Data Strategy
2020 - 2022

National Archives of Australia

MESSAGE FROM CIO

Data is a key enabler for the National Archives of Australia's important work of collecting the records of Australian government decisions and actions as evidence.

We do this to connect Australians with the nation's memory, their identity, and history. We collect, store, and provide access to data from across the Australian Government.

Data is a fundamental component of providing access to our collections. We want to reuse data in different ways to provide more opportunities to explore our collections.

Ultimately, the value of our data is not about how much we collect, but what we can learn from it.

To do all this, the National Archives needs a modern strategy for its data.

Through implementing this strategy, we will strengthen our data governance and digital capabilities. It will also help us achieve the principles outlined in our new policy, *Building Trust in the Public Record.*

Yaso Arumugam
Chief Information Officer
National Archives of Australia

EXECUTIVE SUMMARY

The National Archives of Australia's Data Strategy is a plan of action to improve our data maturity – how we use, manage, preserve, and access our corporate and collection data assets.

The strategy builds on our ongoing commitment to the Australian Government data agenda, and the policies and the procedures of the National Archives.

The Data Strategy is built on these five components:

➢ Data is F.A.I.R. (findable, accessible, interoperable, and reusable)
➢ Data is fit for purpose
➢ Supportive stakeholders and business owners
➢ Maximising data potential through innovative approaches
➢ Data drives decision-making to achieve improved outcomes

The goal for NAA Data Strategy is to ensure that the National Archives benefits from its data and data assets.

OUR ENVIRONMENT

The National Archives provides leadership in best-practice management of the official record of the Australian Government and ensures that Australian Government information of enduring significance is secured, preserved and available to government agencies, researchers and the community.

We continue to lead the development of standards and frameworks for best practice information management.

The Data Strategy builds on our ongoing commitment to the Australian Government data framework and agenda.

Our decisions are also guided by and aligned with the policy *Building Trust in the Public Record*, and the National Archive's I&T Strategy and Information and Data Governance Framework.

Australian Government Data Framework agenda

Building Trust in the Public Record

I&T Strategy

Information and Data Governance Framework
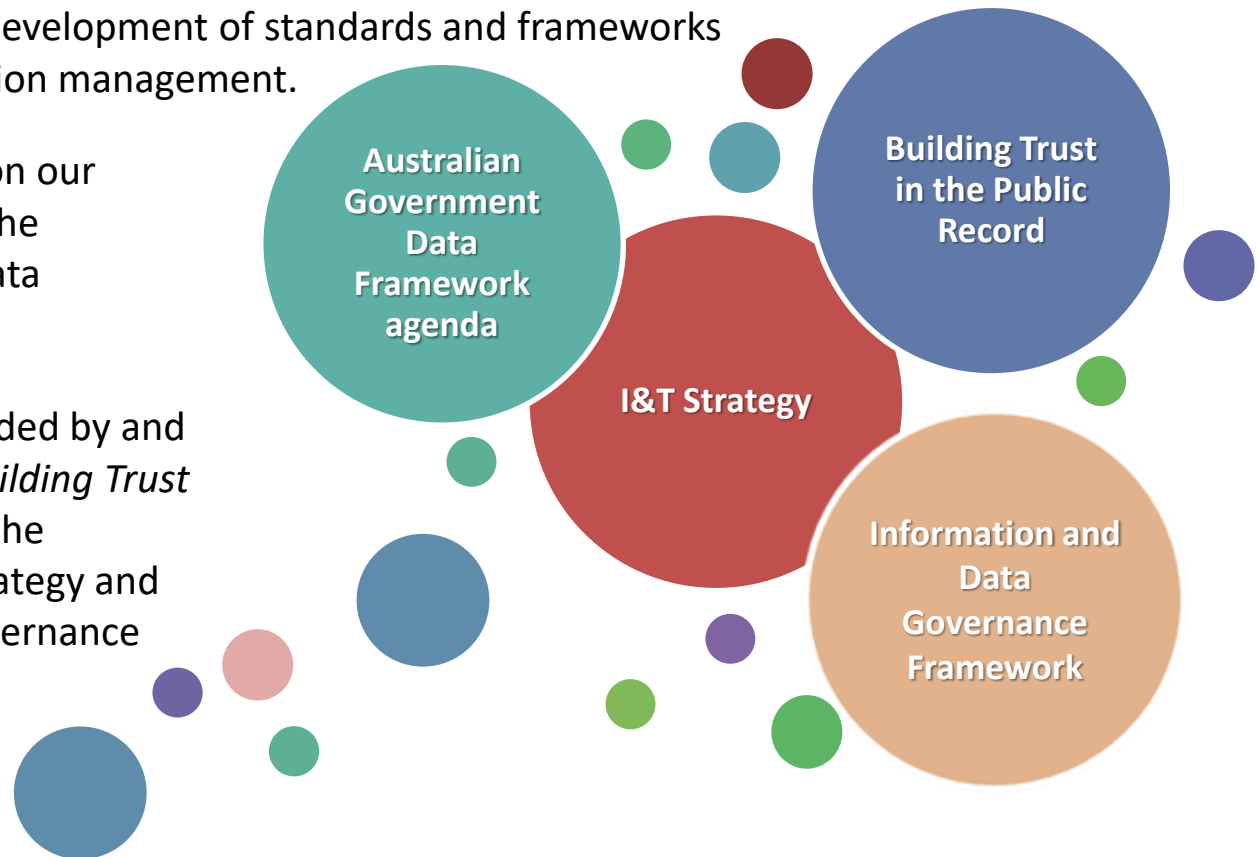
PURPOSE

The National Archives' Data Strategy aims to strengthen our capability as an exemplar agency for information and data governance.

It has been developed to define a plan and a path to maximise data maturity and achieve quantifiable benefits from our data assets for all stakeholders.

We will achieve this by:

➢ maximising data maturity
➢ prioritising and understanding data in our processes, systems and projects
➢ creating insights through data to support research, public engagement and improve operational efficiency
➢ improving data interoperability and accessibility
➢ investigating and leveraging of Artificial Intelligence (AI) for better use of data for innovative outcomes, such as the Business and Research and Innovation Initiative (BRII) Challenge
➢ integrating data systems such as the Integrated Archival Management System (IAMS), and
➢ engaging with practitioner-level Commonwealth data initiatives as an agency with experience in data tools and advice.
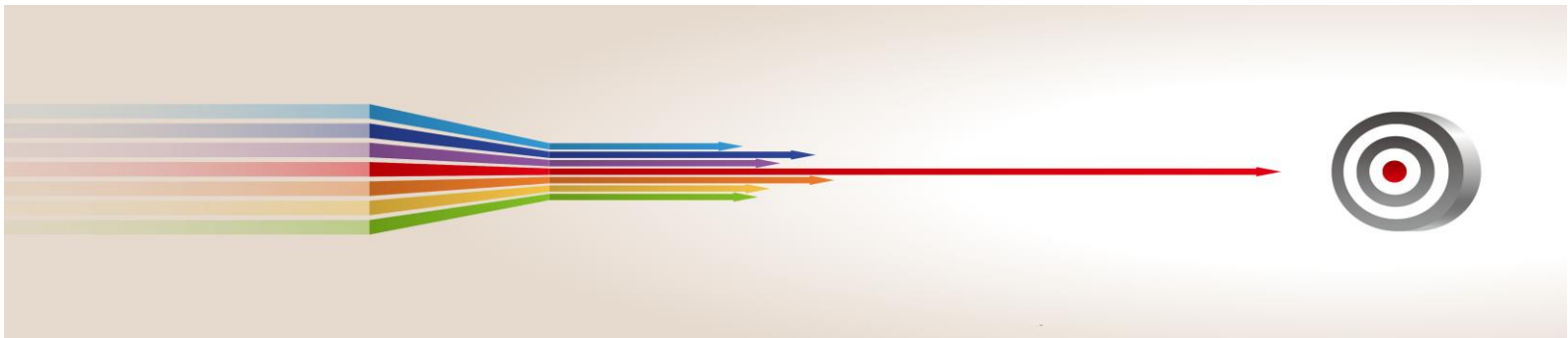
SCOPE

The National Archives Data Strategy Data applies to the:

Corporate data assets still in active or semi-active use
to inform, implement, document and communicate our business activities, decisions and approvals.

Collection data to be preserved and used on demand
as the most valuable records of the Australian Government designated for permanent retention by the National Archives.

DEFINITIONS

Data may be numeric, spatial, statistical, structured or unstructured information (unprocessed or processed) communicated in a form, such as by text, numbers or multimedia. Data is capable of being analysed, processed and communicated by an individual, a computer or electronic device. At the National Archives, **data** is a type of information.

A **dataset** is a structured collection of data generally associated with a unique body of work, a particular subject, or created for a specific purpose such as the WWII service records from the collection.

A **database** is an organised collection of data stored as a single or multiple datasets. These are generally stored and accessed electronically from a computer system that allows the data to be easily accessed, manipulated, and updated. A business system may comprise of one or more databases such as the Recordkeeping System or RecordSearch.
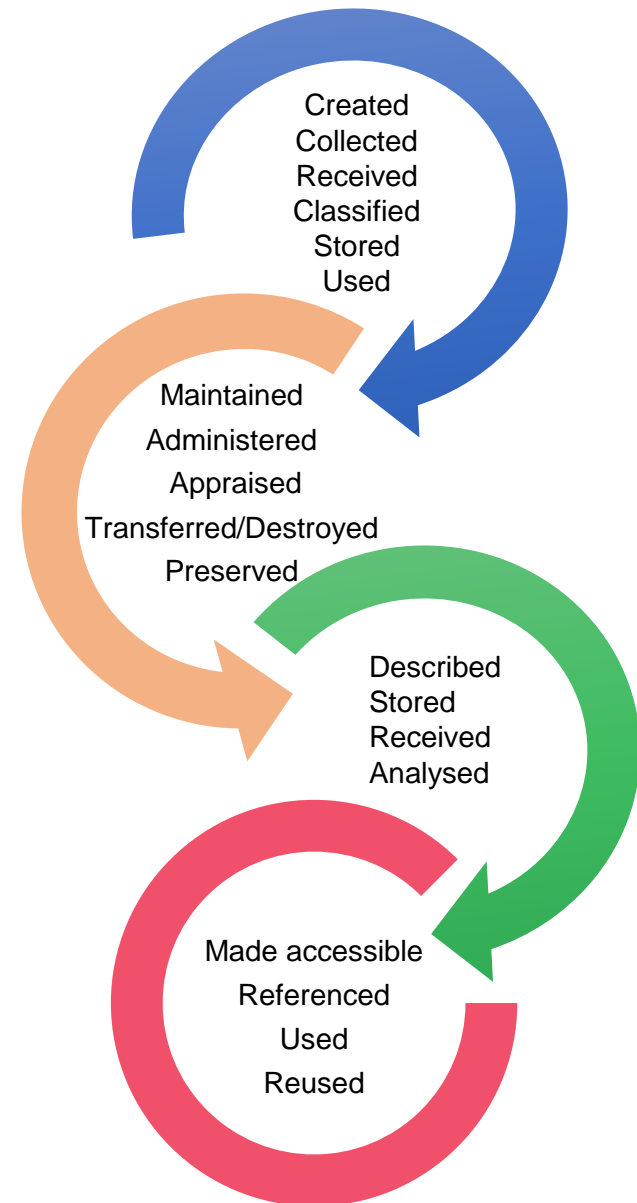
VISION

The National Archives' vision is to be a world leading archive in this digital age.
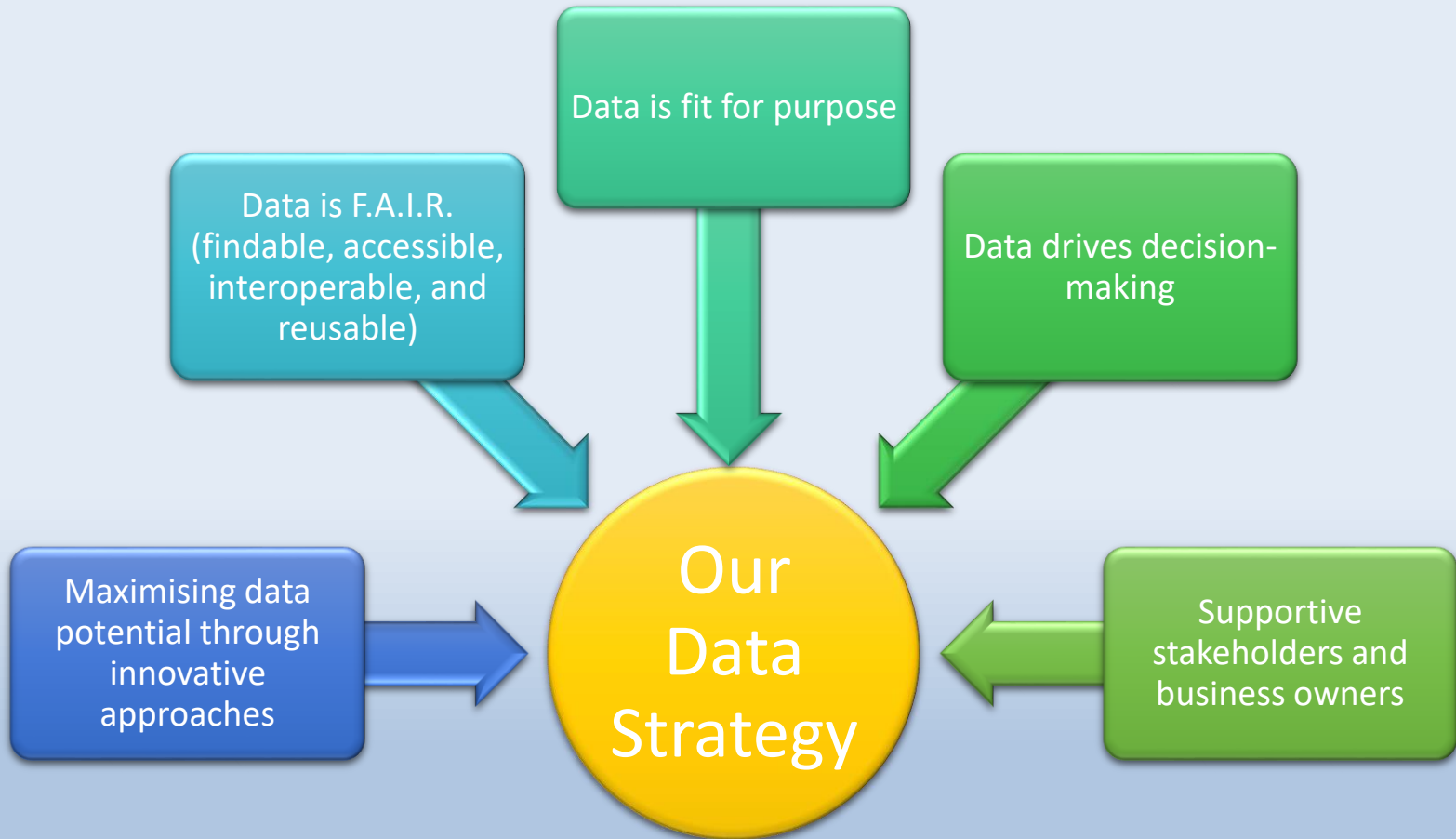
The National Archives' Data Strategy defines a plan to maximise data maturity and ensure a data focus.

DATA LIFECYCLE

Data exists within a lifecycle which includes processes that create or obtain data, those that move, transform, and store it and enable it to be maintained and shared, and those that use or apply it, as well as those that dispose of it. Throughout its lifecycle, data can be cleansed, transformed, merged, enhanced, or aggregated. As these processes occur often new data is created which form interconnecting processes.

Created
Collected
Received
Classified
Stored
Used

Maintained
Administered
Appraised
Transferred/Destroyed
Preserved

Described
Stored
Received
Analysed

Made accessible
Referenced
Used
Reused

1. DATA IS AN ASSET

Data is F.A.I.R.
(findable, accessible, interoperable, and reusable)

WHAT THIS MEANS…

Data has value for the National Archives.

Data meets the F.A.I.R. Data Principles.

Data is available in a format and timeframe needed by business owners or external stakeholders.

Data assets have adequate and standardised descriptive metadata.

Data is used and shared safely and consistently.

INITIATIVES

➤ Continue to digitise analogue records for accessibility
➤ Testing of application programming interfaces (APIs) to link data sources, breaking down data silos
➤ Continue to develop staff data capability through the learning programs offered for National Archives staff through PM&D

WHAT SUCCESS LOOKS LIKE

❑ Increased access to the National Archives' collection
❑ APIs are investigated to aid data discoverability and access for external stakeholders
❑ A culture of continuous data learning across the National Archives
❑ Continued support of Chief Information Governance Officer (CIGO)

2. DATA IS TRUSTED

Data is fit for purpose

WHAT THIS MEANS…

Data has value and we need to rely upon it.

Data quality is the accuracy, completeness, and fitness for purpose, in a given context.

Poor quality data makes data difficult to locate and use.

Data that is trusted and relied upon, requires a risk based approach for security and protection.

INITIATIVES

➤ Access and implement high capacity and durable data storage for example Enterprise Grade Storage (EGS)
➤ Agreement on data fields in RecordSearch to ensure consistent data input and output

WHAT SUCCESS LOOKS LIKE

❑ Datasets have associated data quality statements that are linked to data
❑ Automated tooling is used to reduce the manual effort of metadata maintenance
❑ Routine quality checking of data is undertaken

3. DATA IS MAINTAINED

Supportive stakeholders and business owners

WHAT THIS MEANS…

There is a shared understanding of data across the National Archives.

There is visibility of data – where it is stored, who has responsibility, and who is the custodian.

The content and context of data is known, and can be verified and be understood.

Maintaining data quality enables a consistent data representation.

INITIATIVES

➢ Development and maintenance of a Digital Assets Register (a single source of truth) of data assets and systems
➢ Assign data owners and business owners to ensure accountable data use and ownership.
➢ Data use rules are established and communicated to improve processes.

WHAT SUCCESS LOOKS LIKE

❑ A Digital Assets Register documents the value of the National Archives assets to business users and include high level management considerations such as sensitivities associated with the data.
❑ We know where data expertise is and draw on it when required
❑ Staff understand what data is, and see themselves as data users

4. INNOVATION THROUGH DATA

Maximising Data potential through innovative approaches

WHAT THIS MEANS…

Innovation enables business transformation.

Innovation is experimentation. It is creation, disruption and improvement. Innovation through data leads to discoveries that may change the way we work.

INITIATIVES

➢ Liaise and collaborate with stakeholders undertaking projects using Augmented Intelligence (AI) and Machine Learning (ML) to improve process and practices such as the Business Research and Innovation Initiative (BRII)
➢ Digital Archives Innovation and Research (DAIR) section develop software Preservation Policy
➢ Modernise Search and Discovery tools

WHAT SUCCESS LOOKS LIKE

❑ We collaborate on data findings and initiatives internally and externally such as BRII
❑ Software Preservation Policy created by DAIR
❑ Implementation of the archival management system, the Integrated Archival Management System (IAMS)
❑ Implement new Search and Discovery layer for RecordSearch

5. BUILD DATA INSIGHTS

Data drives decision making to achieve outcomes

WHAT THIS MEANS...

Enabling new knowledge which enhances and optimises the value of the National Archives collection and corporate data

Continued exploration of emerging forms of data technology and practices

Building data analytics capabilities to deliver insights and foresight

➢ Build and explore data visualisation capabilities
➢ Keep informed on data analytical technologies for future work

WHAT SUCCESS LOOKS LIKE

❑ Data insights are produced to support research and public engagement
❑ Access to more data which leads to actionable insights

Building Trust in the Public Record: Managing Information and Data for Government and Community

Manage information assets strategically with appropriate governance and reporting

Implement fit for purposes information management process, practices and systems

Reduce areas of information management inefficiencies and risk
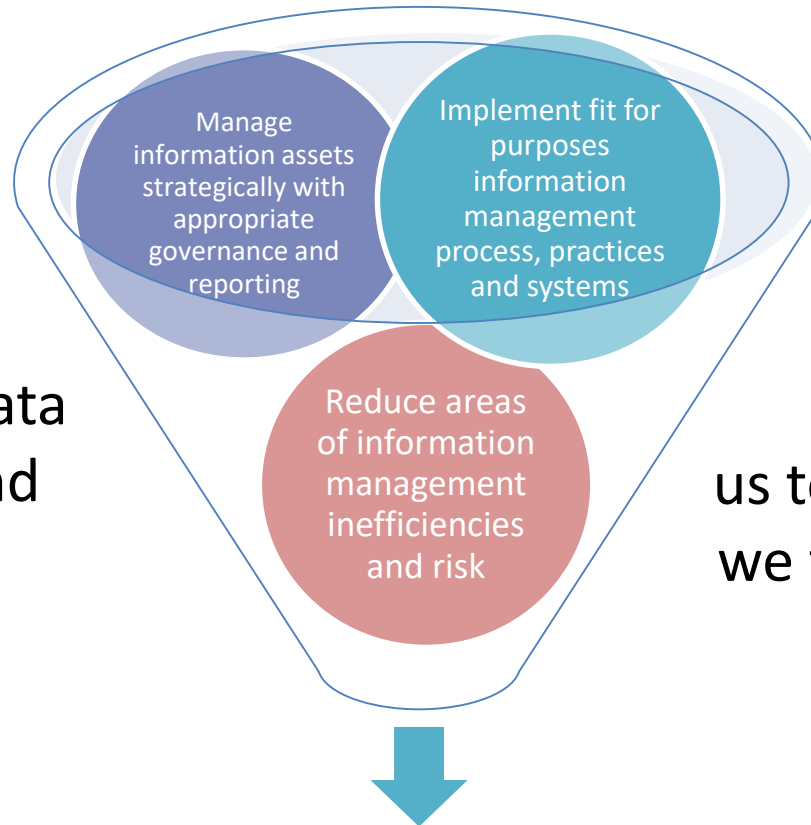
Applying these **three key requirements** will enable us to transform how we think about data

Data is F.A.I.R. (findable, accessible, interoperable, and reusable)

Data is fit for purpose

Supportive stakeholders and business owners

Maximising data potential through innovative approaches

Data drives decision making to achieve outcomes

INTERNAL ALIGNMENT

The Data Strategy has been developed in consultation with staff across the National Archives

➡ The Data Strategy aligns with our internal frameworks, strategies, policies and plans

➡ The Data Strategy initiatives intersect with those in other strategies as we work towards the same vision

I&T Strategy 2019-2022 — Information and Data Governance Framework — Information Security Policy (ISP) — Archives' Risk Management Framework — Digital Assets Register — Business Continuity Policy and Plan — Our Way: Aboriginal and Torres Strait Islander Protocols — Information and Data Management Policy
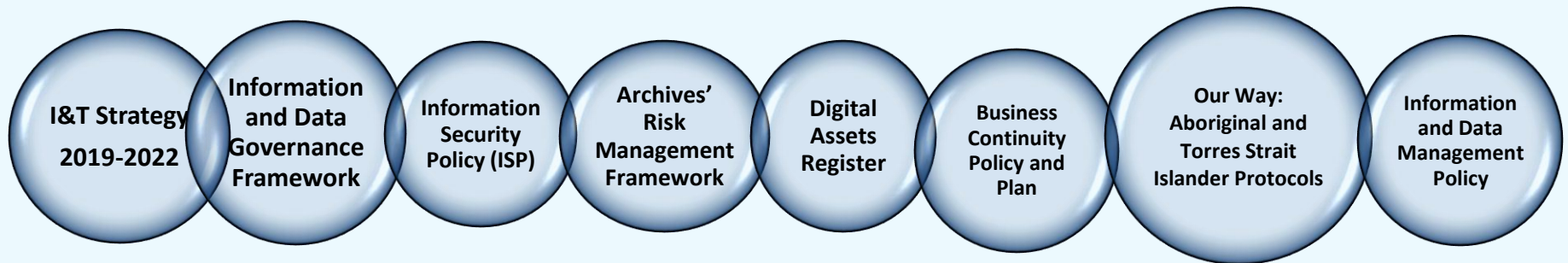
ALIGNMENT WITH GOVERNMENT

Further Government commitments shaping data

- Building Trust in the Public Record: Managing Information and Data for Government and Community

- The Foundational Four – Starting an ongoing data improvement journey

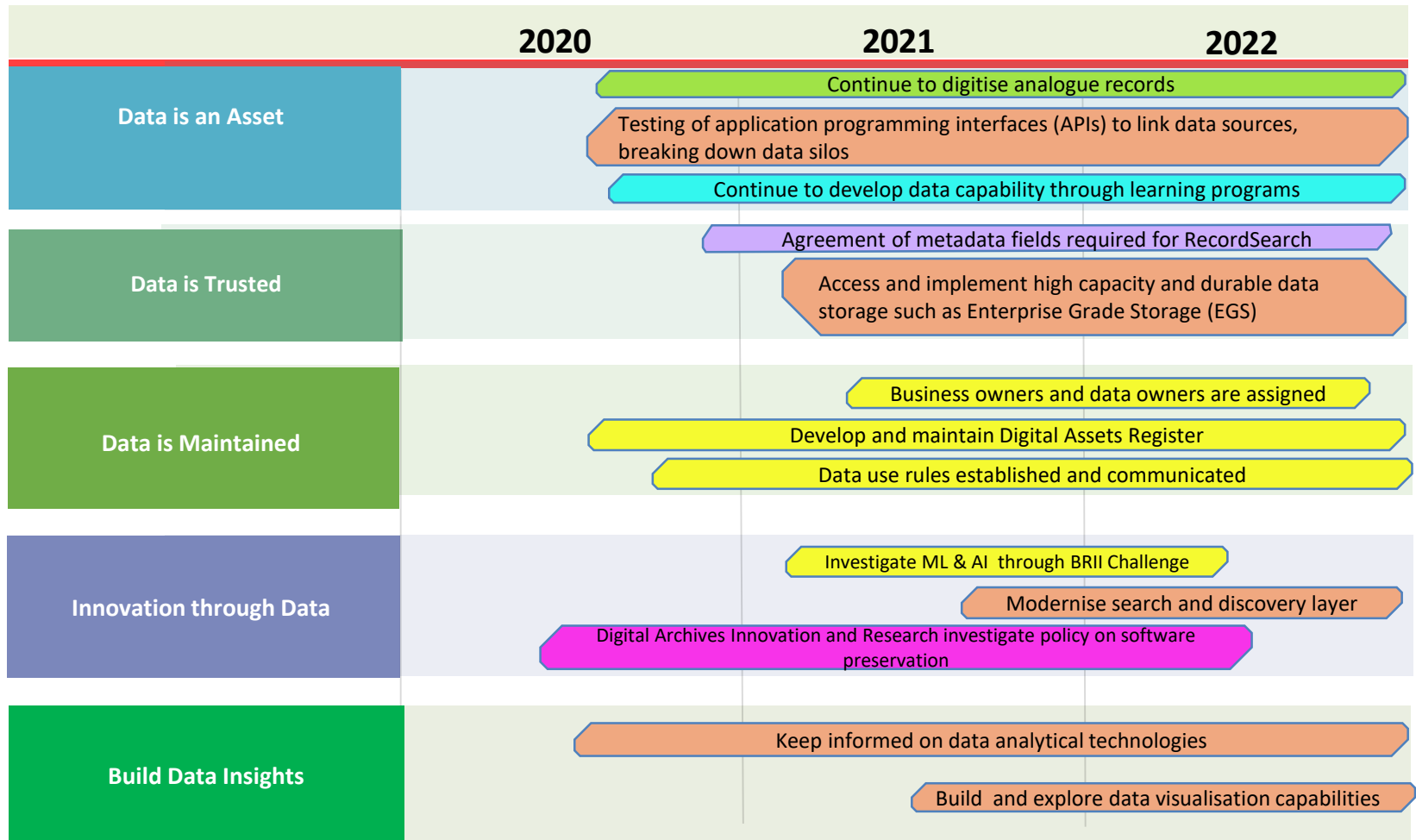- Protective Security Policy Framework (PSPF)

- Australian Government Information Security Manual (ISM)

- Data Availability and Transparency Bill 2020 (Exposure draft)

- Public Data Policy Statement

- Best Practice Guide to Applying Data Sharing Principles

- A Framework for Data De-identification

- Delivering for Australians - A world-class Australian Public Service: The Government's APS public reform agenda

- Privacy Act 1988 and Australian Government Agencies Privacy Code

- Australian Privacy Principles

ROADMAP

	2020	**2021**	**2022**
Data is an Asset		Continue to digitise analogue records	
		Testing of application programming interfaces (APIs) to link data sources, breaking down data silos	
		Continue to develop data capability through learning programs	
Data is Trusted		Agreement of metadata fields required for RecordSearch	
		Access and implement high capacity and durable data storage such as Enterprise Grade Storage (EGS)	
Data is Maintained		Business owners and data owners are assigned	
		Develop and maintain Digital Assets Register	
		Data use rules established and communicated	
Innovation through Data		Investigate ML & AI through BRII Challenge	
		Modernise search and discovery layer	
		Digital Archives Innovation and Research investigate policy on software preservation	
Build Data Insights		Keep informed on data analytical technologies	
		Build and explore data visualisation capabilities	

Legend:
- IG Section Plan 2020-21
- National Archives' 2020 Strategic Plan
- 2019-2022 I&T Strategy
- DAIR Section Plan 2020-21
- PM&D Digital Archival Program
- IAMS Project