment to uplift national data infrastructure. Five domain reports were developed, and a synthesis report focused on common themes and multidisciplinary opportunities and needs. We hope that this project will transition into an ongoing national data policy and strategic planning capability.

Report prepared by Dr Chris Hatherly, Dr Isabel Ceron and Callum Jones.

This project received investment from the Australian Research Data Commons (ARDC). The ARDC is supported by the National Collaborative Research Infrastructure Strategy (NCRIS).

Date of publication
June 2022

Publisher
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 (2022) Australia's Data Enabled Research Future: The Social Sciences.

ISBN: 978-0-908290-13-0

Contents

1. Executive Summary....3
2. About this report....5
   2.1. Goal....5
   2.2. Guiding questions for the environmental scan....5
3. Structure and methods....6
4. Aspirations for a social science research data ecosystem....7
   4.1. The four pillars of a social science research data ecosystem: workforce, governance, tools and data....9
5. Where are we now?....11
   5.1. Workforce....11
   5.2. Governance....12
   5.3. Tools....15
   5.4. Data....16
6. Interim priorities....17
   6.1. The role of the Academy....18
7. In their own words....19
8. Appendix A – Consultation details....21
9. Appendix B – Project team....25
10. Appendix C – Survey results....26

1. Executive Summary

This report was prepared by the *Academy of the Social Sciences in Australia* as part of a series of disciplinary reports on *Australia's data-enabled research future* produced by Australia's Learned Academies in partnership with the Australian Research Data Commons (ARDC).

The aim of this report is to identify the current state, opportunities, and priorities for the development Australia's social science research data ecosystem. This ecosystem includes: (i) large-scale data storage and computation infrastructure, analytic tools and interoperability protocols – referred to collectively as **Research Data Commons**; and (ii) the **workforce** and **governance arrangements** required to enable research data and infrastructure to be effectively used.

Importantly, this reports builds on the ARDC's recent Humanities, Social Sciences and Arts Research Data Commons report (2020) incorporating the perspectives of over 160 social science stakeholders who participated in roundtables or completed a survey for this project. It focuses particularly on:

1. What a fully-developed research data ecosystem could deliver for the social sciences and to Australians more broadly **(Section 4**: *Aspirations for a social science research data ecosystem*)
2. A range of issues that need to be addressed in order for Research Data Commons to operate effectively in the social sciences (**Section 5**: *Where are we now?*)
3. Priorities for investment over the next 5 to 10 years (**Section 6**: *Interim priorities*).

Aspirations. Stakeholders were clear that there is enormous potential to increase the capacity and value of social science research through a more effective and integrated research data ecosystem. Specifically, they noted opportunities to:

- More effectively apply the best available data and research tools to address pressing social problems
- Embed *social systems thinking* into research and policy analysis
- Ensure socially-responsible use of research data commons
- Give value back to communities through an increase in collaborative research and provision of data-enabled, community-oriented online services and applications.

Current challenges. Much of the data ecosystem envisaged in this report is linked to the deployment of Research Data Commons that connect large and diverse datasets with powerful computational and analytic tools and robust protocols. However, the development and implementation of large-scale Research Data Commons is still at an early stage in the Australian social sciences and stakeholders identified the issues needing to be addressed as follows:

- *Research Data Infrastructure.* Adapting traditional qualitative research methods to work with big data and related capabilities; increased funding to build advanced social research capabilities (and underpinning infrastructures) and the rapid deployment of the *Data Availability and Transparency Act 2022*. Stakeholders also noted the need to prepare the sector (our people and facilities) to continue to produce value in a future characterised by data volume rather than scarcity.
- *Workforce.* A need to increase awareness and understanding of the use and benefits of Research Data Commons infrastructure; embed data literacy into university curricula and researcher training across all social science disciplines; and introduce better mechanisms and incentives to utilise big data in social research, policy, and translation.
- *Governance.* Ensuring effective policies and protocols for access, privacy, ethics, and stewardship of research data, as well as necessary funding mechanisms. This includes governance and sovereignty issues that need to be acknowledged and addressed with respect to Indigenous data.

Priorities. The report concludes with seven priorities for action:

(1) Developing a Decadal Plan for social science research infrastructure

(2) Ensuring appropriate funding for Indigenous data commons and governance

(3) Advocating for increased investment in advanced research capabilities for the social sciences

(4) Ensuring greater coverage of social science research priorities in future National Research Infrastructure Roadmap

(5) Scaling up qualitative social science methods for big data

(6) Fast-tracking deployment of the Data Availability and Transparency Act 2022

(7) Future-proofing social science Research Data Commons.

With respect to the first of these priorities, the **Academy of the Social Sciences in Australia** is proposing to co-lead the development of said Decadal Plan in partnership with the ARDC, and to lead and collaborate on other future initiatives, as appropriate.

2. About this report

2.1. Goal

This environmental scan summarises the findings from the *Australian Learned Academies Data Interworking Network* (ALADIN); a capability-building project undertaken by the *Academy of the Social Sciences in Australia* in collaboration with Australia's four other Learned Academies and ACOLA, and with co-investment and technical guidance from the ARDC. The goal of ALADIN was to build the capability of Australia's Learned Academies through:

(i) support and development for Australia's Learned Academies to provide leadership, advocacy and planning for data-enabled research
(ii) sharing of approaches among Academies for a coherent data policy and planning environment
(iii) identifying opportunities to align national research data infrastructure with the strategic needs of Academy research domains
(iv) identifying opportunities for joined-up Academy agendas in support of data-enabled research
(v) establishing a network of data policy and planning capabilities across Academies.

2.2. Guiding questions for the environmental scan

The scope of enquiry of this exercise was defined by the following questions:

- What kind of data-enabled research future should the social science aspire to? (**Section 4**: *Aspirations for a social science research data ecosystem*)
- Where do the social sciences currently stand with respect to data-enabled workforce, governance, tools and data? (**Section 5**: *Where are we now?*)
- What are the key gaps and priorities for the social science research data ecosystem (**Section 6**: *Interim priorities*).

What are the social sciences?

Social science is research on and knowledge about society: its institutions and structures, its histories, and its people. Social science disciplines use systematic methods to understand, describe, educate, predict, and influence the social world, but **their defining feature is a focus on the 'social'**. 'Social', in this context, refers to events, objects, rules, patterns, and other things that emerge spontaneously or by design when humans interact in groups. Think friendship, families, religion, language, politics, schools and hospitals, legislation, markets, armed forces, and elections. Or traffic rules, slavery, poverty, crime, corruption, cooperation, justice, homelessness, activism, social media, consumerism, and so on. A social system can be as small as a family or a football team, and as big as a group of nations.

The current list of social science disciplines recognised by the Academy include Anthropology; Communication and Media Studies; Criminology; Cultural Studies; Demography; Design; Development Studies; Economics; Education; Gender Studies; History; Human geography; Indigenous Studies; Law and Legal Studies; Linguistics; Management, Commerce, Tourism and Services; Policy and Administration; Political Science; Psychology; Public Health; Social Work; Sociology; Statistics; and Urban and Regional Planning.

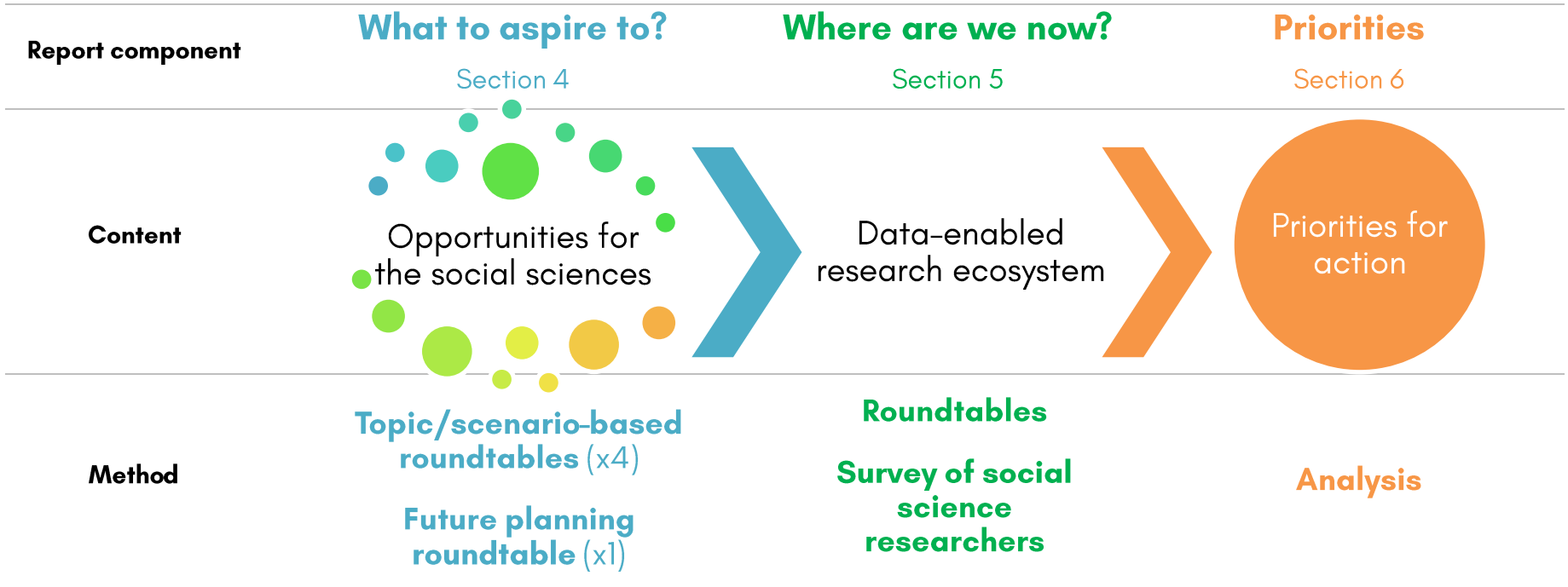
From State of the Social Sciences 2021

3. Structure and methods

This environmental scan is structured as a gap analysis, where identified gaps provide a basis to identify priority actions.

The information presented in **Sections 4** and **5** about the potential future state and current challenges are based on stakeholder consultation undertaken between 2 February and 5 April 2022. Consultation involved 162 social science stakeholders from across the government, academic, and community sectors. 39 of these stakeholders attended one of four scenario-based online roundtables, 29 attended a future planning roundtable, and 94 responded to an online survey. Further details about the roundtables and survey are provided in **Appendix A – Consultation details**.

The project was overseen by an expert steering group, and the consultation and drafting were undertaken by Academy staff with ongoing support of a senior policy advisor from the ARDC. Team members' details are provided in **Appendix B – Project team**.



4. Aspirations for a social science research data ecosystem

What could or should the social sciences aspire to with respect to the research data ecosystem within which we operate? That is, what infrastructure, tools, training and policies would support current and emerging best-practice and innovation in data collection, storage and stewardship, data sharing, analysis and insights, and – perhaps most importantly – the application of research to address real-world social issues? Stakeholders consulted for this report identified four major opportunities for the future:

- Applying the best available data tools and technologies to answer important questions and solve real-world issues.
- Embedding social systems thinking more firmly into research and policy domains.
- Ensuring socially responsible use of research data commons.
- Creating value for communities.

Applying the best available data tools and technologies

Research data tools and technologies have evolved rapidly over recent years. Enhanced capacity for data collection, storage, access and linkage have been complemented by innovation in quantitative and qualitative methods and 'big data' technologies (including machine learning and artificial intelligence) to create opportunities to undertake social research that would have been impractical if not impossible in previous decades.

"It's time we move away from anecdote, into the real evidence of what's happening, at scale."

Specifically, researchers today have unprecedented capacity – at least in theory – to address important research questions and provide insights and solutions to end users in more effective and efficient ways.

There are a number of challenges that will need to be overcome to realise these benefits, including training and workforce development, data access and linkage, and the need for participatory and collaborative engagement with research stakeholders, including communities. In addition, there will be a need in some social science disciplines to overcome a small-team culture that worked well in previous decades, but which can create barriers to large-scale collaborative and interdisciplinary research.

"The social sciences are called to help in the framing of problems of all kinds. It's about understanding problems well before jumping on to solutions and avoiding unintended consequences."

Much of this could and should be achieved over the coming decades through the development and rollout of a future-ready Research Data Commons infrastructure that connects new and existing large datasets (public, scholarly and privately-held) to high-performance computing and analytics tools supported by robust policies and protocols as well as rigorous training and support of data specialists.

With such a data commons in place, the social sciences will be well prepared to provide new insights and solutions to the social, economic, environmental and human questions and challenges of the future.

Embedding social systems thinking more firmly into research and policy domains

Social systems thinking refers to the integration of social dimensions of an issue (i.e., the *demographic*, *economic*, *political*, and *cultural* factors) into analysis, problem-solving or decision-making processes. The transition to electric vehicles, for example, is a technological approach to solving an environmental problem, which can only succeed if policies appropriately consider and incorporate the human and social components of the problem (e.g., vehicle costs, willingness to pay, attitudes, and beliefs around car ownership and environmental problems, drivers of behaviour, and so on).

Data commons infrastructure provides an ideal environment for social systems thinking: capable of handling large amounts of data from potentially disparate sources or domains (e.g., health, finance, housing, education, infrastructure, environment), and able to link and potentially simulate the dynamics of real-life social systems (e.g., families, firms, regions). In this context, a social science ecosystem that is well supported by data commons could add considerable value to society. More specifically, it could lead to better framing of the social components in research and policy problems in a variety of domains (e.g., technology, health), and to solutions that appropriately consider and draw on existing and well-understood social structures and phenomena (e.g., beliefs, behaviour, social capital).

"An emerging opportunity in the food and nutrition space is to reward precise, personalised data with tailored diets. How can we empower individuals to take charge of their health trajectories, while dealing with those chunky challenges around privacy and data protection? The capabilities are here now, but we don't yet have the constructs to make the best use of it. All sorts of things are coming our way, really big social science questions, in my view. How will we resolve that boundary between privacy and the power of personalised information?"

Ensuring socially responsible use of data commons

Social science expertise is needed to inform the safe, ethical and socially acceptable design and use of data commons systems, processes, and policies. This is particularly important in the context of new technologies at the core of data commons infrastructure (big data, artificial intelligence). There are specific issues that need to be considered and protocols that need to be adopted in the context of data commons that connect to or access Indigenous data in any form.

Creating value for communities

Lastly, data commons infrastructure could support critical value-add initiatives for communities, including:

- Collective memory and cultural heritage
- Public record and accountability
    - Big-picture analyses (current affairs, policy)
    - Evidence-based futures
- Individual wellbeing
    - Personalised advice and information.

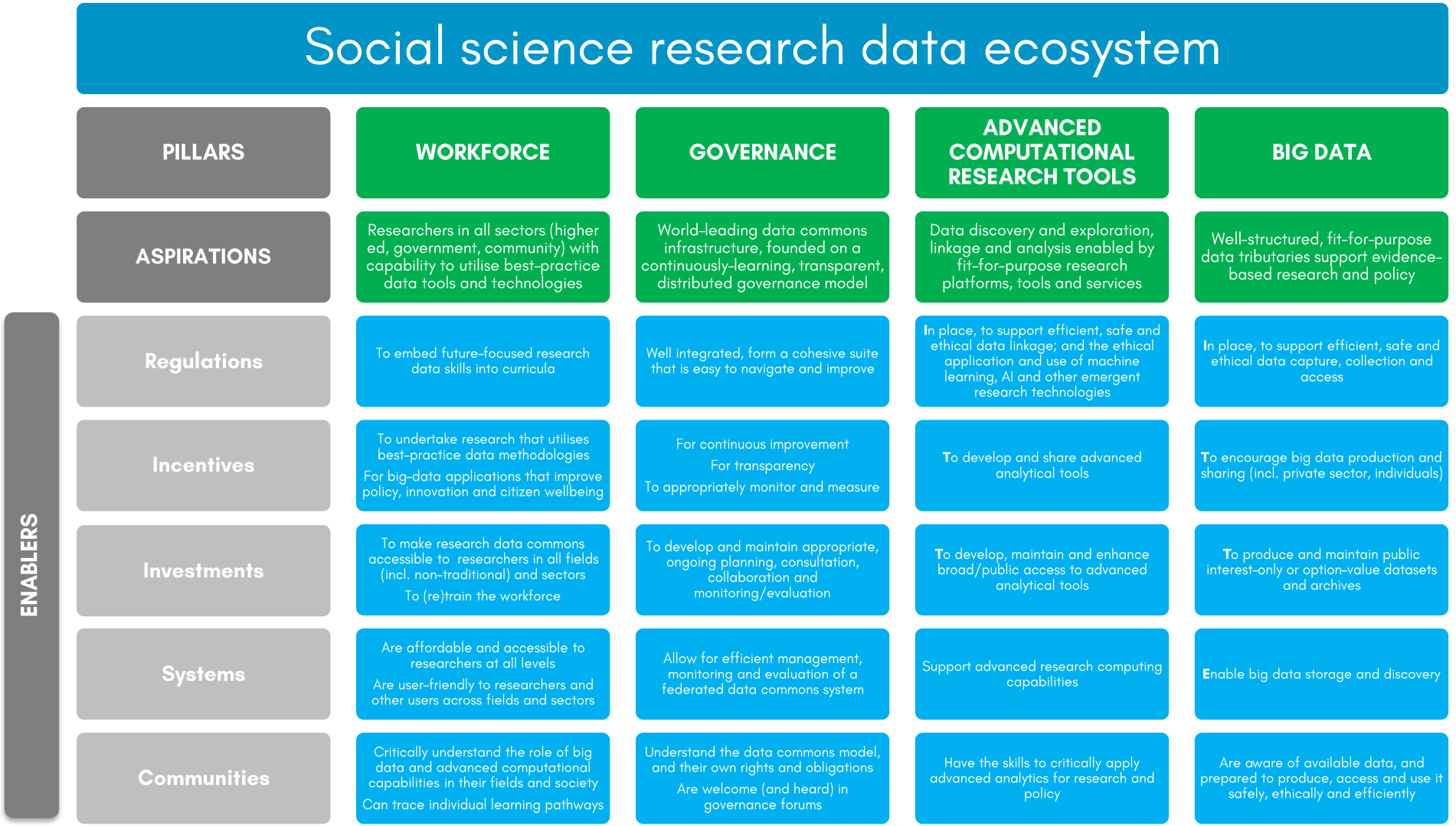
4.1. The four pillars of a social science research data ecosystem: workforce, governance, tools and data

A fully-developed research data ecosystem in the social sciences will need to be built on policies, capacity, and resources in four key domains:

- **An appropriately trained and incentivised research workforce**, including:
  - Future-focused undergraduate and postgraduate university training in data science tools and methodologies, including data ethics and governance, participative and collaborative research with stakeholders, predictive models, machine learning and AI, aimed at ensuring a generally consistent baseline of data literacy among social science graduates across disciplines;
  - researcher training and professional development opportunities that support the acquisition of new data knowledge and skills across the existing research workforce;
  - incentives to conduct social science research in ways that contribute (back) to the data commons, the research community and broader society (e.g., incentives to produce FAIR datasets, to share newly developed workflows or software, to develop applications that benefit the community).
- **Appropriate governance models**, including:
  - support for federated governance and management of data commons that allow transparent and efficient transactions can be seamlessly run across multiple agencies or organisations;
  - research ethics and data sharing standards and regulations that adopt both FAIR (Findable, Accessible, Interoperable and Reusable) and that include serious, system-wide changes in practice and attitudes with respect to Indigenous data – including the CARE principles (Collective benefit, Authority to control, Responsibility and Ethics);
  - government provision of ongoing funding and support for critical commons infrastructure and investments that prioritise Australia's most vulnerable groups;
  - a research community that understands the data commons model, and their obligations and rights within it.
- **Advanced computational research tools**, including:
  - systems and standards that support data linkage across data sets and research domains;
  - systems that facilitate timely discovery of social problems and are agile/responsive in times of crisis;
  - tools and techniques to take advantage of new and existing supercomputing facilities applied to social science data.
- **Access to big data,** including:
  - improved access to data held by government agencies taking advantage of the recently adopted *Data Availability and Transparency Act* (2022);
  - appropriate incentives and platforms/environments to encourage data-sharing with industry, community organisations and individuals (e.g., citizen science).

Figure 1 (over page) provides a visual summary of these four pillars identified by stakeholders as critical to the social science research data ecosystem, along with the various levels of policy and other instruments that are required to make this vision a reality ('enablers').

Figure 1. What does it take to achieve a functional research data ecosystem? Four pillars and key enablers



5. Where are we now?

This section outlines the current state of social science data commons infrastructure in Australia, as identified through desktop research and reported by the stakeholders participating in the consultation.

5.1. Workforce

Australia has a large and diverse social science research workforce distributed across universities, government research and policy agencies, private sector organisations, and a range of not-for-profit, public- and private-sector cultural, social and research organisations (see **Table 1**).

There is enormous variability within this workforce with respect to research methodologies and training. Much of this is based on disciplinary approaches and curricula, and much is based on changes in research training over time. What is clear is that variation in training, terminology, cultures and capacities across the social sciences can and often does act as a barrier to realisation of a functional social science research data ecosystem.

Deliberate policies and programs designed to accelerate sector awareness, understanding and adoption of research data commons over the coming years will be required

Table 1. Current state: workforce

Enablers	Comments
Goal: Researchers in all sectors (higher ed, gov, community) with capability to utilise best-practice data tools and technologies	
Regulations to embed future-focused research data skills into curricula	Significant variability in research data across disciplines and study programmes
Incentives to undertake research that utilises best-practice data methodologies; and for big-data applications that improve policy, innovation and citizen wellbeing	• Recent Cth government investment ($8.9 million) in a humanities, arts, social sciences and indigenous research data commons (led by ARDC), largely to support the development of system-type enablers • Ongoing support for digitally-enabled social science research through various schemes, including NCRIS capabilities, ARC LIEF and ARDC project grants. These schemes generally limited in scope and scale. • More systemic change would see incentives embedded throughout the research pipeline (from topic selection, to funding, publication, and translation or commercialisation)
Investments to make research data commons accessible to researchers in all fields (incl. non-traditional) and sectors; and to (re)train the workforce	
Systems are affordable and accessible to researchers at all levels (inc. HDRs, EMCRs); and user-friendly to researchers and other users, in all fields and sectors	
Communities critically understand the role of big data and advanced computational capabilities in their fields and society; and can trace individual learning pathways	• The online survey of social science researchers (n=94) indicated low awareness and understanding of the principles and skills surrounding data commons (e.g., FAIR) • Many social science researchers feel overwhelmed by the technological change, and uncertain about where to start their up-skilling journey.

5.2. Governance

The mechanisms needed to govern a social science data commons are taking shape, with key pieces already in place or well advanced, some still missing, and the entire suite yet to come together as a well-integrated, easy-to-navigate whole (see **Table 2**). Key challenges over coming years will include:

- Developing an overarching or umbrella policy, which seamlessly integrates the various pieces into an intelligible whole, and is itself well-integrated into national research infrastructure planning agendas
- Explaining to all relevant stakeholders the governance-critical elements for a federated data commons
- Making the case for appropriate funding of governance bodies and mechanisms, where these cannot simply be absorbed by existing agencies (for example, the governance of Indigenous Data Commons[1]).

Table 2. Current state: governance

Enablers	Comments
Goal: Data commons founded on a continuously-learning, transparent, federated governance model	
Regulations are well integrated, and form a cohesive suite that is easy to navigate and improve	• Key pieces already implemented or well underway, such as: FAIR and CARE principles, Mayam Nayri Wingara Principles, and the Data Availability and Transparency Act 2022. • Some key governance pieces, such as FAIR principles, remain non-binding (e.g., still not a requirement to research or publish), leading to slow adoption or stagnation • A revamp of ethics regulations currently in place in higher-education settings is often referred as urgent, to address data management bottlenecks (access, storage, publication), or to incorporate changing views (e.g., Indigenous data governance) • The priorities recognised in the National Research Infrastructure Roadmap are heavily aligned with technology and industry outcomes, to detriment of more socially-oriented priorities. This effectively limits the opportunities to direct Cth investment into social science research infrastructure.
Investments to develop and maintain appropriate, ongoing planning, consultation, collaboration and monitoring/evaluation	
Incentives for continuous improvement, transparency, and to appropriately monitor and measure performance	[Too early]
Systems allow for efficient management, monitoring and evaluation of a federated data commons system	
Communities understand the federated data commons model, their own rights and obligations; and are welcome (and heard) in governance forums	• Insufficient understanding of the federated data commons model

[1] Discussed in the recent *Indigenous Data Governance and Sovereignty Roundtable*, 9-10 June, Melbourne.

5.2.1 Indigenous data capability and sovereignty

Indigenous data sovereignty refers to *the right of Indigenous Peoples to govern the collection, management, access, interpretation, dissemination and reuse of data related to them* (Kukutai and Taylor 2016; Snipp 2016).

In recent years, the push for Indigenous data sovereignty has led to the development of two sets of complimentary governance principles related to Indigenous research and data: CARE (by the Global Indigenous Data Alliance) and Mayam Nayri Wingara (or *To Welcome Good Knowledge,* by the Aboriginal and Torres Strait Islander Data Sovereignty Collective), outlined in the table below.

Table 3. Governance principles to enable Indigenous sovereignty over Indigenous data

CARE Principles (2018)	**Mayam Nayri Wingara Principles (2018)**
• Collective benefit to Indigenous peoples. • Authority to control, recognising data rights, interests and governance. • Responsibility to the providers for expanding Indigenous capability and fostering positive relationships. • Ethical data and processes, that minimise harm, maximise wellbeing and ensure justice.	• Exercise control of data ecosystems and data collection and use. • Data that is contextual and disaggregated. • Data that is relevant and empowers self-determination and self-governance. • Data stewardship and structures that are accountable to Indigenous peoples. • Data that is protective and respectful of individual and collective interests.

The Academy's recent State of the Social Sciences 2021 report concluded that true reconciliation with Australia's First Nations is a top priority for the social sciences, with sustained effort, leadership and resourcing required from all parts of the system to halt and reconcile the damage that has been done to Indigenous communities and Indigenous scholars through and by the social sciences.

This historical and ongoing damage includes the enormous and largely un-rewarded or compensated burden shouldered by Indigenous academic leaders. It includes the harm to communities of deficit focused research – both past research driven by overtly racist theory, and more insidious research biased towards the dominant pejorative narratives about First Nations Australia rather than the research First Nations need for nation rebuilding. Other harms include the dispossession of Indigenous Peoples from their own information, when such information is collected, stored and used by non-Indigenous researchers or government agencies without the full, informed and enduring consent of First Nations' Peoples.

For these reasons, it is imperative that data commons and data infrastructure initiatives in the social sciences explicitly focus on Indigenous data and the likely or potential impact on or interaction with First Nations People/s and communities. This applies both to research with Indigenous people specifically, and more general research that may involve Indigenous participants in community-based samples or in administrative or population level data. This focus needs to be coupled with resources and training to increase capacity of researchers across the spectrum of Australian social science to understand and apply CARE principles and relevant ethical and practice codes including those curated by AIATSIS and the National Health and Medical Research Council.

Notably, as recognised by the over 30 stakeholders attending the Future Planning roundtable (Canberra, April 5th 2022), and which were non-Indigenous in majority, getting data infrastructure and governance arrangements right for Indigenous peoples would not only benefit Indigenous Australians, but raise the

standards of practice for researchers across all domains (i.e., STEM, health, social sciences, and humanities).

Lastly, it is likely that any initiatives will encounter the same systemic difficulties associated with progressing Indigenous agendas in other research areas, particularly, the over consultation (exhaustion) of the relatively small number of Indigenous researchers working in the sector. Indigenous research leadership is the ambition, but where this is not possible, it is insufficient to wish for Indigenous participation in infrastructure planning processes: that participation needs to be enabled through the opening of Indigenous work placements specifically to fulfill those roles.

Improving Indigenous Research Capabilities: An Indigenous Research Data Commons

The HASS RDC and Indigenous Research Capability (IRC) Program, led by ARDC with NCRIS funding, includes the development of an Indigenous RDC, led by the Indigenous Data Network (based at the University of Melbourne). This national Indigenous–led project seeks to expand Australia's Indigenous research capability in ways that leverage and link new and existing data assets within an agreed framework of Indigenous Data governance principles and processes. Specifically, it aims to:

(1) Develop the necessary social architecture for Indigenous data governance and sovereignty

(2) Develop a technical architecture that builds the foundations for an Aboriginal and Torres Strait Islander Research Data Commons

(3) Create core National Indigenous Data Assets through an Aboriginal and Torres Strait Islander spatiotemporal Framework.

5.3. Tools

Big data and related capabilities (e.g., machine learning, A.I.) create exciting possibilities for research in all scholarly domains. However, the majority of social science researchers are unfamiliar with big data research approaches and methods. Qualitative methods, in particular, were largely developed to deal with the comparatively smaller samples that were traditionally available or able to be processed. The rapid expansion of unstructured qualitative data, for example, through social media, provides significant but as-yet-unrealised opportunities for widespread adoption and uptake of advanced data tools and methodologies.

To this point, systems- or platform-oriented development initiatives in the social sciences, such as the IRISS project, have focused primarily on quantitative research (i.e., readier for large-scale deployment). It is paramount for the sector to bring qualitative social sciences methods to a similar readiness state, as quickly as possible. This effort will require dedicated, multidisciplinary efforts across various disciplines and methodologies, ethical and sovereignty considerations, and on the technical implementation of solutions and their integration into processes and policies[2].

Table 4. Current state, advanced computational research tools

Enablers	Comments
Goal: Data discovery and exploration, linkage and analysis enabled by fit-for-research platforms and tools	
Regulations in place, to support efficient, safe and ethical data linkage; and ensure ethical application and use of machine learning, AI and other emergent research technologies	• Key pieces ready (e.g., FAIR and CARE principles, Mayam Nayri Wingara principles) but non-binding • Other key pieces still under development, such as the specific rules and protocols that will ensure privacy is preserved when datasets are shared or linked at the level of the individual or family
Investments to develop and share advanced analytical tools	• $8.9 million Cth Government investment into the HASS RDC and Indigenous Research Capability (IRC) Program, while a significant landmark for the social sciences, is still effectively pilot funding in comparison to the investment that will be required. • Continued funding for the HASS RDC and Indigenous Research Capability (IRC) Program, or their future large-scale deployment is not guaranteed
Incentives to develop, maintain and enhance broad/public access to advanced analytical tools	• [Too early]
Systems support advanced research computing capabilities	• The IRISS project, along with the Linguistics and Indigenous Data Commons projects (under the HASS RDC and Indigenous Research Capability (IRC) Program umbrella) are well underway to design and proof a suite of platforms, environments, standards and workflows required to enable an RDC for the social sciences • Higher-end computational methods still make a small proportion of social science research
Communities have the skills to critically apply advanced analytics for research and policy	• The social science research community is yet to define a concerted way forward for qualitative research in the new environment (big data, new objects of study and data formats, new analysis capabilities, etc)

[2] One researcher compared the challenges ahead to the work undertaken years' back in the health community to develop the Human Genome Project.

5.4. Data

A lack of access to large social datasets has been a major roadblock for Australian social science research for many years. However, this situation is rapidly changing. For example:

- The Data Availability and Transparency Act 2022 (passed in April 2022) allows for access to and reuse of data held by government agencies for research purposes, where research is demonstrably in the public interest.
- Twitter has made significant volumes of data openly available to anyone willing (and able) to utilise it.

While other barriers to data access remain (much data of interest to the social sciences are held by corporations, for example, some rests with individuals, and some are yet to be captured), there is a growing appreciation that the data scarcity that characterised social science research in the past will no longer be a problem in many areas. We are, instead, approaching an era of data abundance, where the key issues will involve choosing the best research tools and research questions. Identifying which datasets constitute 'nationally significant' assets, for example, will weigh heavily on future public funding allocations[3].

Over the coming years, the sector needs to continue to amplify and strengthen researchers' access to human data while simultaneously asking the hard questions about what commodities, if not data, will carry *value* in the future. Some researchers are, for example, speaking of *code* being the future source of value to circulate through data commons. Both aspects will be critical to good decision-making around investments in infrastructure. For example, some of the data-centric public infrastructures, such as archives, libraries, and data agencies, are not purpose-built or adequately funded to deal with big data or emerging digital formats; but simply expanding on the traditional archive model could get the sector into path-dependent infrastructure investments that could quickly become outdated.

Table 5. Current state, data

Enablers	Comments
Goal: Well-structured, fit-for-research data tributaries support evidence-based research and policy	
Regulations in place to support efficient, safe and ethical data capture, collection and access	• [See 'Regulations' in **Table 4**] • The *Data Availability and Transparency Act 2022* is an important step towards regulating use and access to government data through legislation. Appropriate use and access to data by private firms and individuals could be next • Incentives for data production and sharing largely lacking
Incentives to encourage big data production and sharing (incl. private sector, individuals)	
Investments to produce and maintain public interest-only or option-value datasets/archives	• The sector is pending to define what data assets and other tools are critical to research and society, and warrant funding and protection as public goods
Systems enable big data storage and discovery	• Archive facilities largely fragmented and underfunded • Experts recommend thinking beyond 'storage', and into higher-level issues such as discovery and linkage instead
Communities are aware of available data, and prepared to produce, access and use data safely, ethically and efficiently	• The Office of the National Data Commissioner is building a catalogue of government-held data. The sector would benefit from replicating for all nationally-significant data. • [Refer back to **Table 1**, 'Communities']

[3] Identified as a key activity in Tindall and Duncan (2020) *Humanities, Arts and Social Sciences Research Data Commons*, published by ARDC.

6. Interim priorities

This environmental scan led to identify the following interim priorities. They are proposed as interim, as they are still work in progress, yet to be tested and revised by the sector's stakeholders.

1) **A Decadal Plan**
   Develop a vision statement and roadmap for social science data commons, and back-cast milestones required to achieve the anticipated outcomes. The idea of a Decadal Plan (2023-33) for social science infrastructure was welcomed by stakeholders at the future planning roundtable (Canberra, 5th April 2022). A Decadal Plan should simultaneously look after the identified four pillars (workforce, governance, tools and data), and bring together governments, academic institutions, data producers and other key stakeholders around agreed priorities

2) **Funding for Indigenous commons and governance**
   Ensure Indigenous data governance and sovereignty are woven into data commons infrastructure initiatives across all domains, and their leading agencies and governing bodies appropriately funded. After all, Indigenous self-determination is as much a policy/culture problem, as a resourcing one

3) **More funding towards the development of advanced research computing capabilities for the social sciences, such as the IRISS project and others currently underway through the HASS RDC and Indigenous Research Capability (IRC) Program**
   Advocate for continued AND increased Cth funding for the HASS RDC and Indigenous Research Capability (IRC) Program

4) **Inserting social priorities into the NRI Roadmap**
   Advocate for better inclusion of social priorities in the National Research Infrastructure Roadmap and related planning processes

5) **Scale up qualitative social science methods**
   Enlist and appropriately fund initiatives that fast-track the adaptation of social science qualitative research methods (from data collection, through to analysis) to big data (and related capabilities – e.g, machine learning, A.I.), and new research substrates and formats

6) **Fast-tracking deployment of the *Data Availability and Transparency Act 2022***
   Identify opportunities to accelerate researcher access to government data, to take advantage of the recently adopted *Data Availability and Transparency Act 2022*

7) **Future-proofing the social science commons**
   Identifying and understanding the trends that will shape social science research and supporting data common systems in the medium- and longer-term futures; and integrating them into investment plans and decision-making frameworks.

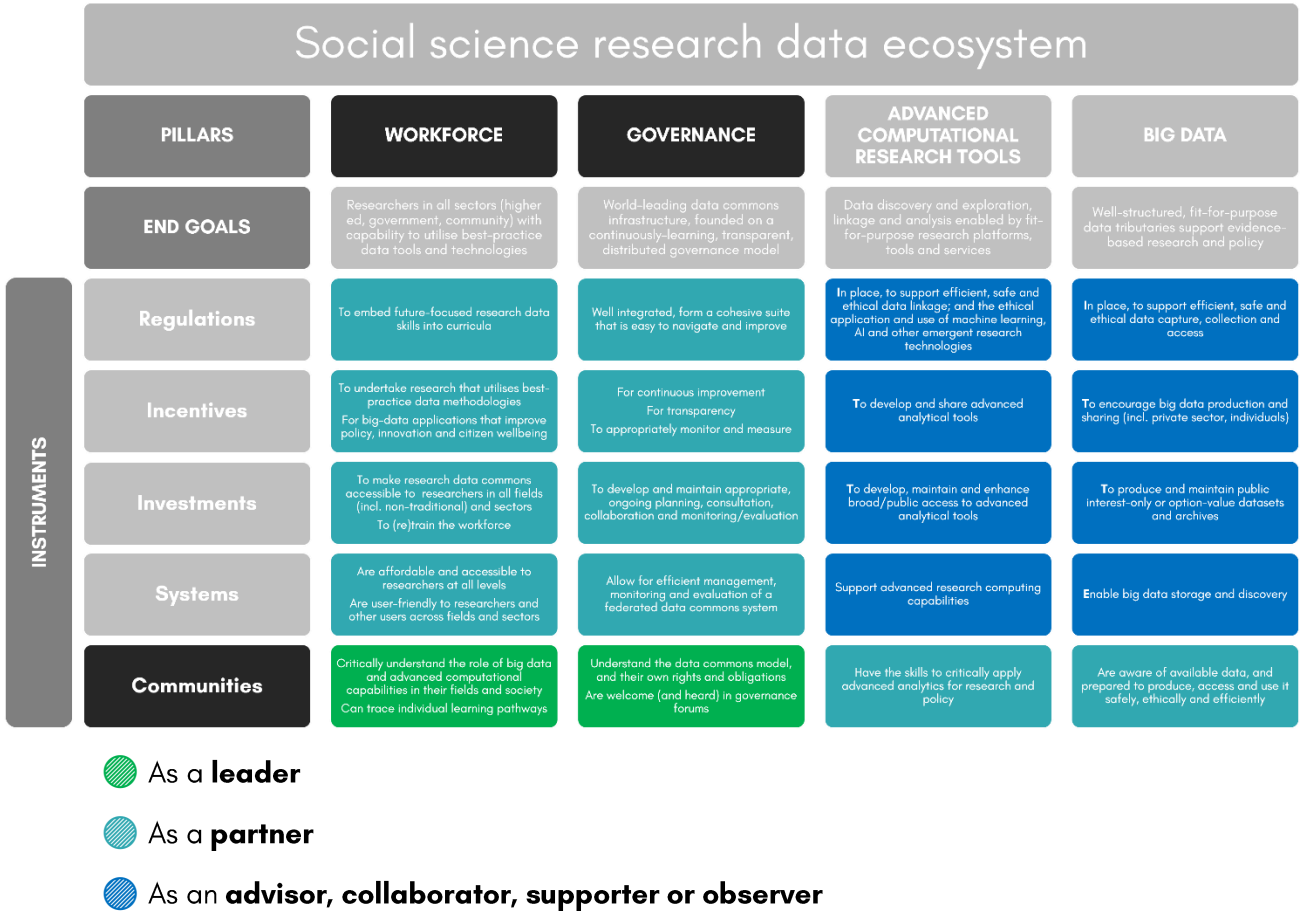
To the extent possible, efforts should restrain from building anew, and look instead at strengthening existing institutions, organisations, platforms, programs and alliances.

6.1. The role of the Academy

Because of its nature, the Academy is best placed to (**Figure 2**):

- **Lead** initiatives in the **workforce and governance pillars** (which directly relates to its constituents and its advocacy function), particularly in the **communities stream**
- Be an **active partner** in initiatives in workforce and governance pillars in other streams (regulation, investment, incentives or systems)
- Act as an **advisor, collaborator, supporter or observer** in all other areas, in the appropriate capacity (e.g., conduit to research experts, advocacy organisation).

Figure 2.Where is the Academy best placed to contribute?



6.1.2 A Decadal Plan for Social Research Commons 2023-33

As an immediate first step, the Academy is proposing to lead, in partnership with the ARDC, the development of the aforementioned Decadal Plan for social science data commons infrastructure. Having a Decadal Plan (with sufficient stakeholder buy-in) means any future substantive initiatives will take place within an integrated, purposeful whole, and higher chances to secure funding and resourcing.

7. In their own words

Key extracts from stakeholder consultation:

"One of the really big challenges in humanities and social science research is the fragmentation of research methods. If you think about biology, the Genome Project, for example, that's a massive international research effort, made possible by shared methods and protocols for recording, annotating, sharing and distributing data. In the humanities and social sciences we all do different sorts of things in terms of accessing and recording our data."

"It wasn't until I started working closely with a librarian that I learned about the importance of metadata, as the basis for indexing and searching. It's something fundamental: thinking about how we structure our data collections, make them information rich, and able to talk to one another."

"None of my skills in digital humanities has come from systematic or formal training. I've just gone to the archives, work out how that works. [...] A lot of the training offered to research students is still very focused on traditional data sets, and traditional ways to query the data. The questions we ask now are very different to those researchers made 20 years ago"

"We can get data on land quantities but, what about qualities such as, is that land irrigated, or rain fed? Does it have good soils? Good or bad climate? And how are those qualities changing over time? These are the data we need to think about food production, consumption and equity into the future. I find that inadequately handled at the moment".

"Right now, you couldn't tell the social, political, or cultural history of the last decade without reference to social media. That data is everywhere, but it's also elusive. [...] How will we make it accessible for future researchers?"

"For a lot of audio-visual materials, the only copy is something uploaded to YouTube, by a hobbyist who got a VCR tape. None of it is being collected in a systematic way. Yet, television and film are fundamental to anyone studying the history of the post war period".

"We want to create access but can't possibly keep records of everything. So the question is, how do we go about choosing what to preserve for the future? It's a question that we have to tackle and for which we don't have a clear answer now."

"We're not going to have governments handing over billions of dollars to pursue all possible research agendas or mass digitisation projects. It's just not going to happen. We're always going to be working in a context of scarcity."

"How many times does the same individual access a critical service? We made a decision in the past to provide our services in a completely confidential way, to protect people's privacy. But that means we don't have the means to identify repeat users of the service, who might be in a critically vulnerable situation"

"Leadership in this area will not only benefit the social sciences, but set the standards for ethical use of data in the corporate world."

"Coming from the United States, I can't help but notice the whole infrastructure to support social research in Australia is very tailored to traditional social science methods. Data coming from interviews, or focus groups, for example, is well supported. But when someone is trying to do big data research, it just... doesn't work well at all. We need to become much more efficient at dealing with the ethical questions arising from doing research using social media or other big data assets."

"Current ethics processes create significant challenges for open research... I sit on an external ethics board. And my perception is that, across different universities, they really don't want us to deposit qualitative research data that could potentially have privacy implications around sensitive topics. For example, if I wrote in my ethics application that I want to deposit data related to gender-based violence, in an anonymous open-source way, it would just continually come back to me through the ethics review process. What's going to happen to this data we're collecting on gender violence, that we currently cannot share?"

"Having gone through the process of seeking access to public records across Australia, we found that every state, every department, every organisation had their own objections. It often depended on who we knew, who we could get to review our application. In my experience, these organisations use privacy as a way to say: 'Well, you can't access these materials that are potentially revealing quite hard things, for example, about child sexual abuse' or other kinds of record keeping."

"We need to work more on our visions of the data futures. I have not read any reports that give me a clear vision of the data futures that we'd like to see in 2030. A vision we can back-cast from, to figure out what steps we need to take today. That's number one."

"We need to understand what is possible in this new environment, because I don't fully get that. Sort of a profile of the kind of research, questions and tasks we could be tackling, and how."

"On the infrastructure roadmaps, there is a series of priority areas, and social science and humanities priorities are not there. It's all very informed by industry, manufacturing and STEM. We need to put our own agendas among those national priorities, because then you get attention from various bodies and organisations, and then the funding and the strategy follow that."

"Even in long term crises like climate change, it's taken a long time for research bodies to respond. Because they're either not allowed to or not funded to, and the funders themselves have long cycles." [...] "Each time there's an event, a tornado... takes half of Armidale... and we're struggling to capture them. And with climate change, it's only going to get worse."

8. Appendix A – Consultation details

Details about the four roundtables and online survey are shown in Error! Reference source not found.**Table 6** below.

Table 6. Consultation methods (details)

Consultation method	Guiding questions	Participant details
Roundtable Scenario-based	**Food futures** We could maximise our chances at achieving equitable nutrition, while minimising ecological nutrition and maintaining economic prosperity if we: • Produced or had access to [these datasets] • Along with [these tools] (e.g., for analysis, prediction, management) • General or discipline-specific training on [these skills] • Supported by [these policies].	**Session 1** 2 February 2022 | 9am AEDT | 90 min 9 participants | Zoom Facilitator: Shannon Callaghan (ARDC) **Session 2** 4 February 2022 | 9am AEDT | 90 min 9 participants | Zoom Facilitator: Shannon Callaghan (ARDC)
Roundtable Scenario-based	**Collective memories of gender** To preserve and future-proof our collective memory of gender: • We must develop or gain access to [these datasets] • Along with [these tools] (e.g., search, display, processing, management) • General or discipline-specific training on [these skills] • Governed and supported by [these policies].	**Session 1** 2 March 2022 | 9am AEDT | 90 min 10 participants | Zoom Facilitator: Chris Hatherly (Academy CEO) **Session 2** 4 March 2022 | 9am AEDT | 90 min 11 participants | Zoom Facilitator: Isabel Ceron (Academy Policy Analyst)
Roundtable Future planning	**Social Sciences Research Infrastructure Roundtable** • Current state of social science infrastructure (data, platforms, tools, services, governance) • Where are the key gaps? • Opportunities for leadership • A long-term strategy for social science research infrastructure?	5th April, ANU, Canberra 26 Participants (in-person and online) 3 Observers Presenters: Jenny Fewster (ARDC), A/Prof. Steven McEachern, Prof. Mark Western FASSA Facilitator: Chris Hatherly (Academy CEO)

Consultation method	Guiding questions	Participant details
Online survey General skills and knowledge	• How familiar are you with each of these data terms: FAIR principles, CARE principles, structure data, open data, and controlled vocabularies? (5-point rating scale) • Which of these skills and methodologies do you apply in your research: Descriptive statistics, inferential statistics, data-based simulations and scenarios, coding, and machine learning? (4-point rating scale)	94 responses Administered via Google Forms Period: 21 February – 7 March Promoted via direct email and social media.

Details about roundtable session participants are shown in **Table 7** below. Attendees participated under condition of anonymity.

Table 7. Roundtable sessions, participants

Roundtable session	Participant details
Food futures Sessions 1 & 2	**Researchers** | Food systems, innovation, nutrition, agriculture, and climate change (11) **Not-for-profit sector** | Nutrition-related organisations, board member representatives (3) **Industry** | Peak body representatives (2) **Government** | Data and research managers, agriculture-related agencies (2)
Collective memories of gender Sessions 1 & 2	**Researchers** | History, anthropology, philosophy, law, business & leadership, and information studies (15) **Galleries, libraries, archives, and museums (GLAM) sector** | (3) **Not-for-profit gender-related organisations** | (3)
Future planning	**26** leaders from Australian social science research institutes and Commonwealth government data agencies, including: • Australian National University • The University of Queensland • Melbourne University • University of Tasmania • Macquarie University • RMIT University • University of Western Australia • National Library of Australia

Roundtable session	Participant details
	• Australian Bureau of Statistics • Australian Bureau of Agricultural and Resource Economics (ABARES) • Australian Institute of Health and Welfare (AIHW) • Department of Education, Skills and Employment (DESE) • Australian Academy of the Humanities • Australian Institute of Family Studies • Australian Council of Social Services • Australian Education Research Organisation.

The survey was specifically targeted to researchers in social science disciplines. Details about the respondents to the online survey are displayed in **Table 8**, **Table 9**, and **Table 10** below.

Table 8. Survey respondents, career stage

Respondent type	Count	%
Senior researcher	36	38%
Mid-career researcher	23	24%
Early-career researcher	23	24%
PhD student	6	6%
Other	6	6%
Total	94	100%

Table 9. Survey respondents, university affiliation

University	Count	%
The University of Queensland	**18**	**19%**
The Australian National University	**13**	**14%**
The University of Melbourne	**12**	**13%**
Central Queensland University	5	5%
The University of Western Australia	4	4%
James Cook University	3	3%
The University of Sydney	3	3%
Monash University	3	3%
Victoria University	3	3%
Griffith University	2	2%
University of Technology, Sydney	2	2%
Western Sydney University	2	2%
UNSW Sydney	2	2%
University of Newcastle	1	1%
Charles Darwin University	1	1%
Edith Cowan University	1	1%
Flinders University	1	1%

University	Count	%
Queensland University of Technology	1	1%
University of Tasmania	1	1%
Macquarie University	1	1%
University of South Australia	1	1%
La Trobe University	1	1%
RMIT University	1	1%
University of the Sunshine Coast	1	1%
Other, not listed	11	12%
Total	**94**	**100%**

Table 10. Survey respondents, discipline

Discipline	Counts	%
Economics	21	22%
Sociology	8	9%
Public Health	6	6%
Linguistics	5	5%
Management, Commerce, Tourism and Services	5	5%
Political Science	4	4%
Psychology	4	4%
History	4	4%
Demography	4	4%
Policy and Administration	3	3%
Education	3	3%
Criminology	3	3%
Human geography	3	3%
Communication and Media Studies	3	3%
Anthropology	2	2%
Urban and Regional Planning	2	2%
Development Studies	1	1%
Statistics	1	1%
Gender Studies	1	1%
Other	11	12%
Total	**94**	**100%**

9. Appendix B – Project team

Members of the project team included:

Expert Steering Group

- Prof. Mark Western, FASSA | Director, Institute for Social Science Research (ISSR), The University of Queensland
- A/Prof. Steven McEachern | Director, Australian Data Archive (ADA), Australian National University
- David Kalisch, FAICD FASSA | Former Australian Statistician (2014-19), Consultant (DWK Consulting)
- Prof. Maggie Walter, FASSA | Professor of Sociology and Pro Vice Chancellor Aboriginal Research and Leadership, University of Tasmania & Commissioner, Yoo-rrook Justice Commission
- Prof. Sarah Medland, FASSA FAHMS | Mental Health Research Program, Psychiatric Genetics, QIMR Berghofer
- Prof. Sarah Pink, FASSA | ARC Centre of Excellence for Automated Decision-Making & Society, Monash Energy Institute, Monash University
- Prof. Stuart Barr, Director, Australian Urban Research Infrastructure Network (AURIN).

Academy staff

- Dr. Chris Hatherly, CEO
- Dr. Isabel Ceron, Senior Policy Analyst
- Callum Jones, Policy Analyst.

ARDC guidance and support

- Shannon Callaghan, Senior Policy Advisor.

10. Appendix C – Survey results

By Callum Jones

Participant demographics

In total, 94 respondents participated in the survey. These respondents represented a wide array of humanities and social science disciplines, all of which, and their proportion of the overall sample.

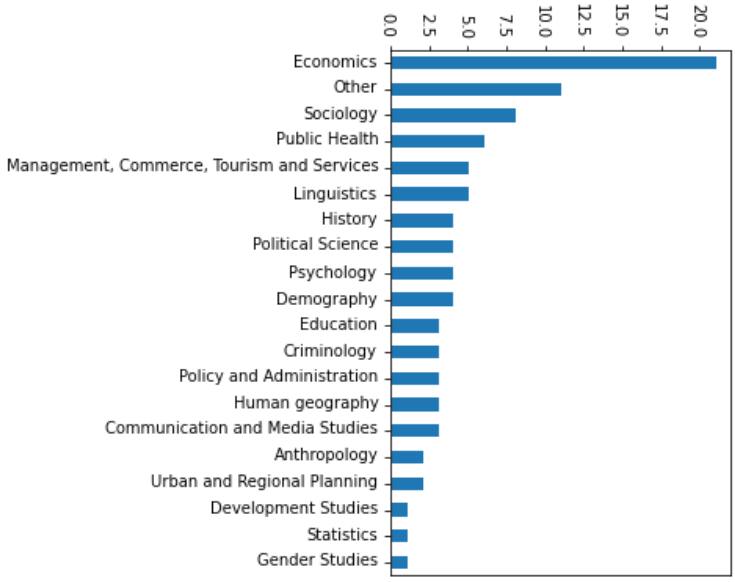
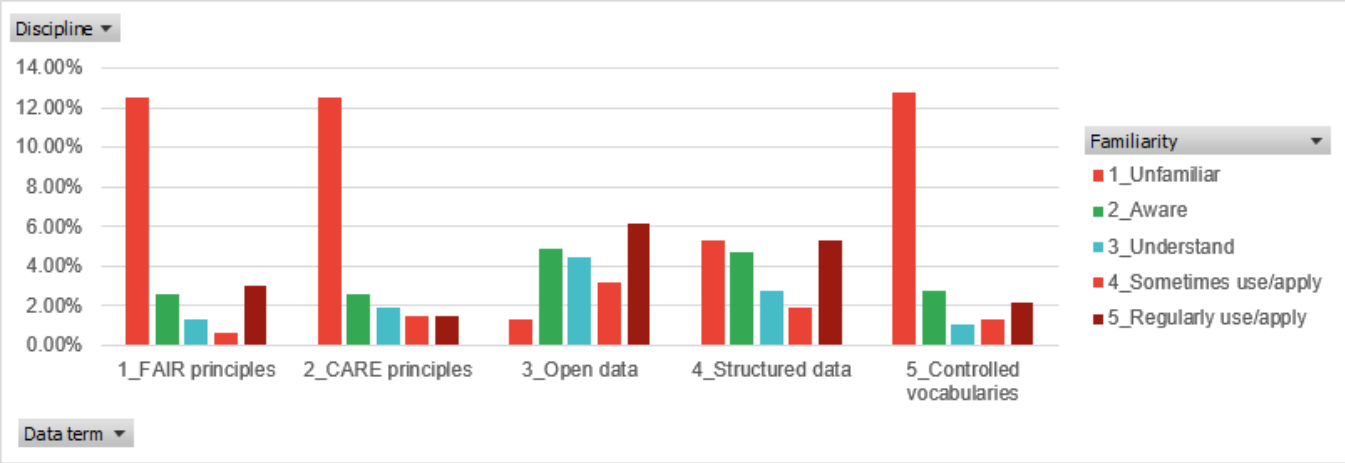


Figure 3. Academic disciplines of respondents

Familiarity with data commons principles

Respondents were asked to indicate the degree to which they were familiar with a range of data principles, types and methods of storage and structure. As can be seen in the figure below, respondents are not widely familiar with the data principles underpinning the current data commons model.

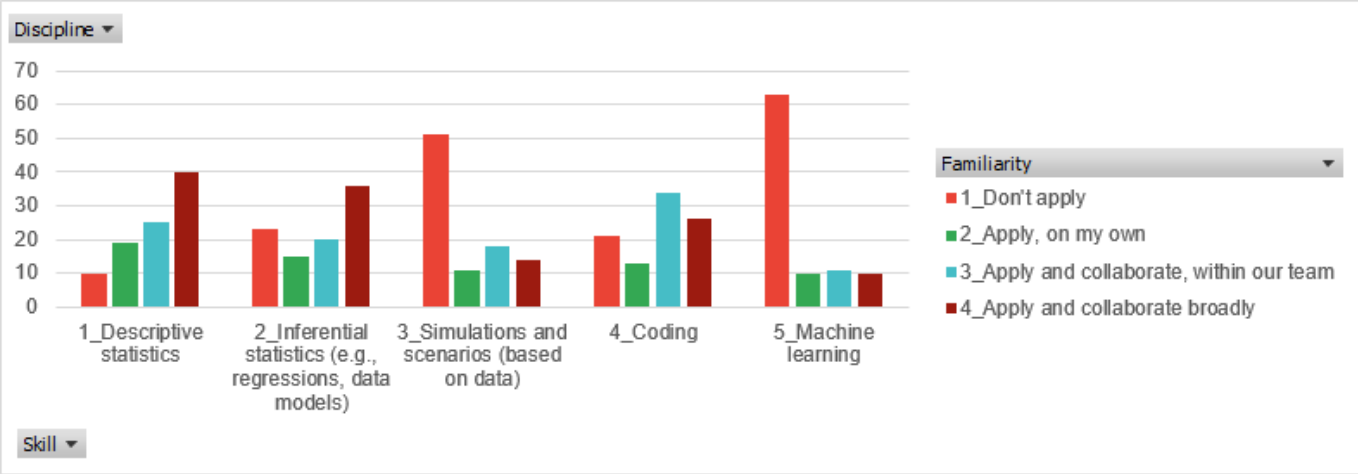
Figure 4. Researchers' level of familiarity with data commons principles



Familiarity with advanced computational research methods and tools

Respondents were asked to indicate which methods and tools they use. Results show respondents were far more likely to apply different forms of statistical analyses, as opposed to manual coding approaches, and computational methods, such as simulations and scenarios and machine learning.

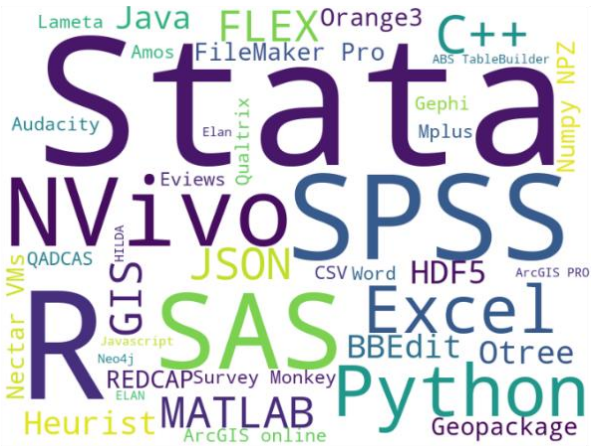
Figure 5. Researchers' level of familiarity with advanced computational research methods and tools



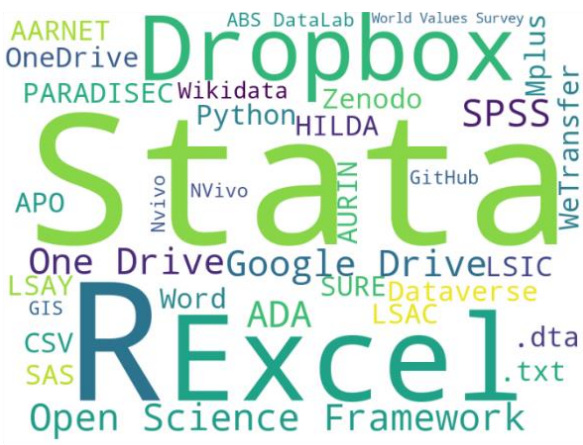
Interestingly, as explored in greater depth in Section 3.1.1, the computational approaches that were identified by respondents as being those that they do not apply on their own, nor in collaboration, are those also identified by respondents as the methods they would be most interested in learning how to apply.

Tools to work individually and tools to collaborate

Respondents were asked what tools (e.g., software, other) they use, and whether they use them individually or for collaboration. The figures below show the tools most often used by researchers in individual (left) and collaborative (right) research. Statistical software tools are the most widely used software type amongst surveyed social scientists, accounting for 53% of the tools used (the most popular being Stata, R, SPSS, SAS and Excel, in that order). Qualitative-specific software such as NVivo was used by respondents, but nowhere near as widely (of all the software listed by respondents NVivo accounted for just 5%); although this could be a result of the greater balance of respondents from academic disciplines such as economics, political science and psychology.



Tools applied in individual projects



Tools applied for collaborative research

Barriers and limitations

Insufficient training was the top barrier or limitation cited by respondents, when asked what was limiting greater use of big data and related capabilities in their research.

Biggest developments

Respondents were asked to identify what they thought to be the biggest data developments in their discipline or area. Top responses included machine learning, big data, social media data and increasing levels of data linkage.