



Australian Government
Office of the National Data Commissioner

The Foundational Four

Starting an ongoing data improvement journey



Contents

Executive summary	1
Introduction.....	3
Why is public sector data important?.....	3
The <i>Foundational Four</i>	4
Using this Guide.....	5
Leadership – A senior leader is responsible and accountable for data across the agency.....	6
What is the role of a Senior Data Leader?	6
Putting Leadership into practice – Appoint a Chief Data Officer or equivalent.....	7
Where can I find more information?.....	7
Data Strategy – An agency has a clear vision and plan for using data to achieve objectives	8
What is a Data Strategy?	8
What is the difference between a Data Strategy and an Information Strategy?.....	8
Putting Data Strategy into practice – Develop and implement a Data Strategy.....	9
Contents	10
Endorsement and implementation	13
Implementation.....	13
Where can I find more information?.....	14
Governance – Mechanisms exist to oversee data management.....	15
What is Data Governance?.....	15
What is the relationship between Data Governance and Data Management?.....	15
Putting Governance into practice – Develop a Data Governance Program.....	16
How is a Data Governance Program established?.....	16
Where can I find more information?.....	19
Asset Discovery – Data assets have been identified and recorded	20
Why identify data assets?.....	20
What is the difference between a data inventory, register, catalogue and repository?.....	21
Putting Asset into practice – Establish a discoverable Data Inventory	21
What makes a data inventory effective?.....	21
Next steps: from inventory to catalogue.....	24
Where can I get more information?.....	24
Next Steps	25

Copyright Statement PM&C

The Foundational Four

© Commonwealth of Australia 2020

ISBN 978-1-925364-22-4 The Foundational Four (online)

Copyright Notice

With the exception of the Commonwealth Coat of Arms, this work is licensed under a Creative Commons Attribution 4.0 International licence (CC BY 4.0) (<https://creativecommons.org/licenses/by/4.0/>).



Third party copyright

Wherever a third party holds copyright in this material, the copyright remains with that party. Their permission may be required to use the material. Please contact them directly.

Attribution

This publication should be attributed as follows:

© Commonwealth of Australia, Department of the Prime Minister and Cabinet, The Foundational Four.

Use of the Coat of Arms

The terms under which the Coat of Arms can be used are detailed on the following website:

<https://pmc.gov.au/cca>

Other uses

Enquiries regarding this document are welcome at:
www.datacommissioner.gov.au/form/contact

Executive summary

Government data is valuable, but it needs to be managed effectively to realise its value. While some agencies are more advanced in their management and use of data, others are still developing their capability.

Improving data capability can be a daunting task, with a range of risks and costs to consider. Without simple advice, it can be difficult knowing where to begin.

Better data management is not an end in itself, but is needed to meet the government's policy and service delivery ambitions. These goals have been articulated in a number of policy statements and reports, including:

- The Productivity Commission's [Data Availability and Use Inquiry](#)
- The [Public Data Policy Statement](#)
- The [Data Availability and Transparency legislation](#)
- [The Digital Continuity 2020 Policy](#)
- The [Open Government National Action Plan](#)
- [Delivering for Australians. A world-class Australian Public Service: The Government's APS reform agenda](#)

These reviews and policy statements all recognise that the APS needs to shift its culture and build data capability. Culture is an outcome of organisational settings, behaviours and activities. Each of the *Foundational Four* have an important role in improving data capability and through that, shifting culture. Though it will require a contribution from APS-level staff, champions for change are needed at the Senior Executive Level together with leadership from more experienced APS agencies.

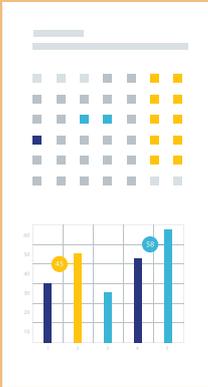
The *Foundational Four* are:

- **Leadership:** A senior leader is responsible and accountable for data across the agency
- **Strategy:** An agency has a clear vision and plan for using data to achieve objectives
- **Governance:** Mechanisms exist to oversee data management
- **Asset Discovery:** Data assets have been identified and recorded

The APS has been investing in strengthening its data leadership strategies to better use and manage data. Agencies with more mature practices recommend that priority be given to investing in data governance. If effective data governance mechanisms are not in place, the remaining elements of the *Foundational Four* will have little or no impact.

How well the interests of all Australians are served through delivery of government programs and policy, easy to access services, and an innovative and strong research and development sector, is ultimately dependant on the public service meeting the challenge of effectively managing data.

We ask agencies to test themselves against the *Foundational Four*. For those who have some or all of the elements in place, it is a good opportunity to reflect on how effective they are and whether there are opportunities for improvement. If you don't have all of the elements in place, then action is needed.



Introduction

Why is public sector data important?

Government agencies are custodians of a great deal of data that reflects life in Australia: from people to animals, plants and fungi; from schools to businesses to hospitals; from weather patterns to Australia's location on planet Earth.

Data held by government agencies is important. It shapes the lives of Australians through:

- enabling everyday business and service operations of agencies
- shaping policy and program development
- assisting policy and program evaluation
- improving service delivery
- delivering insight to Australian people, businesses, environments and culture
- facilitating research and development, and
- providing transparency of government activities

Government decisions must be based on the best evidence available and good quality data is critical to good quality evidence. It is important agencies that are the custodians of data manage it well so it can help deliver the best outcomes for Australia. It will also help build public confidence and trust in government's ability to hold, handle and use data, as well as meeting community expectations that data assets are managed professionally, securely and effectively.

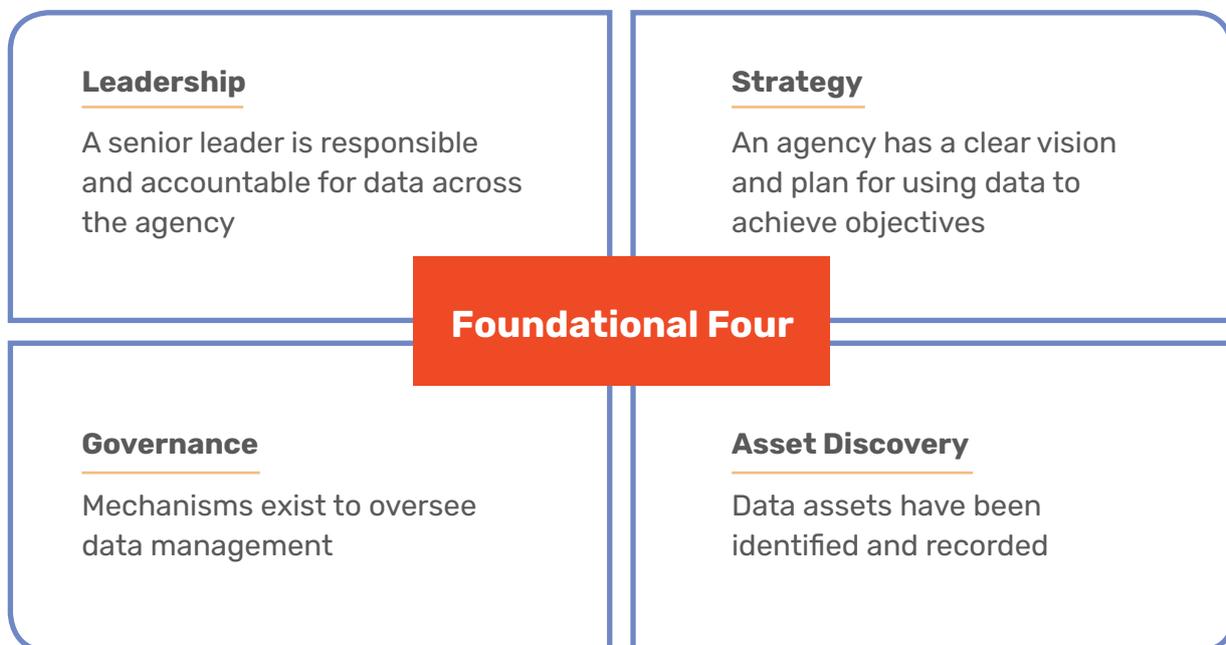
To help achieve this, the Office of the National Data Commissioner has developed the *Foundational Four*.

The *Foundational Four*

Most agencies are aware of the importance of data, and many are already taking steps to improve their data practices. What can be difficult is knowing where to start, particularly if there are multiple issues within an agency that need addressing.

The *Foundational Four* provides a clear starting point and some simple, useful steps to take towards managing data more effectively. They have been chosen based on their importance as foundational first steps, so that future initiatives can build upon them. The *Foundational Four* complement existing advice and experience from across the Australian Public Service and draw on international best practice approaches.

Figure 1 - The '*Foundational Four*'



Using this Guide

The aim of the *Foundational Four* is, primarily, to provide agencies who are beginning their data journey with a starting place to improve their data practice. They can also be used as a reminder for agencies looking to refresh their practices. For agencies that have already implemented the *Foundational Four*, these guidelines can serve as a positive reminder of progress.

It is important that agencies adapt the practices described in this guide to their own context, culture and business/policy drivers. There is no one-size-fits-all approach to the *Foundational Four* and there is no set order in which you need to implement the *Foundational Four*.

Cultural change is an ongoing journey and the *Foundational Four* are not an end point. They are the foundations for agencies to start moving in the right direction. Once the *Foundational Four* are implemented agencies can continue to improve their data culture and capability through other activities.



Leadership

A senior leader is responsible and accountable for data across the agency

What is the role of a Senior Data Leader?

A Senior Data Leader is a business role responsible and accountable for enhancing the strategic value of an agency's data assets. If data is managed successfully as an asset and its use is maximised, agencies can achieve their vision and deliver their functions more efficiently and effectively. In most agencies, this role should ideally be assigned to a member of the Senior Executive Service (SES).

One of the key responsibilities (and challenges) of a Senior Data Leader is to be able to identify, measure and communicate the tangible business outcomes that have been achieved through changes in an agency's use and management of its data. Evidence of the impact of data is important to create a culture that understands and values data.

The Senior Data Leader will:

- lead, and be accountable for, the development and implementation of the agency Data Strategy
- set the standards and approach to data management
- improve data governance

- oversee and coordinate data functions such as data literacy, data quality, data architecture, business intelligence, data security and data protection
- drive data culture change
- build positive relationships with other agency leaders
- support and promote staff who work with and manage data

The Senior Data Leader also plays a role in improving community trust in the government's handling of public sector data. The Senior Data Leader should ensure appropriate safeguards to protect data are in place, as well as improving transparency of the agency's data activities. A Senior Data Leader may need to consider agency data activities not simply in terms of 'what can we do?' (e.g. from a legal or IT systems perspective), but also 'what should we do?' (e.g. from an ethical or community expectations perspective).

Having a designated Senior Data Leader does not decrease the importance of the roles that other staff play within an agency. Leadership for data change can come from all levels, all business areas and within all projects. A designated Senior Data Leader provides a central point of leadership to coordinate and focus data initiatives and decision-making.

Putting Leadership into practice – Appoint a Chief Data Officer (CDO) or equivalent

Ideally, all agencies with significant data holdings will have a Senior Data Leader as a separate role (i.e. a designated Chief Data Officer). However, creating a new specialist role is not always possible. The functions of the role are of greater importance than the title or where the position sits in an organisation, and a CDO role could be assigned to the senior executive responsible for one or more related functions such as:

- information services
- data analytics
- digital strategy
- data management
- Privacy Champion and/or Privacy Officer (Australian Government Agencies Privacy Code)
- Chief Information Governance Officer (Digital Continuity 2020)

How does a Chief Information Officer differ from a Chief Data Officer?

A Chief Information Officer (CIO) holds a different role in an agency to a Chief Data Officer (CDO), though in the absence of a specific staff member to undertake a CDO or equivalent role, a CIO can often take on some of their responsibilities.

Traditionally, a CIO leads the IT strategy for an organisation and is responsible for managing technology assets. The remit of a CIO does not traditionally include responsibility for data outside the IT systems. A Chief Data Officer (or equivalent) is responsible for the effective and efficient use and management of an agency's data holdings. A CDO is not a technological role and should not be focused on the technology side of data. However, being aware of an agency's data infrastructure, including IT, is important. The CDO aims to understand the strategic value of data, the safeguards required to protect it and what is required to maximise its use. The role should also guide technological decisions where the protection and security of data are concerned.

The CIO and CDO should work collaboratively as data relies on technology and a great deal of data is associated with technology. Together these two roles can resolve a variety of data-related issues, bringing the perspectives of their expertise to the solution.

Where can I find more information?

Below are a few resources about CDOs in particular and the important role they play in modern organisations:

[Chief Data Officer in Government: A CDO Playbook \(Deloitte\)](#)

[A Month in the Life of a Chief Data Officer \(Dataversity\)](#)

[What is a Chief Data Officer? \(Dataversity\)](#)

[The Chief Data Officer Rises \(TDAN\)](#)

[What is a Chief Data Officer \(CIO\)](#)

[The Chief Data Officer Playbook \(IBM\)](#)

[Chief Information Governance Officer \(National Archives of Australia\)](#)

[Australian Government Agencies Privacy Code \(Office of the Information Commissioner\)](#)

Questions to ask:

- **How does a Senior Data Leader role fit with our existing governance structures?**
- **What type of Senior Data Leader role makes sense for our agency?**
- **What will the role be called?**
- **What specific duties and responsibilities will the role have?**
- **How will the role be formally endorsed?**
- **How will the role collaborate with other senior executives, executive and APS level staff to advocate for better use and management of data?**

Data Strategy

An agency has a clear vision and plan for using data to achieve objectives

What is a Data Strategy?

A Data Strategy is a detailed plan for an agency's data and how it can be best used to meet the agency's broader business objectives and priorities.

It can provide an opportunity to refresh current practices around data collection and use, and to implement new technologies and innovations to get the best use out of an agency's data. It identifies a target state for agency data use and a pathway to get there.

A well-rounded Data Strategy should consider all data processes, what is required at each step of those processes and where improvements need to be made and efforts focused. It may also consider other measures, such as skills and capability building, and cultural change activities.

Ideally, all agency data would be within scope of a Data Strategy. However, due to resource constraints, an agency may choose to initially limit the data in scope of the strategy. For example, an organisation may choose to focus on its program data, but not its corporate data (e.g. Human Resources data).

The scope of the Data Strategy can be expanded in a second or third-phase Data Strategy as an agency's practices mature.

What is the difference between a Data Strategy and an Information Strategy?

A Data Strategy is closely related to an Information Strategy but narrower in remit. An Information Strategy 'describes your agency's planned approach to information management to meet current and future organisational needs and regulatory requirements'¹. It includes all of an agency's information, including data and all files and records.

Data is just one of many different types of information an agency will hold. A Data Strategy only focuses on data and associated practices, but the two strategies should align. Some agencies may choose to have a single Information Strategy incorporating information and data, others may choose to have separate but related strategies.

1 <https://www.naa.gov.au/information-management/information-governance/establishing-information-governance-framework/developing-information-management-strategy>

Putting Data Strategy into practice – Develop and implement a Data Strategy

There are some key considerations when developing a Data Strategy:

Value proposition

Before you embark on a Data Strategy you should understand the value proposition: 'Why are we doing this?' and 'What's in it for me?'. For some areas of an agency, the need for and benefits of a Data Strategy are obvious. However, this is not the case for everyone, particularly when a Data Strategy may divert resources from one stream of work to another. This is where a value proposition can be beneficial. It will clearly articulate what the benefits of the Data Strategy are for all parts of the agency.

The different perspectives, values, concerns and priorities of stakeholders will need to be considered to develop value propositions. Spending time defining what value looks like for stakeholders and tailoring messages is important to gain buy-in.

State analysis

To inform your Data Strategy you need to have some understanding of what your agency's data capability is; your agency's current state.

Once you know your current state you can identify what you want your future state to look like and what pathway can get you there.

Agencies may use a formal maturity model to conduct their state analysis (see 'Where can I find more information?' for details) or they may use some other method, such as workshops or focus groups. A risk analysis may inform your state analysis.

A data audit or inventory may also help shape the Data Strategy as it allows agencies to determine the extent of their data holdings and how they are stored, managed and used.

The following article outlines the components of a number of data maturity models:

[Data Governance Maturity Models \(Lights on Data, 2018\)](#)

Engagement, consultation and communication

A Data Strategy aims to drive transformational change across an agency. Engagement and consultation will be essential to develop buy-in from staff and drive change through the strategy.

A stakeholder analysis can be undertaken to determine who will be involved and impacted by the Data Strategy, and the level of influence each stakeholder has. Other elements to consider are the frequency with which you engage, using a co-design approach, ensuring you engage across all staffing levels, using a variety of engagement channels to communicate progress, and demonstrating you have heard the input from staff including how it has shaped the strategy.

Size

There is no right 'size' for a Data Strategy. It needs to be as long or as short as necessary to achieve desired outcomes. An agency can focus on small improvements, or big transformation. The scale of the Data Strategy will depend on resources, culture, appetite and what is realistically achievable for an agency.

Timeframe

There needs to be a balance between choosing a timeframe that is too short to achieve meaningful change, or one that is too long to respond to the changing data landscape. Typically, an agency's Data Strategy would span three years, with regular 'check-ups' or progress reports to assess how the Data Strategy is performing.

Contents

The contents of a Data Strategy can vary, but some useful elements to include are:

- Vision statement
- Alignment with Government data policy
- Alignment with internal policies / strategies
- Strategic objectives
- Initiatives/areas of focus
- Measurable outcomes/indicators of success
- Roadmap

Vision statement

This outlines where an agency desires to be and what the Data Strategy should ultimately achieve. This should be short, catchy and future-focused. If an agency already has a strong vision statement, it may be helpful to relate the Data Strategy to this.

Examples of Data Strategy vision statements include:

- Data is F.A.I.R (findable, accessible, interoperable, and reusable)
- Service delivery is enhanced through maximising the value of our data
- Data drives decision making to achieve policy outcomes
- Public trust is enhanced through transparent collection and use of data

Strategic alignment with Government data policy

There are many initiatives within Government that may help to shape a Data Strategy. As government agencies, it is important that any activities related to the use of data reflect the direction of current Government commitments. A Data Strategy should be compliant with mandatory government legislation and policies, and align to the greatest extent possible to non-mandatory policies.

Relevant Government data policies include:

- [Public Data Policy Statement](#)
- [Government response to the Productivity Commission's Data Availability and Use report](#)
- [Digital Continuity 2020 Policy](#)
- [Protective Security Policy Framework](#)
- [Australian Government Information Security Manual](#)
- [Privacy Act 1988 and Australian Government Agencies Privacy Code](#)
- [Data Sharing Principles](#)
- [De-identification Decision Making Framework](#)
- [Delivering for Australians - A world-class Australian Public Service: The Government's APS public reform agenda](#)

Internal policies/strategies

A Data Strategy may struggle to get traction if developed in isolation. It should complement and leverage existing internal strategies, policies and roadmaps, such as the Corporate Plan. This is particularly relevant in the area of ICT. The Data Strategy should consider the technical requirements needed to support its implementation such as data warehousing capability. These requirements will need to align with your agency's ICT or enterprise architecture directions.

Examples of relevant internal policies include an agency's:

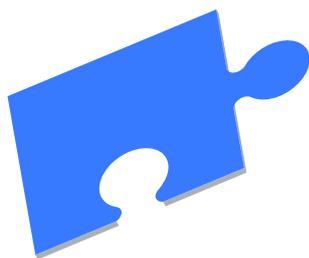
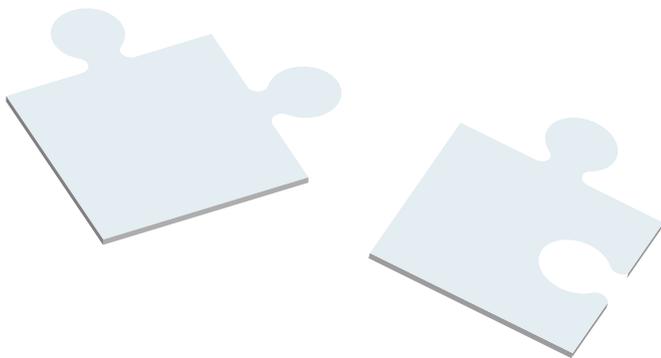
- Mission or vision statement
- Business strategy
- Corporate plan
- Enterprise Architecture strategy / roadmap
- ICT strategy
- [Information Management Strategy](#)
- [Information Governance Framework](#)
- People / Workforce / Capability strategy
- Privacy policy

Strategic objectives

Strategic objectives define what you want to get out of the Data Strategy. They should be supported by explanatory text so it is clear to all staff members what they mean in practice. The strategic objectives should align with the Vision Statement and also reflect the outcomes of the State Analysis.

Examples of strategic objectives include:

- Build first-class analytics capability
- Manage and use data to improve client outcomes
- Create a data culture
- Future-proof data infrastructure
- Break down data silos
- Ensure transparency in data collection, access, and use
- Manage privacy proactively using a privacy-by-design approach



Initiatives/Areas of Focus

These are the heart of the Data Strategy. The Initiatives or Areas of Focus identify what an agency is going to do as part of its Data Strategy.

There are many different areas an agency can focus on as part of the Data Strategy. The particular areas of focus will depend on the outcomes of the state analysis, the scale of data activities within an agency and the resources available.

Examples of Initiatives/Areas of Focus include:

- Improving data governance
- Establishing career pathways for data scientists
- Building skills and capabilities
- Ensuring consistent data management across the data lifecycle
- Supporting IT (infrastructure, architecture)
- Remediating legacy data
- Facilitating data migration
- Improving process management and transformation
- Developing and implementing standards and metadata
- Improving the culture and environment around data management
- Ensuring appropriate data security
- Establishing privacy governance to guide management of personal information

Each Initiative or Area of Focus should be associated with one or more of the Strategic Objectives. If an initiative cannot be aligned to a Strategic Objective, then it is likely not a priority.

It is important to identify who within the agency is responsible and accountable for each of the initiatives. It may be a position within a team or it may involve collaboration between different areas. To ensure accountability and responsibility, ownership of initiatives must be clearly indicated so all staff know who is responsible and accountable for each initiative and it can be factored into business planning.

Example 1:

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

Strategic Objective 1: Ensure transparency in data collection, access, and use

Initiatives to meet Strategic Objective 1:

- Increase data management skills across the data lifecycle
 - Responsibility: Learning and Development team
- Release data as open where appropriate
 - Responsibility: IT, Data teams, Publishing team
- Use data platforms to increase data accessibility
 - Responsibility: Data team, IT, Web Services team
- Ensure all data is catalogued to increase discoverability and promote reuse
 - Responsibility: Publishing team, Cataloging team, IT, Data teams

Example 2:

Vision Statement: Data drives decision making to achieve policy outcomes

Strategic Objective: Build first class data analytics capability

Initiatives:

- Establish career pathways for data scientists
 - Responsibility: Learning and Development Team, Senior Management, Supervisors
- Establish a Community of Practice for analytics
 - Responsibility: Data teams, Agency Data Champions
- Sponsor data 'hack-a-thons' involving analysts and policy officers
 - Responsibility: Senior Management, Agency Data Champions

Measurable outcomes/Indicators of success²

A good Data Strategy should include ways to measure success or progress. These performance indicators are useful evidence for management to ensure ongoing support, as well as identifying where changes or resources are needed.

The measures should directly reflect the initiatives within the strategy. Measures can be qualitative or quantitative, noting that data management lends itself well to quantitative measures. Quantitative indicators are also easier to measure and compare across the lifespan of the Data Strategy. They can also be persuasive for senior management.

Regular 'check-ups' or review points should be incorporated into the roadmap so progress can be tracked and adjustments made if objectives are not being met.

External stakeholders and data users may also be considered when developing success indicators.

Examples:

Quantitative

- 50% of data remediated
- 65% of staff completed data management training
- 95% data catalogued internally

Qualitative

- Staff understand the value of data to the agency
- Staff are confident in their ability to access data required for their job
- Agency is referred to as a 'data leader' by stakeholders

Roadmap

A Roadmap is a complementary product to the Data Strategy. The Roadmap details the various initiatives and the timelines for completion. This allows an agency to 'map out' the flow of activities and ensure they occur

² There are some established frameworks for developing measures of success that agencies can use, such as OKR (Objectives and Key Results), KPI (Key Performance Indicators) or SMART (Specific, Measureable, Achievable, Relevant, Time-bound) goals.

in a systematic order, particularly if activities are dependent on others. It is useful to identify those responsible and accountable for each initiative on the Roadmap and the relevant timelines, so people can see visually who is responsible and accountable for what and when.

The Roadmap provides an overarching structure to tasks and will form a core component of work programs moving forward. How each individual task is completed is up to the individuals and areas responsible for that initiative.

Endorsement and governance

To influence cultural change, the Data Strategy must be supported by the agency executive. This can be done through a number of methods, such as clearance and endorsement processes (e.g. Executive Committees), but also through forewords/introductions, branding and launches, or a combination of these.

Governance and oversight of the Data Strategy is important to ensure progress and accountability. Agencies may wish to set up a new governance structure specifically for the Data Strategy, or may use existing governance structures.

Implementation

There are a number of factors that are critical to successful implementation of a Data Strategy:

- Ensure the Data Strategy is aligned with the agency's mandate, vision and/or business objectives. This will ensure the Data Strategy remains relevant and change will occur in conjunction with other agency strategies/directions.
- Have a champion, or several champions, within an agency who can bring everyone along on the journey and promote the Data Strategy. If your agency has a CDO and/or Data Champion then they may be the best placed person to lead this work.
- Aim high, but be realistic about what can be achieved. Some strategic objectives and initiatives may be suited to future versions of a Data Strategy.

- Change and improvements take time. Don't be disheartened if it takes some time for initiatives to gain traction within your agency.
- Showcase successes. Focusing on successes, even small ones, provides evidence of progress. This is a compliment to those who have made the changes and also helps bring 'nay-sayers' along on the journey as they can see benefits of the Strategy.
- Review the Data Strategy periodically and don't be afraid to adjust initiatives if they aren't working. A Data Strategy is a medium-to-long term plan and it's normal for plans to require adjustments as time passes and new information is discovered.
- Test the environment. Large agencies, in particular, may have differing levels of data usage and maturity, dependent on roles, business processes and systems that govern different work areas.
- Data Strategies will need to iterate over time and agencies should consider transitioning to a new Strategy as practices improve.

Finally, it is useful to share and compare Data Strategies with other agencies. The challenges faced by one agency are likely faced by others. When beginning work on a Data Strategy it can be helpful to approach other agencies for a copy of their strategy. To help foster transparency and trust in government, agencies may also consider making their Data Strategy, or portions of it, publically accessible as appropriate. There are a number of Australian Government agencies and international agencies, which have already established and published a Data Strategy:

- [Department of Industry, Innovation and Science](#)
- [Geoscience Australia](#)
- [Department of Employment, Skills, Small and Family Business](#)
- [Australian Securities and Investment Commission](#)
- [Office for National Statistics \(UK\)](#)
- [Statistics New Zealand](#)

Where can I find more information?

There are interest groups across the APS that can provide useful advice:

- Data Champions Network, facilitated by Department of the Prime Minister and Cabinet.
- Data Governance Special Interest Group, facilitated by National Archives of Australia.
- The Australian Government Linked Data Working Group (AGLDWG) is a community of Commonwealth Government experts and champions, with invited non-voting participation of individuals, corporations and other entities.
- There is an Open Data Tool Kit on data.gov.au, which provides information about publishing open data, including guidelines for an Open Data Strategy.
- Queensland Government has provided information on their audit of information management maturity. This is one example of a maturity assessment in practice.

The National Archives of Australia runs the Check-Up Plus for agencies every year as part of annual reporting requirements. Agencies could use the results of this survey to inform improvements within their Data Strategy.



Questions to ask:

- **Who should lead or create the Data Strategy?**
- **Who needs to be convinced of, and promote, the benefits of our Data Strategy?**
- **Who needs to be involved in developing our Data Strategy?**
- **What is an appropriate and achievable vision for data in our agency?**
- **How does our Data Strategy align with government data policy?**
- **How does our Data Strategy align with internal policies / strategies?**
- **What is the current state of agency data maturity and capability? How will this assessment be undertaken?**
- **What are the overarching strategic objectives related to our agency that will guide the strategy?**
- **What initiatives or focus areas are most relevant and achievable for our agency?**
- **Do the vision statement, strategic objectives and initiatives or focus areas align well with each other?**
- **Has responsibility and accountability for each initiative or focus area been allocated and accepted?**
- **Have measures and indicators of success or progress been defined in a meaningful way?**
- **Who needs to endorse our Data Strategy for it to take effect?**
- **Does a roadmap exist to guide the implementation of the Data Strategy?**
- **What activities in the strategy will support cultural change within the agency?**
- **How will we promote the Data Strategy and ensure continued engagement?**

Governance

Mechanisms exist to oversee data management

What is Data Governance?

Data Governance is the ‘exercise of authority and control (planning, monitoring and enforcement) over the management of data assets’³.

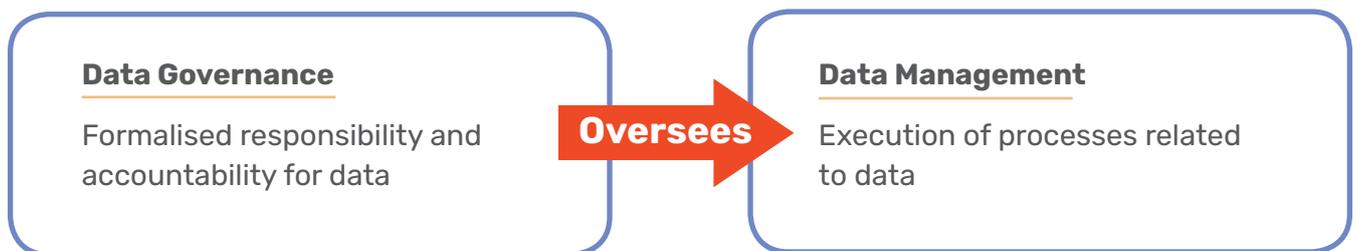
What does this mean? Data Governance is the oversight mechanisms that formalise responsibility and accountability for data and its management in an agency. The purpose of Data Governance is to ensure data is properly managed, according to policies and procedures developed by the agency, and that people understand their responsibilities. It enables an agency to understand, manage and reduce risks, including security and privacy risks relevant to the data it holds.

What is the relationship between Data Governance and Data Management?

Data Governance is a process for guiding and overseeing Data Management. It is a hierarchical relationship.

Data Governance oversees the activities related to data. It ensures there is accountability for data and that the appropriate policies are followed when working with data. Data Management refers to the agency’s practices that relate to the actual data itself. It is not a single activity, but a process related to all aspects of data usage, including acquisition, processing, storage, security, releasing, archiving and disposal. Effective data management ensures data is accessible, discoverable and useable when you need it.

Figure 2 – Relationship between Data Governance and Data Management



To ensure there is robust and consistent oversight of Data Management processes, agencies can formalise the governance structures through a Data Governance Program.

3 DAMA International (2017).

Putting Governance into practice – Develop an Ongoing Data Governance Program

Establishing an ongoing Data Governance Program may involve:

- identifying what level of governance is required
- establishing oversight structures
- developing policies and procedures for data
- defining roles and responsibilities
- identifying agency data monitoring and audit topics to assess how well data practices are being implemented
- identifying risks and management/mitigation strategies

Developing an ongoing Data Governance Program demonstrates that an agency understands that its data has value and needs to be looked after accordingly. A Data Governance Program must have a clearly defined purpose for it to be beneficial. It should be established and formalised to the degree that is appropriate for an agency's needs and culture.

A Data Governance Program does not necessarily require a whole new governance structure be created. Indeed, a Data Governance Program is easier to establish and generally more effective if it is integrated with, and leverages, existing governance processes.

How is a Data Governance Program established?

There are several key elements to a successful Data Governance Program:

- Appropriate style of governance
- Senior executive support
- Staff buy in
- Scope of governance program
- Defined roles and responsibilities
- Relevant policies and procedures
- Clear implementation and review plan, including a Data Governance Framework

Consider the style of governance

There are many different styles of governance. A style that suits one agency may not suit another; a mixed approach may suit somewhere else. The following are some examples of Data Governance styles that could work within an agency:

Examples of Data Governance styles include:

- **Centralised Data Governance** – All data governance activities are overseen and set by a central body or authority, across all business areas.
- **Replicated Data Governance** – The same method of governing data is adopted by multiple business areas, but there is no overarching authority.
- **Federated Data Governance** – The data governance is coordinated across multiple business areas to ensure consistency.

(For more information see: [DAMA International Data Management Body of Knowledge](#))

- **Non-Invasive Data Governance** – The data governance is applied to existing policies and procedures, without imposing new ones or unnecessary burden. The principle is to augment existing practices. Responsibilities are formalised where they already exist and not assigned to anyone who is not already undertaking the role in their day-to-day work (Seiner, 2014).

(For more information see:

[The Data Administration Newsletter](#))

Senior Executive support and champions

The Data Governance Program needs to be supported by Senior Management, particularly a Senior Data Leader, so it receives the attention it needs to be successful. Having many data champions within an agency, beyond the Senior Data Leader, helps bring all staff on the journey and promote the benefits of a Data Governance Program.

Staff buy-in

Data Governance is most successful when it is collaborative. Areas with data holdings should be supportive of a Data Governance Program.

However, there may be resistance, particularly where the Program is perceived as creating more work or diverting resources. Consultation and communication can help and with persistence, resistance will often dissipate when benefits are shown.

A Data Governance Program should be implemented through a collaborative effort between:

- those responsible for formalising Data Governance (the Data Governance team or Data Policy team where they exist, or the equivalent if such teams have not been created within an agency),
- business areas (those responsible for managing and working with data on a day-to-day basis),
- corporate areas (those responsible for other levels of corporate governance)

Through working collaboratively, the team responsible for Data Governance can discover what data practices are currently in place, determine whether those practices are fit-for-purpose and where there are gaps. This team can then work with the business areas to formalise and enhance their processes, accountabilities and responsibilities.

Scope of Governance Program

A Data Governance Program must help ensure that an agency's data can support business activities and agency goals. Having a Business Strategy and/or a Data Strategy in place will outline the direction an agency wishes to take in improving its data practices and gives purpose to the Governance Program.

It is important to consider Data Governance in the context of the existing governance processes within an agency. Many agencies will already have governance processes for information and records management, information and IT security, project implementation, organisational decision making, etc. A Data Governance Program can 'lean on' these existing processes and build upon them to ensure data is effectively governed.

It is also important to consider what needs to be governed: what data processes need oversight, what practices are currently in place, what existing practices need formalising. Using a Data Strategy and supporting state analysis may assist in this process.

In considering data governance, the question of 'how much control or governance?' needs to be answered. The level of oversight to apply to the data depends on its sensitivity and the legislation, rules and guidelines that apply. Does the agency deal primarily with sensitive data that needs to be tightly controlled so decisions about the data need more consideration and clearance?

Can decisions about data be made locally within data teams? Is the culture of the agency risk-avoidant, or is it more comfortable with risk and in need of additional ongoing oversight?

Understanding what needs to be governed and how much governance is required will help an agency determine what 'style' of governance is needed; it must be the right fit for the culture and context of the agency.

Define roles and responsibilities

Defined roles and responsibilities for data ensure accountability. The definition and implementation of roles and responsibilities should be a collaborative effort with the business and potentially IT areas. It may also be more effective to formalise the role of someone who is already actively managing data assets for which they are responsible (i.e. formally appoint them as 'Data Steward'), rather than imposing a role upon someone else.

Agencies may wish to consider training to support staff assigned to certain roles, for example, data stewardship training for data stewards. This helps staff understand the responsibilities, accountabilities and expectations of their role.

Develop policies and procedures

Policies and procedures outline the rules for working with data. These should be developed in collaboration with areas responsible for implementing and complying with the policies. It will ensure they are workable and their implementation is supported. It may be helpful to engage with other agencies when developing policies and procedures. You can learn from their experience,

it can lead to similar approaches for similar data types or situations being adopted, and it will help promote consistency across the public service.

Data policies can be high level, but due to the varied nature of data work, some procedures may need to be more specific and localised.

Examples of areas that may require formalised data policies and procedures include:

- Data access
- Data security
- Data storage
- Data retention and disposal
- Metadata management
- Cataloguing data
- Data quality management
- Handling sensitive data
- Releasing data

Data Governance Framework

A Data Governance Framework is a written document that defines the context for governing data within an agency. It may reference:

- legislation relevant to an agency's business processes
- whole of government data policies and initiatives
- strategies and policies internal to the agency
- committee structures and reporting relationships
- roles and responsibilities of various data stakeholders
- the policies and procedures related to data, developed by the agency

The Data Governance Framework should align with an agency's Information Governance Framework. The Information Governance Framework will be broader than the Data Governance Framework as it will include all of an agency's information and records (emails, documents, etc.). The Data Governance Framework will specifically focus on data. Some agencies may choose to have a single Information Governance Framework covering information and data, others may choose to have separate but related frameworks.

A Data Governance Framework can follow the same structure as an Information Governance Framework. The National Archives of Australia's Information Governance Framework⁴ provides a useful structure for a specialised Data Governance Framework.

The Data Governance Framework should be endorsed by Senior Management to provide authority for the Framework.

Progressive implementation

Implementing an ongoing Data Governance Program will take time and involves cultural change, so a progressive roll out may be beneficial. The Data Governance Program can be developed as a whole, but then implemented piece by piece. For example, if consistent data cataloguing has been identified as an area to improve, an agency can start by developing a policy to mandate the registration of all data an agency holds and then develop consistent procedures for cataloguing data.

To ensure successful implementation, the Data Governance Program will need to be promoted and staff kept informed about the program and its benefits. This task is easier if the program has been developed through a collaborative and consultative approach.

4 <https://www.naa.gov.au/information-management/information-governance/establishing-information-governance-framework>

Implementation will also be easier if the program has received Senior Executive endorsement and there are champions within the agency to promote and drive the data governance agenda.

Periodically reviewing and evaluating the Data Governance Program helps to ensure it remains fit-for-purpose. It is important to test whether elements of the program are working effectively and continue to meet agency needs.

Where can I find more information?

[Australian Government Information Management Standard \(National Archives of Australia\)](#)

[Data Governance and Management \(National Archives of Australia\)](#)

[Information Management \(Queensland Government Chief Information Office\)](#)

[Data management and use: Governance in the 21st Century \(British Academy for the humanities and social sciences, and The Royal Society\)](#)

[Data Management Framework \(Office for National Statistics\)](#)

[New Approach to Data Governance \(Statistics New Zealand\)](#)

[Data Management Body of Knowledge \(DAMA International\)](#)

[Non-Invasive Data Governance \(Seiner, 2014\)](#)

[The Non Invasive Data Governance Framework – The Framework Structure \(Seiner,2019\)](#)

[Data Governance Operating Models Exposed \(LightsOnData, 2018\)](#)

[Working with Data \(Australian National Data Service\)](#)

Questions to ask:

- **Has accountability and responsibility for data and data initiatives in our agency been clearly defined?**
- **Are there senior executive officers responsible for, and championing, data?**
- **Who should be involved in the development of the Data Governance Program?**
- **Who needs to be convinced of the benefits of a Data Governance Program? Has staff support been secured?**
- **What will the role of mid-level staff, like data stewards and custodians, be?**
- **What needs to be governed?**
- **What governance processes are in place and can they be leveraged?**
- **What style of governance is appropriate?**
- **Has a governance framework been developed to support the implementation of the Data Governance Program?**
- **Have roles and responsibilities been clearly defined in the Data Governance Program?**
- **Have policies and procedures been clearly developed? Are they visible and easy to find?**
- **Who needs to endorse the Data Governance Program for it to take effect?**
- **Is there a clear plan and support for implementation?**
- **How can the Data Governance Program be best communicated?**
- **How will the Data Governance Program be reviewed for effectiveness and relevance?**

Asset Discovery

Data assets have been identified and recorded

Why identify data assets?

Difficulty finding and accessing data is an experience faced by many staff in many agencies. Often it is because agencies do not have a central repository or a 'single source of truth' of their data assets. There may be multiple and fragmented catalogues and systems, which do not always interact.

Ensuring data is discoverable has many benefits:

- It can prevent data duplication and reduce costs. Where there is a lack of visibility of data holdings across an agency, data assets can be reproduced multiple times, often at a significant monetary cost or data management burden to the organisation⁵. Duplicated data assets might also be stored in multiple locations across an agency, resulting in increased technical infrastructure costs, data security risks and management overheads.
- It promotes reuse of data for improved outcomes and a reduced unit cost of managing data. The more people that can find out about what data is available, the more opportunities there are to use that data in new or different ways. The resources involved in managing the data are then benefiting many more outcomes.

- It ensures that constraints on data use are known.
- It manages the risk of 'loss of corporate knowledge'. To make up for insufficient technology and a lack of discoverability, staff often rely on their own established networks and connections to search for and discover data (i.e. ask their colleague in another team where to find data). If these relationships did not exist, the data may never be discoverable to anyone outside the direct data managers or stewards⁶.
- It can make managing data quality easier. By not having an understanding of what data an agency holds, the ongoing management of data is difficult as data assets become out of date, or out of sync, with versions and transformations not being appropriately documented⁷. It is also difficult to justify the resources required to manage data assets if the full extent of an agency's data holdings are not known.

Data inventories, registers or catalogues make it easier to find, access, trace, standardise, use, analyse, manage, govern and communicate about data.

Agencies that are very mature in their data management capabilities may decide to automate much of their cataloguing and focus on developing the quality of the data they hold.

5 Data61 Web Geospatial Visualisation Research, 2019

6 Data61 Web Geospatial Visualisation Research, 2019

7 Data61 Web Geospatial Visualisation Research, 2019

What is the difference between a data inventory, register, catalogue and repository?

There are many terms to describe listing and storing data assets – the concepts are often distinct, but there is also overlap involved.

A **Data Inventory** or **Data Asset Register** is a list of data assets. An inventory should record basic information about a data asset, such as name, description, license, owner/custodian, reference period, etc.

A **Data Catalogue** is a piece of technology that, at the most basic level, enables a user to search for and locate data for their specific needs. It relies on well-described metadata to enable search functionality.

A **Data Repository** is a data storage location. A data repository can store data according to discipline (e.g. clinical data, ecological data), or it can be a single enterprise storage location for all of an agency's data.

A data catalogue requires a data inventory to populate it with information. A data inventory or data catalogue may reference a data repository as a storage location for one (or more) data assets described within the inventory/catalogue.

This document focuses on data inventories as a way agencies can identify and record the data resources they have. This can be built on to create a searchable data catalogue.

Putting Asset Discovery into practice – Establish a discoverable Data Inventory

Data cannot be effectively managed by agencies unless they know what they have and where it is held. A data asset stocktake or inventory provides a way of maintaining a 'single source of truth' regarding the location and status of data assets.

An inventory, using a recognised standard, with rich metadata to support it, establishes a baseline from which data can be managed. Data stewards or data managers can access relevant information about a data asset, such as when it was collected or acquired, its licensing

arrangements, data formats, methodology (if data lineage or provenance is captured) and potentially, when the data asset should be updated, archived or disposed. It can also improve understanding of the key stakeholders and users involved with a data asset.

A well-managed data inventory supports broader data governance processes including:

- Data access control. A good inventory should display the record for the data asset, but not necessarily provide access to it. The record should outline who can access the data and under what conditions.
- Assigning and managing responsibility for, or ownership of, the data assets so there is clear accountability (e.g. data stewards, business owners, data custodians, etc.).
- Notifications relevant to information management, such as when data assets need reviewing for archiving or disposal purposes. This enables agencies to comply with the *Archives Act 1983*.
- Maintaining a central list of personal information holdings as required under the Australian Government Agencies Privacy Code (10.5.(b)).
- Tracing data lineage or provenance. This is the ability to record movements, changes and interactions with the data throughout its lifecycle (i.e. as it is acquired, processed, analysed, shared and released).
- Decisions relating to data architecture requirements (e.g. the creation of a searchable data catalogue in future).

What makes a data inventory effective?

A data inventory is most useful when it is more than simply a list of data assets. Knowing what assets an agency holds is important, but there are a number of other factors that will make it a more effective management and discovery tool.

Metadata

Metadata describes a set of information (e.g. date created, date updated, source of data, security classification, etc.) about a dataset.

The metadata attributes, which would form the basis of the Data Inventory, should be the metadata most useful for agency data management and/or for users to identify and request access to data relevant to their work. Reliable access and licensing information can indicate what data can be shared or accessed and by whom, removing the need for staff to make 'judgement calls' and minimising the risk of inappropriate sharing, access or use.

The ability to search the list of data holdings and discover data is dependent on the quality and consistency of the metadata. If asset records contain complete and accurate keywords and dates, clear descriptions, accurate titles and are human-readable, searching will be far more effective.

This means a data inventory is only as useful as the metadata that sits behind it. Ideally, metadata should be based on recognised standards. There are two standards set by the [National Archives of Australian: the Minimum Metadata Standard](#)⁸ and the Australian Government Recordkeeping Metadata Standard⁹. There are several international standards which can also be adopted: the Dublin Core Metadata Initiative¹⁰, DCAT¹¹ or ISO 19115-1:2014¹². Using established metadata standards will enhance the comparability of data inventories or catalogues, within and across agencies.

Examples of metadata attributes for possible inclusion in a data inventory:

- Data asset name
- Description (e.g. purpose, collection method, etc.)
- Topic / key words / themes
- Authority for collecting the data, e.g. the legislation or policy driving its collection
- Business owner / Data steward / Data custodian
- Usage caveats (e.g. software requirements, coverage limitations, etc.)
- Time period covered
- Geographic coverage
- Data dictionary (i.e. definitions of key terms)
- Date created / modified
- Frequency / time series
- Protective marking

Governance

A useful data inventory will be well-governed and support other data governance and management processes. Well-governed means that it will have clear guidelines on what level of detail needs to accompany the attributes describing the data asset. This may include:

- Minimum attributes or 'fields' that must be completed. There should not be too many to be burdensome, or too few that the asset is not adequately described.
- A data dictionary or glossary describing what each attribute means. This will ensure consistency of interpretation of content.

8 <https://www.naa.gov.au/information-management/information-management-policies/digital-continuity-2020-policy>

9 <https://www.naa.gov.au/information-management/information-management-standards>

10 <http://dublincore.org/documents/dcmi-terms/>

11 <http://www.w3.org/TR/vocab-dcat/#describing-catalog-records-metadata>

12 <https://www.iso.org/standard/53798.html>

- Controlled vocabularies or ‘pick-lists’ containing agreed terms describing an attribute. This minimises free text and burden for completing an entry and assists with consistency.
- Guidelines regarding what content is expected in each attribute where free-text is an option. This is essential for describing and understanding the data asset.

A well-governed data inventory will also have a person or team that supports and maintains it on an ongoing basis. Records in a Data Inventory will need to be kept up to date for it to remain useful to the agency. Data stewards should be responsible for updating the inventory as needed, but having a coordinator can ensure this occurs when the task is ‘forgotten’ by stewards. The Data Inventory should also be made available to all staff for ease of maintenance and to realise the benefits of data discoverability.

Coverage

Ideally, a data inventory is built around a comprehensive ‘stocktake’ of an agency’s data holdings, which could include financial data, corporate data and client data in addition to what is traditionally considered to be the operational data unique to the organisation. Whilst comprehensive coverage is desirable, it may be easier to prioritise efforts on operational data first, and then move onto corporate data. This is a decision for agencies to make based on their resources and context.

For agencies with multiple data inventories it may be necessary to consolidate these if they are legacy systems or contain duplicate records. Where multiple inventories are necessary, ensure the purpose and content of each one is clear. In such cases, agencies may wish to have some form of meta-catalogue, or catalogue of catalogues, to enable easy searching.

Integration with existing processes

Recording data in the inventory should not be a discrete activity; it must integrate with other data management processes. It should also align with an agency’s information and record management policies and strategies, as a data inventory is another form of record and information management.

Undertaking a data inventory and establishing standard cataloguing processes is an area where the Senior Data Leader and a CIO can lead together. The CIO and Senior Data Leader can work to streamline all data/record/information management processes to ease the burden on staff and boost information management capability.

As part of the Australian Government Agencies Privacy Code, agencies must keep an inventory (formally described as a register) of personal information holdings. Given much of the personal information agencies hold are in datasets, the data inventory will satisfy this requirement. Alternatively, if no broader data inventory already exists, it may provide a starting point for a data inventory. Teams working on a Data Inventory should investigate, leverage and potentially adapt what already exists in an agency.

Next steps: from inventory to catalogue

As agencies continue to develop their data maturity and capability, they may consider transitioning this inventory into a data catalogue. This opens up further possibilities and benefits, including automating a range of data management and governance processes, improving the discoverability of data, creating externally facing as well as internally facing catalogues and possibly aligning or integrating catalogues across agencies.

There are many off-the-shelf and open source cataloguing options available. It may be hard to know which one will suit your agency as there is no one-size-fits-all option. There are benefits to purchasing a vendor-supported product, but also benefits to using more flexible and cheaper open source software and even benefits to developing an agency-tailored catalogue in-house. These decisions are heavily dependent on an agency's existing data maturity and IT infrastructure.

When selecting catalogue software specifically, some important considerations for agencies are that the catalogue:

- meets the business need
- can integrate with other business systems and/or data infrastructure
- can integrate with other Enterprise Architecture or ICT registers, such as business registers (e.g. capturing business process and legislation, which can then link to datasets associated with that business process/legislation), application registers (e.g. so that datasets can be linked to the applications that house the datasets) and infrastructure registers (e.g. so that the servers that host applications can be linked).
- can appropriately extract and/or maintain metadata for individual data assets, and
- is intuitive to use: this should be judged not only by the technical support, but all users of the catalogue

It may be helpful for agencies to liaise with other agencies that have developed or purchased a data catalogue to understand the pros and cons of different systems and approaches.

Where can I get more information?

Below are a few articles that can be used as a starting point:

[Choosing a Data catalog \(Ekerson Group\)](#)

[Open Government Data Toolkit: Technology Options \(World Bank\)](#)

[Data Discovery and Catalogues \(Bloor Research\)](#)

Questions to ask:

- **What type of data 'stocktake' is most appropriate for our agency? An inventory, or catalogue?**
- **Who needs to be convinced of the benefits?**
- **What needs to be included in the data inventory or catalogue for it to be useful?**
- **Who will be involved in creating the data inventory or catalogue?**
- **How will the data inventory or catalogue be maintained?**



Next Steps

The *Foundational Four* provides a strong basis to build data culture and capability. It establishes a leader to advocate for data, identifies a strategy to set an agency's data directions, sets up programs to drive governance and better use of data, and identifies the data they have custodianship of.

As with any capability, on-going attention is needed to ensure programs and strategy remain functional and useful; they are not 'set-and-forget' activities. The Data Governance Program can be refined over time. A Data Inventory will need updating for new assets, or improving the metadata for existing assets. A second-phase Data Strategy can be established once the first has been completed. A Senior Data Leader will continue to grow into their role and enhance the status of data within an agency and can also provide leadership more broadly through the APS.

As the *Foundational Four* become operational, the next steps an agency takes will depend on the culture, context, priorities and business/policy drivers of that agency. What will be important for one may be less relevant for another agency. Some key questions to ask if you are thinking about next steps are:

- Is your agency driven by policy, service delivery or research outputs? What would help deliver these functions in a more joined up way?
- What outcomes is your agency trying to achieve and what role does data have in supporting those outcomes?
- How can you use data to help build trust in the services and policy your agency delivers?

While each agency's next steps may be a little different, there are some core areas which every agency may consider. A **framework that describes how to value your data assets** can help define how you trade-off the effort of managing data against the benefit derived from the use and reuse of data. It can also help identify business critical data which can then help inform security and business continuity requirements.

Agencies could invest in further **data awareness** and **education** to ensure that all staff see the value of data in their agency and know how to best use it in their daily work. Part of understanding the value of data also requires consideration of the value of your agency's data to others. Agencies can work towards **increasing transparency** of data holdings through **releasing more data as open** where appropriate or publishing details of the data assets so that other parties can seek access to data holdings. To do either of these steps, agencies will need to **adopt metadata standards**, remediate existing metadata and integrate metadata management into data management practices.

Agencies with a policy focus may need to increase capabilities to use data to **better answer complex policy questions** through advanced analytics or data integration. This would require **assessing and improving the quality** of data assets and investing in staff **data and analytics capabilities**. Agency activities will be complemented by the Australian Public Service Commission's work to establish a data profession stream in 2020, an action which emerged from the APS review.

Agencies with a service delivery focus may need to increase capabilities to use **data as an input into business intelligence** to improve service delivery. This would require standardising **data collection methods, improving data quality** at the source of information capture and improving **interoperability of data** to ensure data can be exchanged easily and securely between service delivery agencies.

Agencies should reach out to their peers to gain insights into what others are doing, what current initiatives may be leveraged, or where agencies could collaborate to make cross-agency improvements. Working together in a whole of APS way, will continue to drive a shift in data culture, modernise the public service and enable the benefits of improving data practices to be realised.



