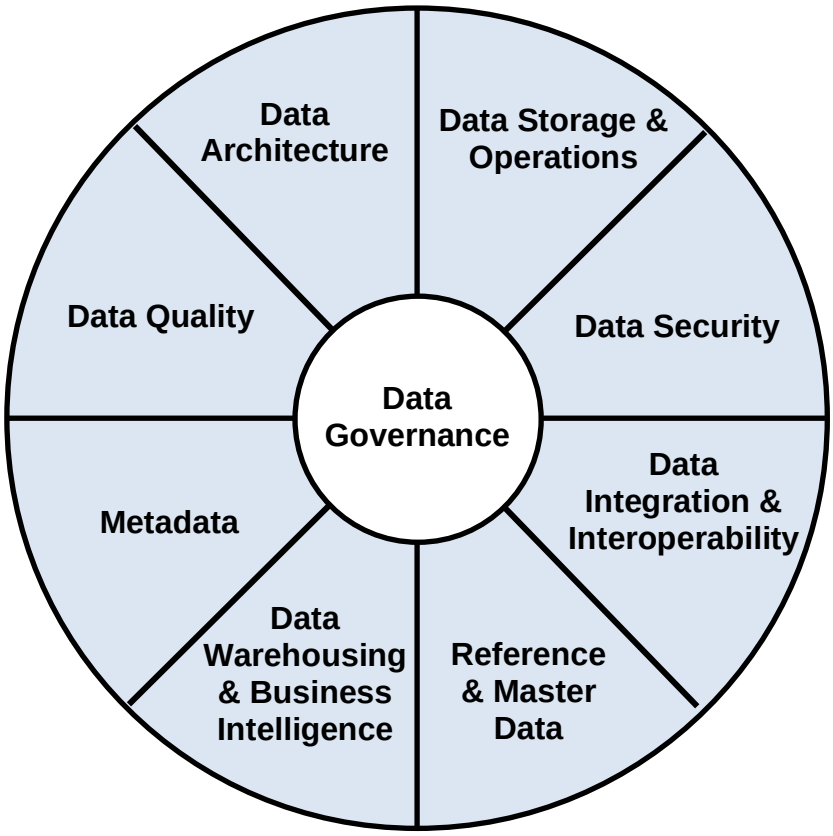
12. Data Management

Where data governance sets the rules of engagement for how data-related decisions are made within an organisation through the creation of policies and processes, data management refers to the planning, execution and operation of these policies and processes.

As illustrated in the following figure, there are 8 core functions of data management which contribute to the effective governance of data:



Source: Adapted from DMBOK Data Management Framework

These functions are adapted from the Data Management Association (DAMA) Data Management Body of Knowledge (DMBOK). While each data management component is important, not all the functions must be included in the first phase of a governance program. For example, some programs will focus more on business definitions (Metadata) initially, while others may emphasise a single view of the customer (Master Data).

12.1 Data Quality Management

What is it?

An enterprise-wide process to manage and improve the quality of an agency's data is a key component of effective data management. Data quality management is a continuous process which involves managing data across the full data lifecycle, from its initial creation to its destruction. It involves the implementation of data quality standards and procedures to address and improve the accuracy, completeness, timeliness, relevance, consistency and reliability of the data.

Why is it important?

The outcome of decisions depends on the quality of the information used to make that decision. Poor data quality can result in poor decisions and unintended outcomes. High data quality can support an agency to achieve desired outcomes by ensuring decision-making is based on accurate and timely information. Data quality methods and procedures are essential to ensuring accurate data is available to decision-makers in a timely manner.

What good looks like:

- **Automated**: the quality of data is managed through automated tools that can automatically detect data quality issues and cleanse and enrich the data.
- **Lifecycle management**: the quality of data is proactively managed across the data lifecycle, from collection through to disposal.
- **Root-cause remediation**: problems with data quality are addressed at their root cause (e.g. fixing the problem at the source).
- **Enterprise-wide**: data quality management is regarded as a responsibility of all staff that have a role in handling data.
- **Standards-driven**: requirements are defined in the form of measurable standards and expectations against which the quality of data can be measured.
- **Monitored**: data quality requirements are enforced through clear monitoring, reporting and issues management processes.

How to achieve good practice:

- **Define data quality requirements for your agency that are relevant to your business needs**. This will ensure consistency across the organisation and help you determine which data to keep, which to get rid of, and which to correct.
- **Measure data quality levels** – a data quality assessment tells you how effectively data is meeting your business needs and stakeholders' requirements.
- **Create and implement a strategy for improving data quality** – this strategy should include:
    - Industry standards for available data
    - Organisational data standards
    - Timeliness for data availability
    - Data quality metrics
    - Goals for data quality metrics
    - Data quality rules for specific fields
- **Develop operational procedures and automated processes to improve data quality** – for example, the NSW Government Data Quality Reporting Tool can be used to understand the different dimensions of data quality and generate data quality statements. Data quality assessments should always be recorded in the metadata. Tools that automate data profiling are also available and can help your agency enrich large amounts of data.
- **Train staff with data responsibilities on data quality rules** – staff are responsible for ensuring that business rules or issues related to data quality are documented, developed and managed in a consistent way, in accordance with the agency's data requirements.
- **Monitor and report on quality levels of your data** – this supports the active management of data quality across the agency and enables the prioritisation of data quality improvement initiatives.

Relevant Standards:

- NSW Government Standard for Data Quality Reporting – the purpose of this document is to establish common principles and protocols for reporting on data quality, so that agencies can create simply data quality statements and users can easily evaluate whether shared or published data is suitable for re-use.

Useful resources:

- Data Quality Reporting tool – this tool is designed to support the NSW Government Standard for Data Quality Reporting. It guides you through a reporting questionnaire to generate a Data Quality Statement. All data should be accompanied by a data quality statement as it helps a user understand how the data can be used.
- ABS Data Quality Framework – NSW has adopted the Australian Bureau of Statistics (ABS) Data Quality Framework to describe the dimensions (or characteristics) of data quality. The framework can assist you with the development of statistical collections to produce high quality outputs.
- ISO 8000 Data Quality – this is the global standard for Data Quality and Enterprise Master Data. It describes fundamental concepts of information and data quality and how these concepts apply to quality management processes and quality management systems.

12.2 Metadata Management

What is it?

Metadata management means maintaining information about data to ensure both the users and systems:

- understand the meaning of data,
- know why data was created and for what purpose; and
- can find data easily when they need to.

By having high quality information that describes the information in data, as well as its storage and origin, staff can understand what the information is, what they can learn from it and how to find it quickly. Depending on the data, metadata may include the lineage, ownership, format, and any rules to be applied to the data.

Metadata management requires a consistent way to capture, manage and publish metadata information. This means controlling the creation of metadata by setting clear standards (and, where appropriate, adhering to well-established industry metadata standards), as well as implementing policies and procedures for metadata management and ensuring they are enforced across the organisation.

Why is it important?

Without metadata, it is very difficult for potential data users to know whether a dataset is available, where it is stored, what the data means, and how accurate it is. A key reason for duplicated data collection and re-work across government is the fact that repositories of what data has been collected are either inadequately maintained or do not exist. Therefore, implementing robust metadata management practices are required to ensure that data can be located, understood and used not only across the agency, but also by other agencies and non-government users.

What good looks like:

- **Valued**: the value of having managed metadata, and its role in improving data quality, is recognised across the organisation.
- **Standardised**: metadata conforms to relevant industry standards to enable data sharing and re-use across the public and private sector.
- **Accessible**: metadata is recorded and maintained on an accessible repository and is freely available at no additional cost with the provision of the dataset.

- **Assured**: the quality of metadata is assured, measured, monitored and improved.
- **Agreed**: changes to metadata are agreed and authorised with due consideration of impacts to other data management functions and business processes.

How to achieve good practice:

- **Identify the metadata associated with priority common data elements**. Determine the level of consistency of metadata for priority common data elements.
- **Measure current metadata effectiveness** and determine the level of consistency needed for efficient agency operations.
- **Define a minimum metadata standard for your agency** in consultation with agency stakeholders and ensure compliance with well-established industry standards.
- **Establish or improve metadata policies, rules, practices and roles** – this can be done by implementing a metadata adoption plan and implementation process across the organisation.
- **Educate staff on the value of metadata**, **as well as on access and use of metadata** – this may include education of staff on their respective metadata management responsibilities.
- **Create a single metadata repository where agency stakeholders can find information** – this can be done by bringing individual repositories together to develop a central (or federated) electronic database that is used to store and manage metadata.
- **Create feedback mechanisms** – to ensure that data users can provide input on the effectiveness of metadata and incorrect or out-of-date metadata.

Relevant standards:

- Metadata Online Registry (METeOR) – Australia's repository for national metadata standards for health, housing and community services statistics and information.
- ANZLIC Metadata Profile Guidelines – ANZLIC – these guidelines provide practical information to better understand and implement the ANZLIC Metadata Profile. The ANZLIC Metadata Profile defines the appropriate content of metadata for geographic information or spatial resources.

- AS ISO 23081.1:2018 – covers the principles that underpin and govern records management metadata.
- AS/NZS ISO 19115:2005 – provides a standardised metadata format for describing geographic information and services.
- AS/NZS ISO 15836:2016 – establishes a standard for cross-domain description and defines the elements typically used in the context on an application profile.

Useful resources:

- Metadata for records and information – NSW State Archives and Records provides a range of advice on metadata, including:
  - The minimum requirements for metadata for authoritative records and information
  - Principles for implementing metadata for records and information
  - What metadata for records and information can achieve
- National Archives of Australia Metadata for Interoperability Guide – this guide provides information on how to develop an organisational Metadata strategy, information on metadata harvesting tools and protocols, tips for building a metadata repository and links to relevant resources and standards.

12.3 Data Security and Privacy

What is it?

Data security and privacy management includes the policies, processes and procedures that are in place to ensure that data is kept safe and secure across all stages of the data lifecycle. Data security and privacy measures are implemented to protect agencies' critical, personal or otherwise sensitive data from unauthorised access and use, and ensure that data can move securely through the organisation. Adherence to privacy legislation, as well as customer and community privacy concerns, is paramount when considering data security and privacy management.

Why is it important?

Data can often contain personal, confidential or otherwise sensitive information that can have serious implications for both the populations the data is about and the organisations storing it. Good data governance practices across your organisation will ensure it is protected against misuse, interference, loss, or unauthorised access, modification or release. Serious physical, emotional or reputational harm to individuals may occur if data becomes compromised. Data breaches can also result in reputational damage and loss of public trust, as well as financial and legal ramifications.

What good looks like:

- **Compliance**: data is managed in accordance with relevant privacy legislation and NSW government and agency-specific security policies, procedures and standards.
- **Clear roles**: roles and responsibilities for authorising and overseeing safeguarding processes are clearly defined and access rights are assigned on a need-to-know basis.
- **Classified**: the safe handling requirements of data are known because each data asset is classified according to the NSW Government Information Classification, Labelling and Handling Guidelines.
- **Privacy-by-design**: privacy measures are built into the design and architecture of information systems, business processes and network infrastructure.
- **Minimised**: data creation and collection processes are designed to ensure that minimum personal information is collected.

- **Transparent**: agencies are transparent and accountable about the procedures used to protect personal data, including the choices made in balancing competing interests.

How to achieve good practice:

- **Define and communicate policies on privacy and security with staff** – ensure alignment with relevant legislation, policies and frameworks. For example, all staff must comply with the *Privacy and Personal Information Protection Act 1998 (NSW)*.
- **Assess current data security risk and define controls to manage risk** – risk analysis should include examination of unauthorised access; human factors such as accidental and intentional errors, omissions, destruction, misuse and disclosure.
- **Implement data privacy and security controls** – including privacy impact assessments (PIAs), privacy breach response procedures, clear arrangements for handling privacy complaints, multi-factor authentication, encryption, logging and monitoring procedures.
- **Training for staff on privacy, confidentiality and data security** – including education on existing industry-based standards for data handling, data minimisation and de-identification, the right for individuals to access and correct their personal information, as well as their role in ensuring data is collected and used only for the intended purpose(s).
- **Monitor, review and revisit data security measures** – continuous monitoring activities include control of IT system components, ongoing assessment of security controls, and auditing user access.

Useful resources:

- NSW Cyber Security Policy – outlines the mandatory requirements for sensitive and classified information.
- Information Classification, Labelling and Handling Guidelines – the Guidelines support the implementation of the NSW Cyber Security Policy. They provide guidance for the application of security classification to prevent government information assets from potential security breaches. This includes how to classify information and the protocols for handling and transmission of information.

- Making data safe for sharing guidance – provides guidance for NSW public sector agencies on how to make data safe (e.g. through data minimisation and de-identification) for sharing and public release.
- Five Safes Framework – provides guidance on how to develop safe data projects and manage disclosure risks according to five 'safe' components and allows data custodians to place appropriate controls on not just the data itself, but the manner in which data can be accessed.
- The IPC Public Interest Test – the Public Interest Test is the practical application of the *Government Information (Public Access) Act 2009* (GIPA Act) and it is designed to help you decide whether or not your data can and should be made open.
- IPC Information Governance Agency Self-assessment tool – enables agencies to conduct an assessment of their systems and policies that ensure their compliance with privacy and information and access requirements.
- Information and Privacy Commission NSW website – provides guidance on implementing privacy obligations under the PPIP Act and the IPPs and/or the HRIP Act and HPPs.

Relevant standards:

- ISO 27001 – sets of the international requirements for an Information Security Management System (ISMS). An ISMS is a systematic approach to managing sensitive company information so that it remains secure. It includes people, processes and IT systems by applying a risk management process.

12.4 Data Warehousing and Business Intelligence

What is it?

A Data Warehouse is a consolidation of data from a variety of sources that is designed to support and optimise organisational decision-making. Its main purpose is to consolidate data and provide readily available and easily accessible information to the business's decision-makers. Business Intelligence (BI) refers to a set of methods and techniques that are used by organisations to enable strategic decision-making. BI leverages technologies and tools that focus on reliable measurement of facts and business objectives to provide better insights to support evidence-based decisions.

Why is it important?

Fragmented, inconsistent and outdated data in multiple databases does not enable informed and strategic decision-making. Data warehousing and BI give business units a way to consolidate and process vast amounts of information and perform more advanced analytics. With appropriate data warehousing in place, systems have the right data available to perform more accurate analysis and get more value from BI and analytics programs. An agency that acts on knowledge gained from BI and analytics can improve operational efficiency and find better ways to innovate based on insights from data.

What good looks like:

- **Current**: the data warehouse supports appropriate information access and is designed to deliver up-to-date information to decision-makers.
- **Business goals:** the data warehouse serves agency strategic priorities and informs the selection of BI solutions.
- **Start with the end in mind**: the business priority drives the creation of the data warehouse content.
- **Once size does not fit all**: use the right warehousing and analytics tools and products for your specific purpose.

How to achieve good practice:

- **Understand your business needs** – start by consulting with stakeholders across the organisation to identify why your agency needs the warehousing and BI solution(s) and what objectives you are seeking to achieve.
- **Avoid one-stop-shop solutions** – while warehousing and BI solutions with all the features may seem enticing, focus on implementing solutions that meet your core business needs.
- **Don't overcomplicate the solution** – while there are endless data warehousing and BI tools available, it is important to focus on getting the basics right before you add additional requirements.
- **Use BI tools for their intended purpose** – rather than trying to find a tool that does it all, look for BI solutions that do something (or a few things) well. It's also important to find tools with user-friendly interfaces that meet the users' needs.
- **Consider how to make the transition easy for staff** – this can be done by ensuring there is a communication strategy for the transition, and that adequate training and guidance is available for staff who are expected to use the solution.

12.5 Reference and Master Data

What is it?

Reference and master data are data that provide a consistent, reliable record for all critical business data across the organisation. Master data can be defined as the “golden record” of critical information that the organisation relies on (e.g. customers, employees, locations, and products). Reference data is a type of master data that is more likely to change and that it less critical to the business. Agencies need to define and manage how master and reference data will be created, integrated, maintained, and used throughout the organisation. The challenges of this are determining the most accurate data values from among potentially conflicting data values and attempting to make that data available wherever needed.

Why is it important?

Definition and management of data assets used across an agency is necessary to meet strategic objectives, reduce risks associated with data redundancy, and reduce the cost of data integration. The management of master and reference data allows agencies to correct data inconsistencies across business units and systems and apply uniform business rules to enable sharing of data assets across agencies and government more broadly.

What good looks like:

- **Interoperable**: Master data and reference data is managed so that it is interoperable across business units and government agencies.
- **Standardised**: Master and reference data should be modeled according to agreed state, national and international standards so the data is represented appropriately.
- **Single view**: Master data is recorded and maintained on an accessible and, where possible, centralised repository to create a single view of the data.
- **Controlled**: changes to reference and master data are agreed and authorised with consideration of impacts to other business processes.

How to achieve good practice:

- **Identify and agree on data definitions** – this involves determining the most accurate data values from among potentially conflicting data values and getting agreement from different parts of the organisation.
- **Collect the master data into a central database** – this database should link to all participating applications.
- **Publish reference and master data** – ensure its use in all appropriate business intelligence and analytics reporting across the organisation, at all levels.
- **Establish maintenance policies and processes**

Relevant standards:

- ISO 8000-115 Data Quality – Part 115: Master Data – this is the global standard for Data Quality and Enterprise Master Data. It describes the features and defines the requirements for standard exchange of Master Data among stakeholders.

12.6 Data Storage and Operations

What is it?

Agencies must ensure data storage environments are secure, comply with relevant legislation, and enable information continuity, sharing and re-use. A number of laws and policies affect how NSW Government agencies can store their data. For example, NSW Government agencies must comply with the *State Records Act 1998* (NSW), which requires agencies to ensure appropriate records storage, maintenance, security and archiving. It is important to note that outsourcing storage does not lessen an agency's obligation to ensure information is stored appropriately.

Why is it important?

Due to its rapidly increasing volume, how and where agencies store their data is becoming increasingly important. Storage environments must be able to manage large volumes of complex data and to provide consistent levels of security, accessibility and functionality. To ensure the long-term continuity and accessibility of data assets, agencies need to find appropriate and secure storage environments that comply with legislative and regulatory requirements.

What good looks like:

- **Digital continuity**: storage environments enable information continuity by ensuring the preservation and maintenance of key data assets.
- **Retention and disposal:** storage environments ensure data is kept and disposed of in accordance with business requirements, protective security requirements, and legislative requirements under the State Records Act, PPIPA and HRIPA.
- **Best practice:** database standards and best practices are understood and applied.
- **Re-use:** storage environments that promote data re-use and integration are preferenced.
- **Migration, transition and decommissioning**: changes to storage environments are agreed and authorised to ensure that data of long-term value is transitioned to new environments or appropriately assessed in decommissioning arrangements.

How to achieve good practice:

- **Identify your agency's storage needs**. All agencies' storage needs will be different so it's important to identify and agree on these requirements.
- **Ensure alignment of business needs to the storage infrastructure**.
- **Ensure storage infrastructure complies with data retention periods** specified under the State Records Act and PPIPA.
- **Ensure the storage infrastructure selected is efficient and flexible**. This means that it is easy to search, query, and store the data.
- **Create and maintain a Data Asset Register/Inventory** that identifies high-value and high-risk data assets and their storage locations.
- **Manage and monitor effectiveness** of the storage infrastructure.
- **Ensure future planning for business continuity** by considering if the storage infrastructure is laying a foundation for future data initiatives and if it can be scaled as needed.

Useful resources:

- State Archives General Retention and Disposal Authorities – outlines the retention and disposal requirements for different types of information, as well as the requirements for storing records outside of NSW.
- NSW Government Cloud Guidance and Policy – provides practical steps to move services to a cloud. This includes information on preparation, contracting and managing, as well as considerations to note when moving to cloud.
- NSW Cyber Security Policy – agencies must abide by the Policy when procuring cloud services. The Policy outlines mandatory requirements to appropriately manage cyber security risks, including the requirement to identify agency 'crown jewels'.
- NSW Internet of Things Policy – provides practical guidance to help agencies design, plan and implement IoT solutions, including guidance on storage options for data generated by your IoT initiative.
- Australian Cyber Security Centre's Cloud Computing Security Considerations – provide detailed cloud security considerations, which include: maintaining availability and business functionality; protecting data from unauthorised access by a third party, the vendor's customers and by rogue employees.

- Government Data Centres Guidance – provides information on the benefits of government data centres and services, including secure data storage and access to services in the cloud.
- National Archives of Australia Outsourcing Digital Storage Guidance – provides advice on outsourcing digital storage, including key risks and consequences of offsite storage location. In addition, the Records Management Risk Assessment Template and the Checklist for Cloud computing and information management provide a helpful understanding of the potential risks and considerations associated with outsourcing storage of your agency's data.

12.7 Data Integration and Interoperability

What is it?

Data integration and interoperability is the ability of systems, organisations and people to exchange data between each other so that they can work together seamlessly, either in the present or in the future. Integration is the ability to consolidate data into consistent forms, either physical or virtual, and interoperability is the ability for multiple systems to communicate. Both are dependent on clear, shared expectations for the context and meaning of data across systems.

Why is it important?

Data integration and interoperability support the use and reuse of government data by allowing agencies to get data where it is needed, when it is needed, and in the form in which it is needed. Having integrated and interoperable data can assist agencies to make better decisions and to provide consistent, coordinated and more timely services by ensuring they have access to the right data at the right time. Lack of interoperability between systems means that government agencies often cannot share information effectively, which contributes to disjointed services, operational inefficiencies and poor citizen outcomes.

What good looks like:

- **Enterprise-wide**: data is stored in agency-wide enterprise architecture, where appropriate.
- **Standardised**: software and hardware conform to defined standards that promote interoperability for data, applications and technology.
- **Understood**: data users understand the meaning of exchanged information through the consistent use of metadata, master data and data quality standards.
- **User-friendly:** interfaces are flexible and generic enough to suit multiple uses.
- **Minimise replication:** data is linked rather than copied.
- **Modularity:** modularity of system design is maintained.

How to achieve good practice:

- Assess current state of interoperability to establish a strong understanding of your agency's business and data management environment
- Build future state vision that defines the requirements for creating new services and systems. Ensure requirements are defined across business functions to ensure the architecture supports the overall business strategy
- Undertake a gap analysis and quantify gaps between current and future state
- Planning and design of solutions to bridge gaps. Avoid boiling the ocean and focus on bridging gaps that are important for your business. Think quick-wins as well as long-term planning
- Implement frameworks, policies and standards and tools to support integration
- Monitor new processes for ongoing improvements

Useful resources:

- The National Archives of Australia has developed the following resources:
  - Interoperability key themes help you understand how interoperability is not just a technical fix, as it also relies on working with your information and data to align your business, security, legal and semantic needs.
  - Interoperability development phases will help you plan and implement solutions to address interoperability hurdles that are visualised in the interoperability scenarios.
  - Your results from using the Business System Assessment Framework (BSAF) can be used to identify:
    - the need to *integrate* business systems or to *migrate/export* data to address risks or gaps
    - system functionality to meet your information and data needs over time
    - what information and data is held in your systems and its value.
- The NSW Government IoT Policy also contains several sections related to interoperability.

12.8 Data Architecture

What is it?

Data Architecture defines information flows in an organisation, and how they are controlled. It relates to incoming data and determines how it is captured, stored and integrated into other platforms across the organisation. It involves understanding business objectives and the existing data infrastructure and assets, defining data architecture requirements, and shaping the enterprise data architecture to provide greater benefits to the organisation. The primary focus of data architecture is to integrate the existing applications and make them interoperable so data can be used across the organisation.

Why is it important?

Like many large organisations which have been in existence for a long period of time, government agencies have many legacy systems which use older technology or bespoke solutions to hold their data. These systems are often difficult to map out and connect with and require tremendous effort to support change.

What good looks like:

- **Aligned**: data architecture is aligned with the organisation's business strategy.
- **Comprehensive**: the data architecture minimises impact of information silos by combining data across the agency's business functions.
- **Integrated**: the data architecture provides a mechanism that documents the relationship among architecture components across domains and their alignment to agency and whole-of-government strategic goals.
- **Scalable**: the architecture can be applied to various organisational levels and scopes (i.e. whole-of-government, cross-agencies, agency, line of business, segments, capability, etc).
- **Flexible**: the architecture supports automation and is designed to meet changing business needs and new technology.
- **Standards**: the architecture adopts best-practice architectural design (such as Reference Architectures) to build and document common business and technical capabilities.

How to achieve good practice:

- **Assess current state architecture** of the organisation.
- **Define future state architecture** of the organisation, within the context of the strategic goals of an agency and its operating model.
- **Perform a gap analysis** between the current state and the future state.
- **Develop a roadmap or implementation plan** that contains a necessary set of actions to transform the organisation from the current state architecture to its target state.
- **Regularly report** on the effectiveness of the roadmap and implementation to the Data Governance Board or Committee.
- **Recruitment and retention of expertise in data architecture**, to guide agencies as they move away from legacy systems and siloed data towards integrated and consolidated data platforms.

Useful Resources:

- NSW Internet of Things Policy – provides practical guidance to help agencies design, plan and implement IoT solutions, including guidance on how to manage ICT infrastructure change when integrating or migrating new systems with legacy systems (see section 6.1.4)

Relevant standards:

- ISO/IEC 42010:2007 Systems and Software Engineering