Enabling institutional environment for data integration - Australia

24 November 2020

Ms Deepa Wright

Director, Data Services

Australian Bureau of Statistics
Informing Australia's important decisions



Australian experience

"What does Australia have in place that makes data integration possible in terms of legal and institutional considerations?"

MADIP - Partnerships

The **M**ulti-**A**gency **D**ata **I**ntegration **P**roject (MADIP) is a secure, person based research data asset combining information on health, education, government payments, income and taxation, employment, and population demographics to create a comprehensive picture of Australia over time..

MADIP Uses

- Answer policy questions
- Program evaluation
- Empirical research on socio-economic issues
- No identification of individuals
- Used only for statistical and research purposes (never for compliance)

MADIP Partner Agencies

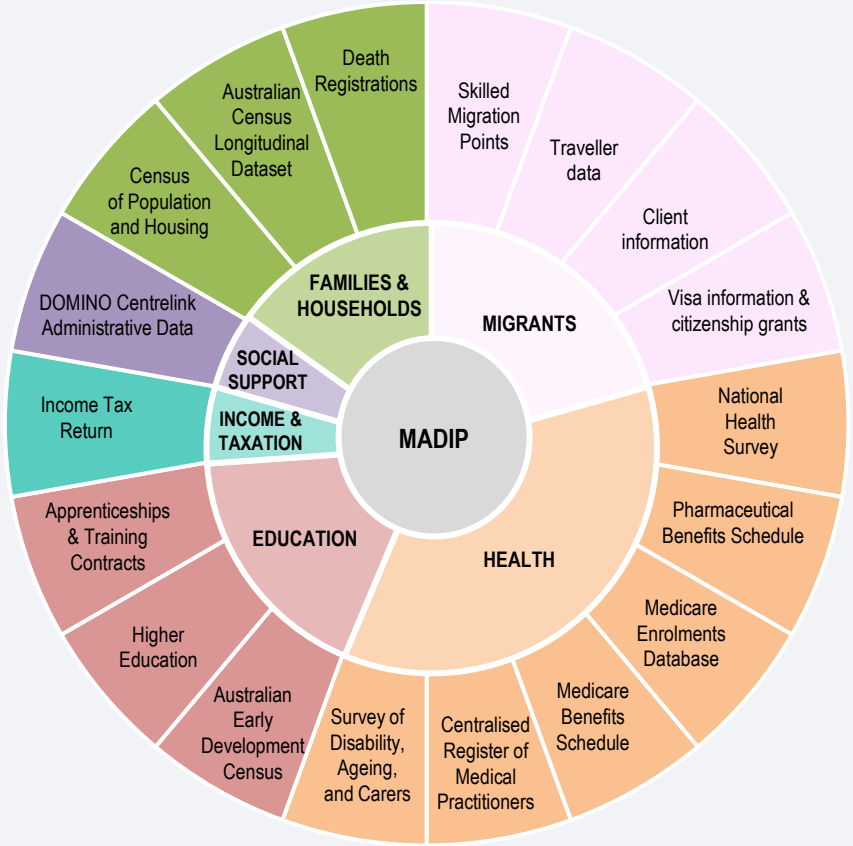
- Australian Bureau of Statistics
- Australian Taxation Office
- Department of Health
- Department of Social Services
- Department of Education, Skills and Employment
- Services Australia

Nationally important datasets



MADIP data is securely held by the ABS – access is only made available to approved researchers for approved purposes.

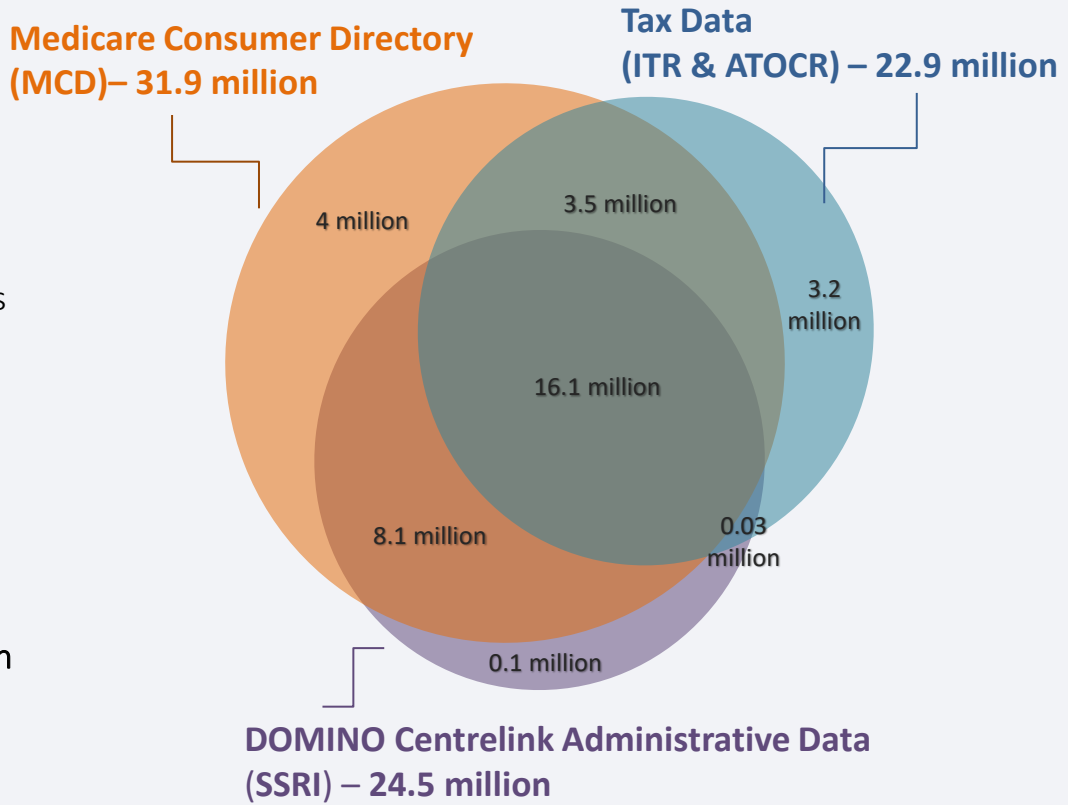
MADIP - Data


MADIP
FAMILIES & HOUSEHOLDS
MIGRANTS
SOCIAL SUPPORT
INCOME & TAXATION
EDUCATION
HEALTH
Death Registrations
Australian Census Longitudinal Dataset
Census of Population and Housing
Skilled Migration Points
Traveller data
Client information
Visa information & citizenship grants
DOMINO Centrelink Administrative Data
Income Tax Return
Apprenticeships & Training Contracts
Higher Education
Australian Early Development Census
Survey of Disability, Ageing, and Carers
Centralised Register of Medical Practitioners
Medicare Benefits Schedule
Medicare Enrolments Database
Pharmaceutical Benefits Schedule
National Health Survey

MADIP – Infrastructure: Person Linkage Spine (Version 3, May 2020)

Spine contains
35.2 million elements

- The Person Linkage Spine is the base dataset to which all other person datasets are linked
- All datasets which are linked to the spine are indirectly linked to each other in a secure and consistent way that protects the privacy of individuals.
- Each element representing a single person resident in Australia at any point between 2006 - 2019



MADIP - Governance

MADIP Operating Model

MADIP arrangements are governed by a set of principles and procedures, approved by the MADIP Board in the MADIP Operating Model.

MADIP Board

Responsible for the operation and strategic direction of MADIP

MADIP Board members (SES officers)

- Australian Bureau of Statistics
- Australian Taxation Office
- Department of Education, Skills and Employment
- Department of Health
- Department of Social Services
- Services Australia

MADIP Working Group

MADIP - Data Custodians capability

DATA SUPPLY:

- Data needs are determined between the ABS, data custodians, and users
- The ABS and data custodians agree on an appropriate data supply agreement. Data custodians determine other internal governance requirements.
- Data custodians are responsible for the preparation of source data and metadata for sharing
- ABS has procedures in place to remediate supply of unsolicited data if this occurs

ACCESS:

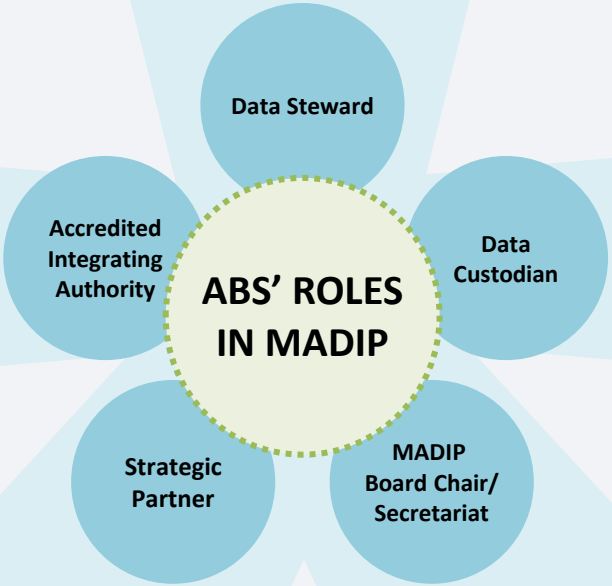
- Data custodian approval is sought for the use of their provided data for every proposed project
- Access to MADIP microdata is subject to approvals by the ABS in its stewardship role **and** the relevant data custodians

MADIP BOARD MEMBERSHIP:

- Data custodian role responsibilities are retained regardless of MADIP Board membership
- The role of MADIP Board membership going forward is to drive the strategic direction of the asset - membership is optional for data custodians
- The Board will consider proposals for new members at its upcoming meetings

MADIP - ABS capability

Our responsibility is to protect and use data for the benefit of the entire community



We have the infrastructure and capability to undertake high risk data integration projects for statistical and research purposes

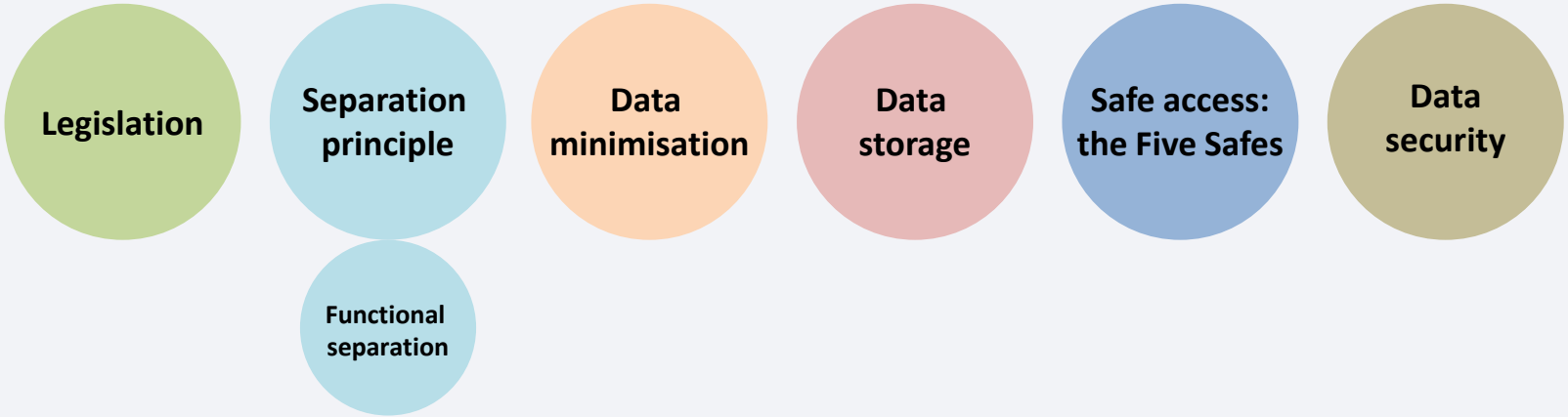
We manage the use, disclosure, protection, and data quality of ABS data used for data integration

We help guide the strategic direction and future of MADIP

We coordinate the key MADIP governance body

MADIP – Protections to keep data safe

We have many protections in place to keep your data safe, including:



Legislation

The Census and Statistics Act 1905

Information cannot be released in a manner likely to enable the identification of individuals and organisations

- ABS and seconded officers are legally bound to uphold the confidentiality of MADIP information.
- The MADIP is collected under this Act, which requires the ABS to publish and disseminate compilations and analyses of statistical information and to maintain the confidentiality of information collected

The Privacy Act 1988

- Data collection, use, and disclosure is authorised by law
- There are protections for data against unauthorised access and disclosure, or loss.

Other agencies' legislation

- Each agency is authorised by law to collect personal information as part of its core functions, to share that information with the ABS for MADIP in order to use it for policy analysis, research, and statistical purposes.

Australian Government standards

- Strong security arrangements for all IT systems: Australian Government Information Security Manual
- Strict control of access to premises: Commonwealth Protective Security Manual

Future Data Availability and Transparency Act (DAT)

When passed, this Act will:

- Promote better sharing of public sector data
- Build trust in use of public data
- Dial up or down appropriate safeguards
- Maintain integrity of the data system and
- Establish institutional arrangements.

The separation principle

SEPARATE TRANSFER

Receipt of analytical information and personal identifiers is in **separate files at separate times.**

SEPARATE STORAGE

Personal identifiers are **stored securely and separately** from the analytical information.

SEPARATE ACCESS

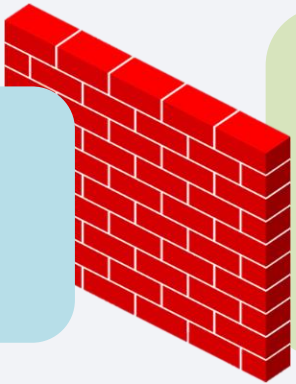
No-one can access both personal identifiers and analytical information at the same time.

Each person working on the project is assigned a role, and is **only able to access the information necessary to perform that role**.

Personal identifiers

Such as:

- Name
- Address
- Date of birth



Analytical information

Such as:

- Occupation
- Income
- Health services use
- Types of Government payments

Functional separation

A person working on a project can only hold **one role at a time**.

This means that **personal information and analytical information cannot be accessed at the same time**, and no person can ever see all of your information together at any point in the process.



Librarian

The ***Librarian*** prepares information for linkage



Linker

The ***Linker*** links information together



Assembler

The ***Assembler*** creates files for analysis



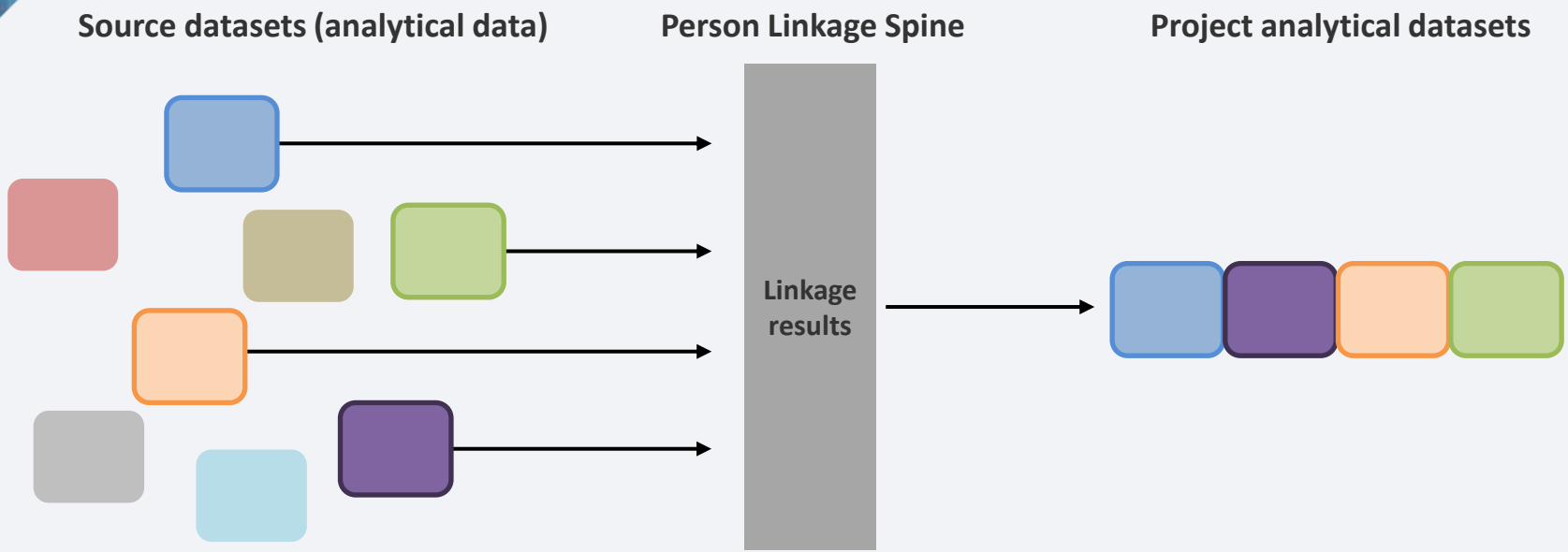
Analyst

The ***Analyst*** analyses linked information

Only accesses data without direct identifiers

MADIP data linkage, assembly and storage

Separate datasets are able to be brought together for specific projects



SEPARATE STORAGE of analytical data

Mechanism for **linkage**

Assembly of analytical data

Access to analytical data in ABS DataLab

MADIP data access and release

Five Safes Framework

- Safe people
- Safe projects
- Safe settings
- Safe data
- Safe output

Is the researcher authorised to access and use the data appropriately?

Is the data to be used for an appropriate purpose?

Does the access environment prevent unauthorised use?

Has appropriate and sufficient protection been applied to the data?

Are the statistical results non-disclosive?



MADIP data security

- IT security arrangements that conform with government standards for information security (the **Australian Government Information Security Manual)**
- **Regular independent reviews** of security arrangements and an ongoing program of **security audits**
- Staff **security checks** and **restricted access** to data
- **Logging** and **monitoring** of access and use of information
- Clear accountabilities and **risk management** processes
- Information is only combined in a **secure environment** within the ABS, by a dedicated team
- Assessed under the **Information Security Registered Assessors Program (IRAP)**
- **Secured** internet gateway
- Data supply via **secure transfer methods** (e.g. Secure Deposit Box)

MADIP – Data services and channels

Microdata – Standard, Custom, New Linkage

	MADIP Standard Products Ie: MADIP Modular Product, MADIP Basic Longitudinal Extracts	**Customised MADIP Extracts**	**New linkage to MADIP data**
What?	Data Items from: • Medicare • Pharmaceutical Benefits Scheme • Social Security • Income tax • Education • Deaths and Census data from 2011-2016	• Creation of new custom extracts, defined by the user	- Linkage of new data source to MADIP data (including interoperability with existing data assets)
When?	• Available now via the ABS DataLab	• Contact us to discuss your data needs	- Timing dependent on data specifications
Who?	• Government employees • Non-government researchers	• Government employees • Government contractors & Individuals sponsored by government • Academics • Researchers from public policy research institutes	

MADIP – Data services and channels

Welcome to
ABS DataLab



Access is only to approved researchers and ABS or government staff



Files remain in the secure ABS environment

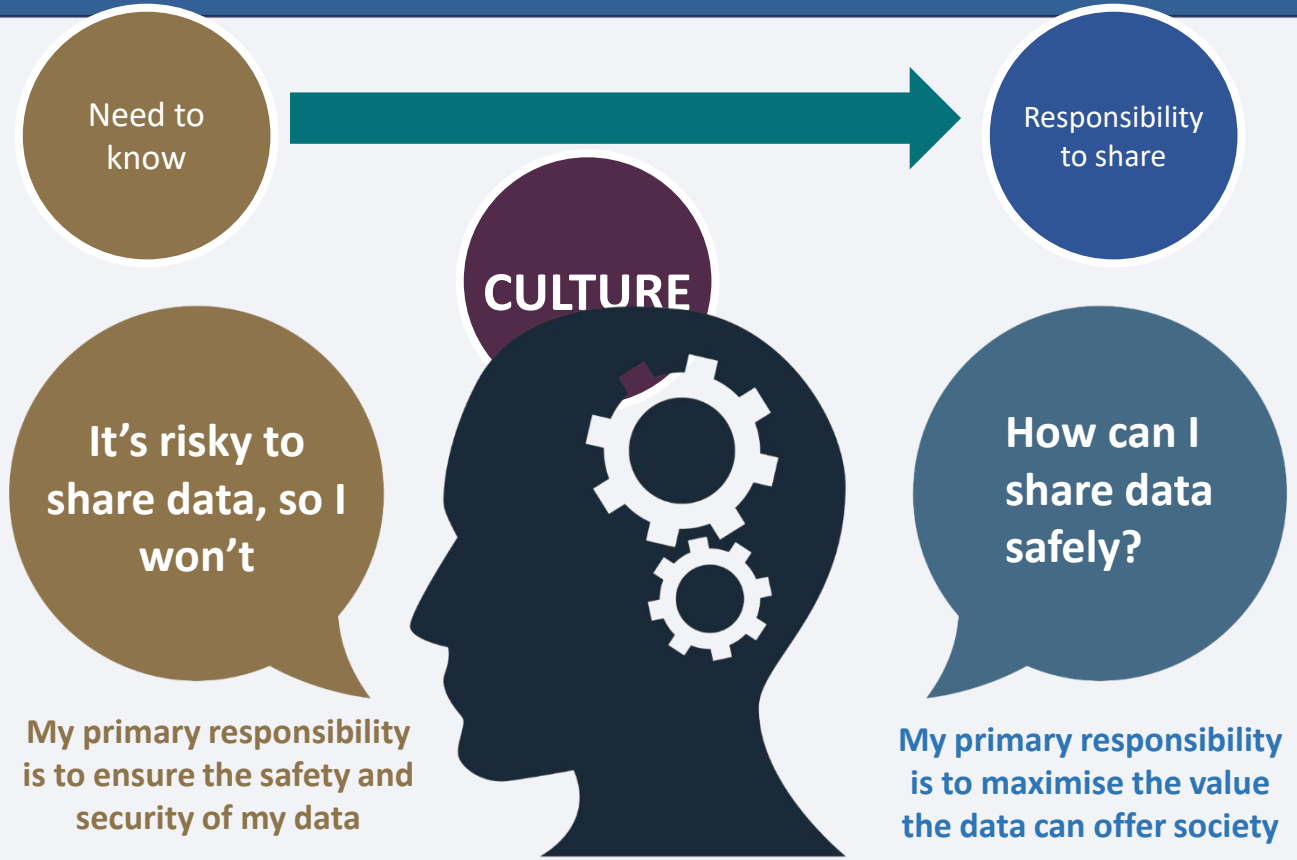


Analytical information does **not include name and address**



All **analytical outputs** are checked by ABS staff

MADIP – Challenges
Cultural change, trust and resourcing



Source: PM&C

Contact us

- Send general enquiries and project proposals to

  data.services@abs.gov.au

- Explore the list of MADIP research projects

  (or via the Data Integration pages on the ABS website)

- For more information contact

  Deepa Wright
  Director, Data Services
  deepa.wright@abs.gov.au

- Explore the MADIP products available

  Microdata: Multi-Agency Data Integration Project, Australia