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

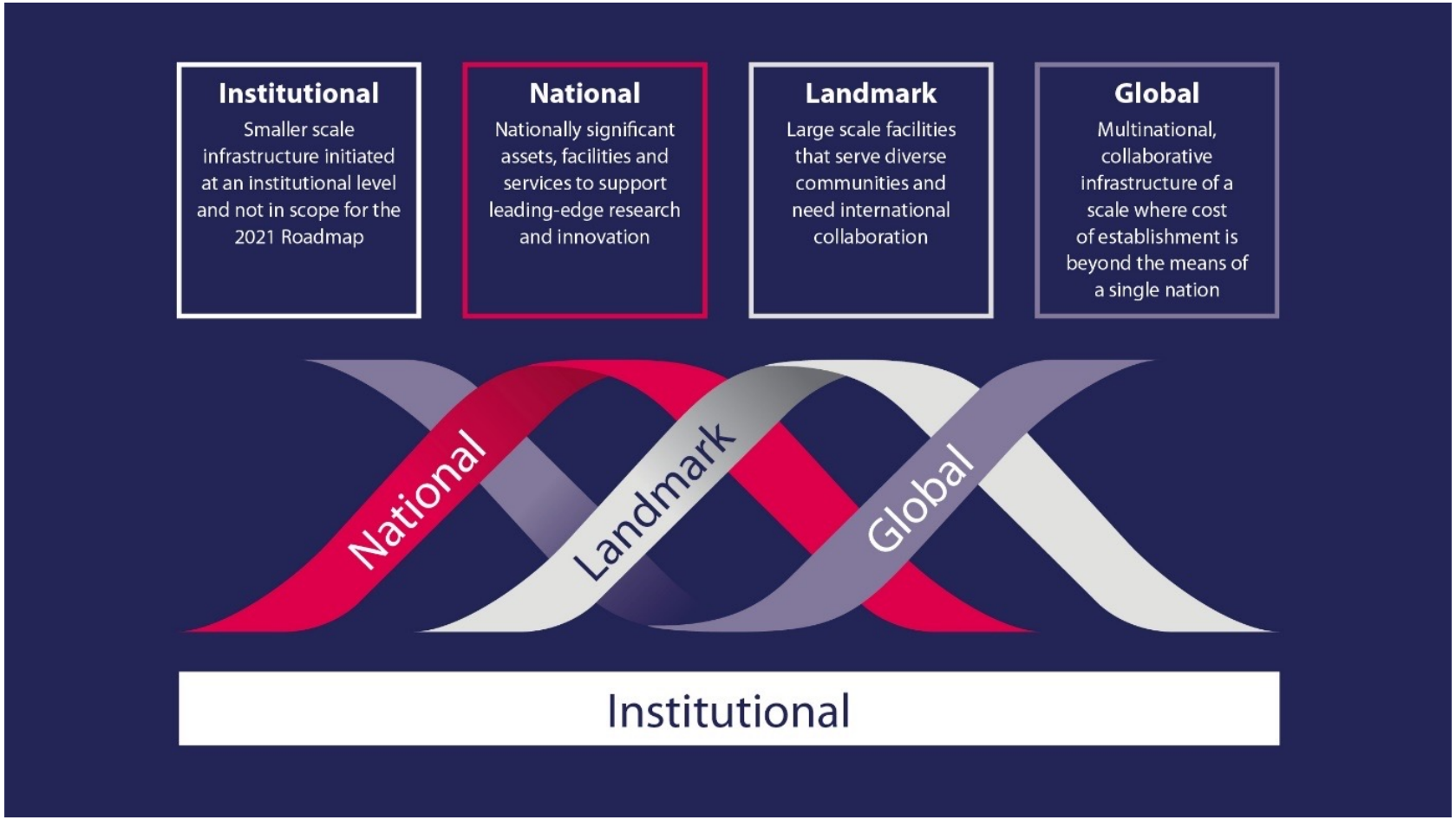
Planning organised through 5 yearly roadmaps (first 2006) & 2-3 year investment plans

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-rdc-indigenous-research-capability/)

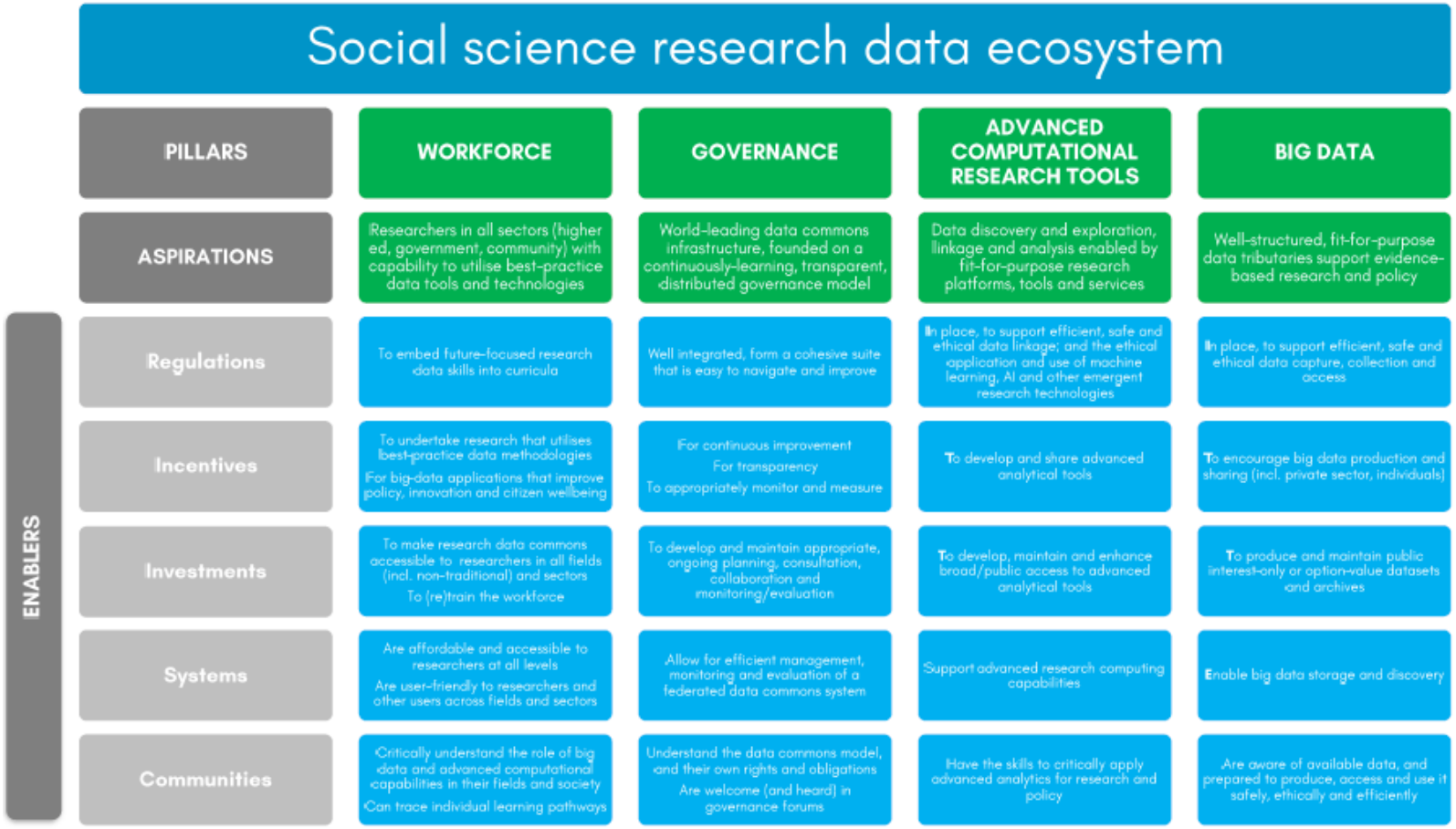
Australian Research Infrastructure ‘Layers’



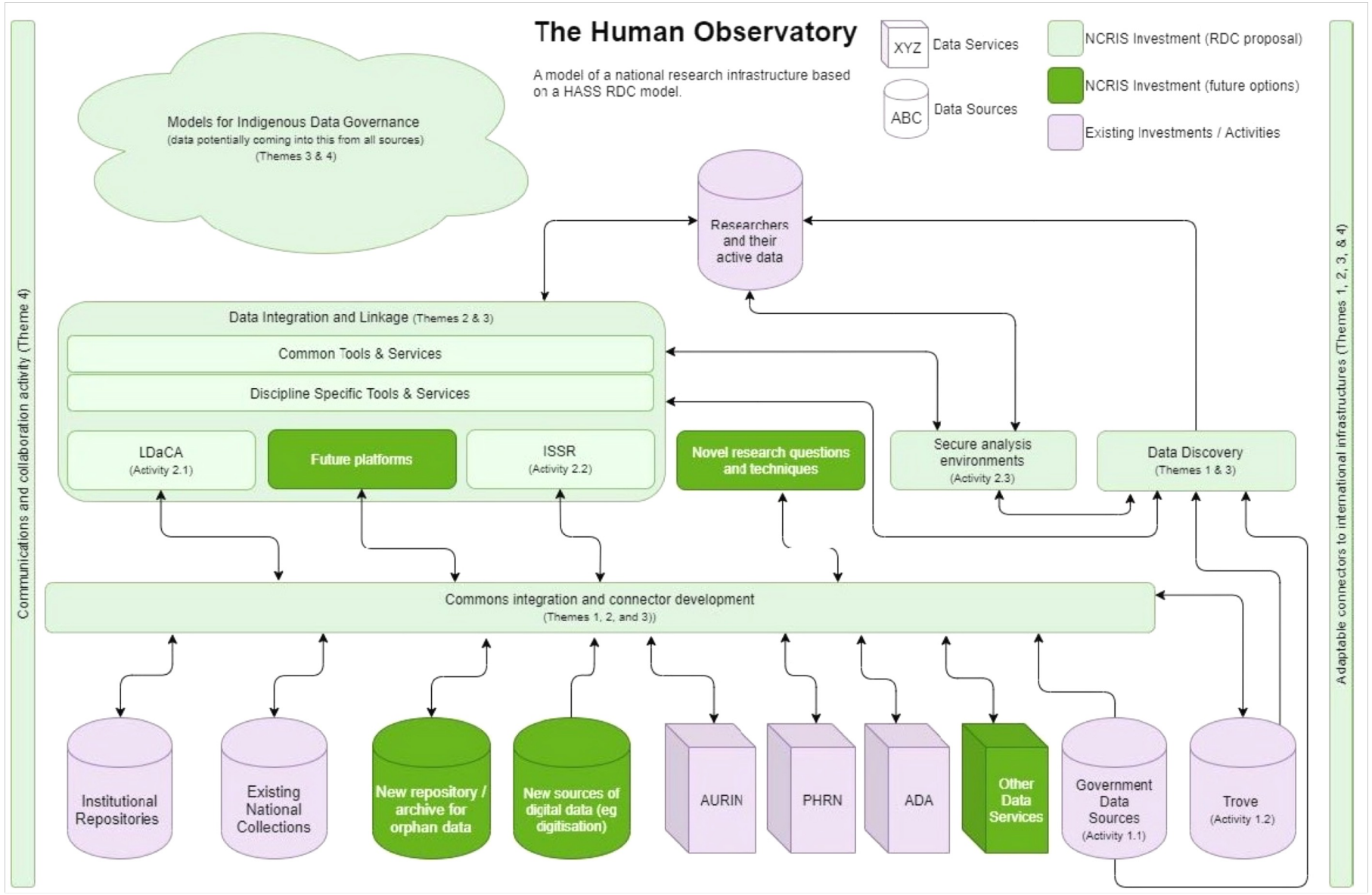
2021 National Research Infrastructure Roadmap

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



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.

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

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, de-identification & 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 re-use**