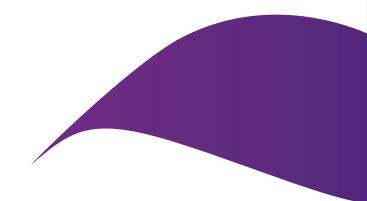


# Data Systems that Work for Data Users

Mark Western, The University of Queensland, Australia

Presented to the Our Data Sources as a Strategic National Asset Symposium, Wellington New Zealand 9 March 2023





# National Consultations & Scoping Studies

- Australian Academy of the Humanities 2020. Mapping International Research Infrastructure for the Humanities, Arts and Social Sciences
- Tindall, A. and Duncan, I. for Australian Research Data Commons. 2020. Humanities Arts and Social Sciences Research Data Commons.
- Academy of Social Sciences in Australia 2021. The State of the Social Sciences Report 2021
- Australian Council of Learned Academies, Australia's 5 Learned Academies and Australian Research Data Commons. 2021. Australia's Data-Enabled Research Future.

 2022-2023. Decadal plan for Australian Social Science Research Infrastructure – current activity by the Academy of the Social Sciences in Australia.



# Why so much activity?

2018 – 2029 \$4B federal investment to support national research infrastructure

Planning organised through 5 yearly roadmaps (first 2006) & 2-3 year investment plans

Social sciences lobbying since 2006 for national research infrastructure capability

2018 research investment plan allocated funding to scope HASS and Indigenous research platforms

2019 Federal budget \$8.9 million HASS Research Data Commons and Indigenous Research capability (https://ardc.edu.au/program/hass-rdcindigenous-research-capability/)



### Australian Research Infrastructure 'Layers'

#### Institutional

Smaller scale infrastructure initiated at an institutional level and not in scope for the 2021 Roadmap

#### National Nationally significant

Nationally significant assets, facilities and services to support leading-edge research and innovation

#### Landmark

Large scale facilities that serve diverse communities and need international collaboration

#### Global

Multinational, collaborative infrastructure of a scale where cost of establishment is beyond the means of a single nation



2021 National Research Infrastructure Roadmap



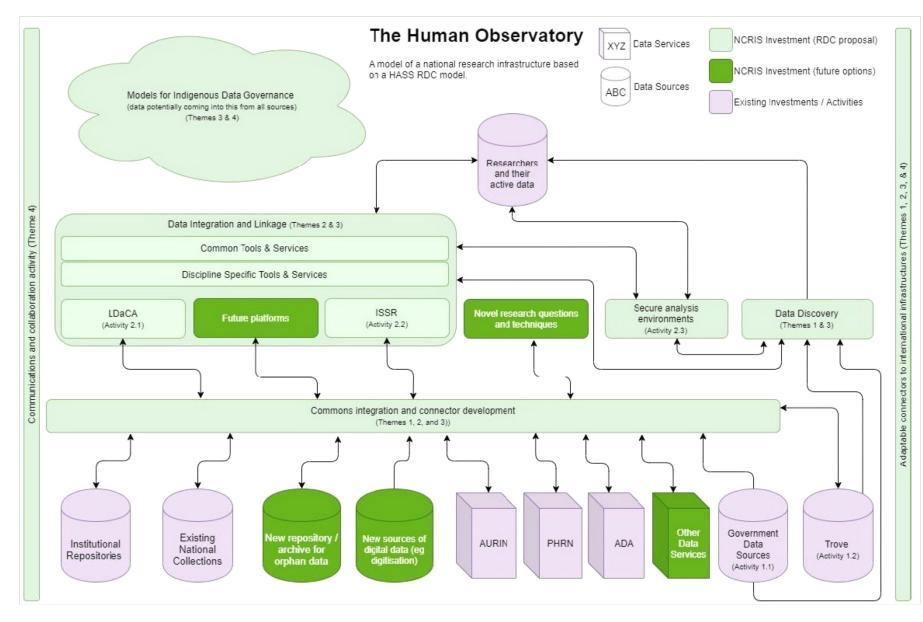
#### What would a national social science research capability look like?

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

	Social science research data ecosystem				
	PILLARS	WORKFORCE	GOVERNANCE	ADVANCED COMPUTATIONAL RESEARCH TOOLS	BIG DATA
	ASPIRATIONS	Researchers in all sectors (higher ed, government, community) with capability to utilise best-practice data tools and technologies	World-leading data commons infrastructure, founded on a continuously-learning, transparent, distributed governance model	Data discovery and exploration, linkage and analysis enabled by fit-for-purpose research platforms, tools and services	Well-structured, fit-for-purpose data tributaries support evidence- based research and policy
ENABLERS	Regulations	To embed future-focused research data skills into curricula	Well integrated, form a cohesive suite that is easy to navigate and improve	In place, to support efficient, safe and ethical data linkage; and the ethical application and use of machine learning. Al and other emergent research technologies	In place, to support efficient, safe and ethical data capture, collection and access
	Incentives	To undertake research that utilises loest-practice data methodologies For big-data applications that improve policy, innovation and citizen wellbeing	For continuous improvement For transparency To appropriately monitor and measure	To develop and share advanced analytical tools	To encourage big data production and sharing (incl. private sector, individuals)
	Investments	To make research data commons accessible to researchers in all fields (incl. non-traditional) and sectors To (re)train the workforce	To develop and mointain appropriate, ongoing planning, consultation, collaboration and monitoring/evaluation	To develop, maintain and enhance broad/public access to advanced analytical tools	To produce and maintain public interest-only or option-value datasets and archives
	Systems	Are affordable and accessible to researchers at all levels Are user-friendly to researchers and other users across fields and sectors	Allow for efficient management, monitoring and evaluation of a federated data commons system	Support advanced research computing capabilities	Enable big data storage and discovery
	Communities	Critically understand the role of big data and advanced computational capabilities in their fields and society Can trace individual learning pathways	Understand the data commons model, and their own rights and obligations Are welcome (and heard) in governance forums	Have the skills to critically apply advanced analytics for research and policy	Are aware of available data, and prepared to produce, access and use it safely, ethically and efficiently

Academy of the Social Sciences in Australia 2022. Australia's Data Enabled Research Future: The Social Sciences





Tindall and Duncan 2020. HASS Research Data Commons



#### 5 Components of HASS RDC & Indigenous Research Capability

- Language Data Commons of Australia vulnerable language collections
- Trove Researcher Platform researcher portals for Trove public heritage site
- Integrated Research Infrastructure for the Social Sciences access, preserve, disseminate quantitative & qualitative data & systems and tools for new & real-time data capture
- Indigenous Research Capabilities: Aboriginal and Torres Strait Islander Research Data Commons – technology, training and governance
- HASS Community Data Lab data access to other collections and archives
- 3 Cross-cutting integration activities
- Access, Authentication and Authorisation
- Gazeteer of Australian Historical Places
- Online Heritage Resource Management



#### Integrated Research Infrastructure for Social Science

A SYSTEM TO	BY MEANS OF	IN ORDER TO
Enhance nationally significant datasets that underpin empirical social science research.	Using best practice data harmonisation techniques, domain informatics and interoperability standards.	Create critical longitudinal, demographic and geospatial data infrastructure for social science researchers to build their evidence base on.
Coordinate data and research infrastructure components that support data-drive social science research.	An integrated workflow between vocabularies, survey tools and data repository service.	Establish a foundation for national research infrastructure for social science research.



### **Project outcomes**

1. Coordinated data governance and integration	Establishing a coordinated governance and integration model for the provision of data and infrastructure in the social sciences and related disciplines in Australia
2. Enhanced research infrastructure	Enhancing the research capacity of Australian social science researchers, through the development of tools and services to enable the creation, dissemination and use of quantitative and qualitative social science data sources
3. New data integration environment	Enabling a cost-effective and accessible data integration environment (for lower risk data)
4. New researcher training opportunities	Providing a new program library for data processing and training packages for the management of social science research data

Source: HaSS and Research Data Commons – Let's Play - IRISS



### **IRISS Work Packages**

- WP1: IRISS project management and coordination (ANU)
- WP2: Vocabulary Access Service for Social Science in AustraLia (VASSAL) (ANU)
- WP3: GeoSocial data integration service (UQ ANU AURIN)
- WP4: IRISS demonstrators
  - D1: Spatial data analysis demonstrator (UQ)
  - D2: Sensitive data analysis demonstrator (UniMelb)
  - D3: Australian Census Digital Collection demonstrator (ANU)
- WP5: Survey Project Integrated Research Environment (SPIRE) (ANU)
- WP6: Curation of Australian Research Data in the Social Sciences (CARDSS) (ANU)

Source: HaSS and Research Data Commons – Let's Play - IRISS



### Two core elements: Data + Framework

7 key data assets based on Federal integrated public sector administrative data

MADIP – health, education, govt payments, income & tax, employment, population demographics – longitudinal

BLADE – tax, trade, intellectual property + business and economic surveys

Australian Census Longitudinal Dataset – linked from 3 most recent population censuses

Linked Employer-Employee dataset – cross-sectional, employer information + personal income tax & payments, jobs

Linked datasets for migrants & temporary entrants



# Framework for integration & sharing

Data Availability & Transparency Act 2022 – Administered by National Data Commissioner (regulator)

- Promote better availability of public sector data
- Enable sharing of public sector data consistent with Privacy Act and security safeguards
- Enhance integrity & transparency of sharing public sector data
- Build confidence in use
- Establish institutional arrangements for sharing

Participants

- Data Custodians Commonwealth bodies who control public sector data
- Accredited users public sector organisations & Australian universities must be accredited
- Accredited data service providers public sector organisations & universities provide integration, deidentification & access services



#### Background

#### **PROBLEM:**

- Fragmented infrastructure landscape for Australian empirical social science researchers
- Limited resources to integrate expand existing data infrastructure and grow user base
- Significant amount of data needs to be kept confidential and/or is subject to privacy considerations

**SOLUTION:** 

- IRISS a new foundation for integration of data, analysis and platforms for social science research
- Integration & interoperability to keep up with leading edge global research and research infrastructures Europe & US
- Starting point: data –acquisition, documentation, harmonisation and dissemination for reuse