WATER SERVICES
ASSOCIATION OF AUSTRALIA

WATER SECTOR DATA PLAYBOOK

JANUARY 2023

WATER SERVICES
ASSOCIATION OF AUSTRALIA

Contents

Introduction

The intent and purpose of the playbook is to provide options and approaches to enhance the management of data by all water businesses. The intent is to provide an overarching approach to better data management, coupled with a menu of options to select to enhance data management and the way data is used within the organisation.

Establishing the enterprise data strategy

Data should be viewed as an asset which needs to be created, managed and appropriately disposed. Doing so requires an enterprise perspective. This is because data is a dynamic yet intangible asset that is used across all of the verticals in an organisation. Good data management enhances the ability of the organisation to optimize processes, respond to customers, meet changes in the external operating environment and deliver on stakeholder expectations. Poor data management can result in significant business costs through poor decisions and legal exposure which may arise from data loss, poor data integrity and inappropriate use.

Management of data requires an understanding of:

1) What problems is the organisation trying to solve and what does it need analytics for.
2) What data sources does it need to use or acquire.
3) Timeliness of the data and scope of the data to provision.
4) The impact on and relation to other data structures.
5) Influences on existing modelled data.

Effective data management requires robust data governance. Key data governance considerations are summarized in the Data Management Body of Knowledge (DMBOK) published by DAMA, on which much of this guide is based. The first part of good data management is establishing an organisational Data Management Strategy. The data management strategy should start with good data governance at the center, then cover the key areas outlined in Figure 1. The elements of each section should be included in the overall data strategy, with roles, accountabilities and expected actions documented.

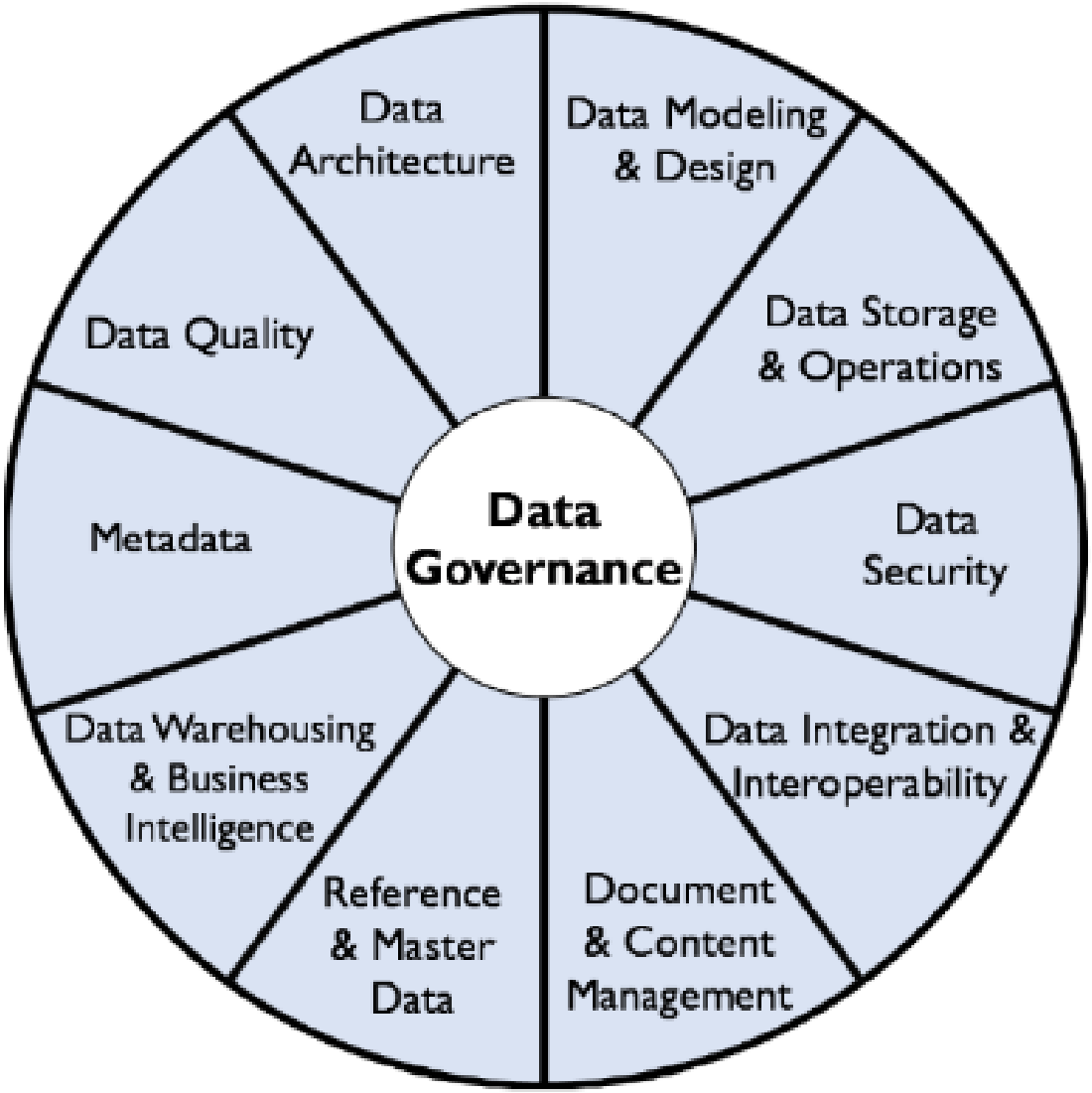

Figure 1: DMBOK data management framework.

The strategy should be designed to manage all aspects of data covered by data governance (Figure 1) across the entire lifecycle (Figure 2). It should be designed to manage data across the data lifecycle with the intent of focusing on the organisation's most critical data and minimising ROT (Redundant, Obsolete and Trivial) data.

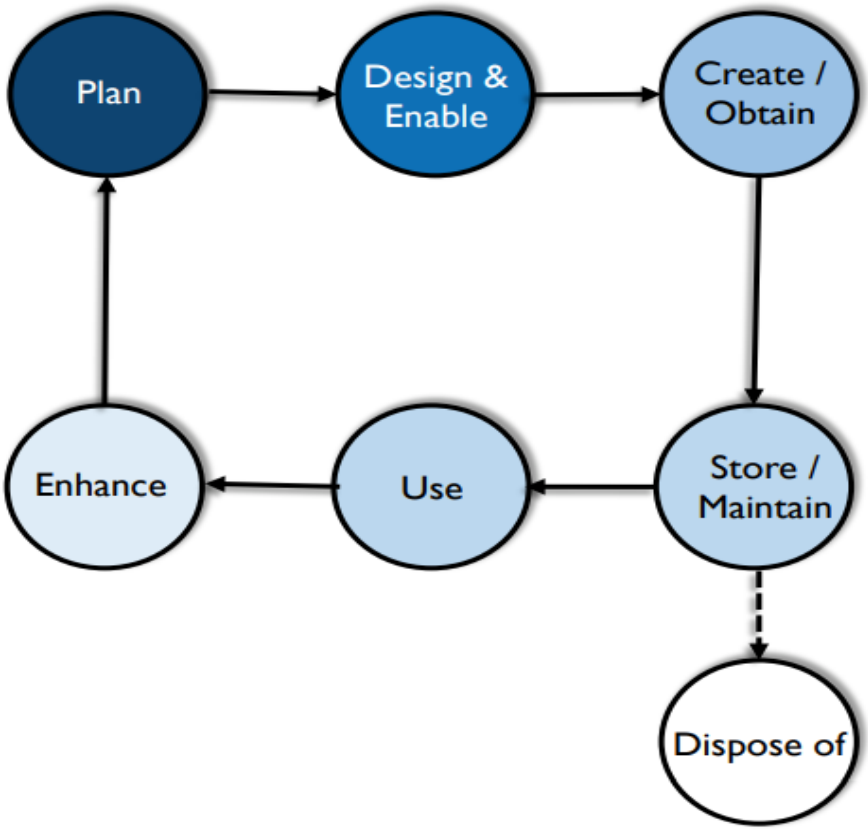

Figure 2: The Data Lifecycle (DMBOK2, P.29)

At a high level, the steps to implementing a data management program are:

- Establish governance
- Understand the current state of data within the organisation.
- Understand opportunities in order to develop a roadmap for data management
- Initiate and organisational change management program to support execution of the roadmap.
    - Set priorities
    - Develop and maintain relevant standards and policies across the data lifecycle
    - Establish communications and knowledge sharing mechanisms
    - Monitoring and reporting on performance
- Share results to build awareness and identify opportunities for improvement

Governance

The keys to ensuring successful data governance are:

- An ongoing leadership commitment.
- Embedding good data management processes into all aspects of the business.
- Measurement of outcomes and implementation effectiveness.

It is important to recognise that data management requires organisational commitment and leadership. This begins with an Executive Champion to enable the adoption of an enterprise approach. The most effective way to implement good data management practices is to nominate a Chief Data Officer or their equivalent. Their role is to coordinate the data governance processes and interact with all departments within an organisation. Along with the Chief Data Officer there needs to clear accountabilities and an operating framework that helps define the interactions within the business. The approach also needs to be principles based.

The governance model, along with the roles and responsibilities depend on the nature of the business and how it operates. Examples of different governance models is provided in the DMBOK2 guideline.

Embedding data governance requires alignment with the business direction and ensuring shared responsibility between the data stewards and technical experts at all levels of the business.

The measurement of data outcomes is a vital and often overlooked aspect of data governance. Details of data outcome measurement are provided later in this guide. Measurement allows assessment of the effectiveness of the data management approaches, facilitating ongoing improvement. It is important to remember that data management is a constantly changing process which requires ongoing management and maintenance.

Data Architecture to managing the data and analytics functions across the business

A critical component of an organisations ability to understand itself its systems, its data and the relationship between business and technical processes is the Data Architecture. Data architecture refers to ensuring an organized arrangement of data in a manner that optimizes the function, performance, feasibility, cost and aesthetics for the business. A strategic approach to overall architecture enables an organisation to make better decisions.

Integration of the data architecture into the business can improve effectiveness by building uniformity in the management and understanding of the current state of its systems, ensure clarity on how data should be organized and managed and promote change towards a desired future state.

There are three main components to delivering an effective data architecture:

- Outcomes - the artifacts - models, definitions, and data flows. These include:
    - The enterprise data architecture which defines the standard terms and the approach to collection, storage, integration, movement and distribution of data. Key aspects of an enterprise data architecture include data models, data definitions, data mapping specifications, data flows, structured data API’s.
    - Standards and guidelines that underpin the elements of the data architecture.

    - Data descriptions – these must include data structures, data specifications along with data flow pathways.
- Activities – actions to achieve the desired outcomes.
    - Translating business needs into data and system requirements so that processes have the data they require when and in the form they need it.
    - Facilitating data alignment across the business
    - Managing complex data flow throughout the organisation
    - Strategically enabling the organisation to evolve data management to enable business agility and transformative change.
- Behaviour – the way people within the organisation collaborate along with the mindsets and skills required for delivery.

Data modelling and flow

Data modelling is critical to data management because data models define key data that are important to the organisation, concisely capture data requirements and clarify rules and relationships that are necessary to manage and ensure the quality of data.

The Enterprise data model

This is a whole of business model that provides a consistent view of data for everyone. It includes key data generating and holding parts of the business, their relationships, critical guiding business rules and key attributes. This provides the foundational requirements for data at all levels including project level data. It needs to be agreed by all key business stakeholders. Building the enterprise data model is typically done in layers, iteratively by increments.

Data flow

Data flow design provides the requirements and master blueprint for storage and processing across databases, applications, platforms and networks. It is developed by starting with the enterprise data model and then mapping data flows across the organisation. The integrations that need to be considered when ensuring smooth data flow across the organisation are between:

- Applications within each business process.
- Data stores or databases.
- Segments of the business computer network
- Business roles, particularly those with responsibility for creating, updating, using and deleting data.
- Locations where local differences occur.

Data quality

Unless data is well managed it will deteriorate. Therefore, curation and active maintenance is essential. A lack of control over data governance results in duplication, poor interface relationships (or interface spaghetti) and uncontrolled data delivery (multiple versions of the truth), which causes inefficiencies and erodes trust in the data.

The data architecture is another vital aspect of data quality. Poor data architecture governance allows systems to evolve without clear direction, rules or control. Resulting in increasingly complex and inflexible system, creating risk for the organisation. Good data quality is achieved through strong data architecture governance with a focus on long term goals, assisted by standardization and structured development.

Driving innovation through data

Good data management enables innovation and the ability to embrace disruptive technologies. It requires a portion of the data architecture to be flexible and responsive to trailing innovative approaches in an agile manner. It also needs effective cross business engagement and a willingness to test new approaches over the short term to evaluate how they can be most effectively integrated into the enterprise data model.

Ensuring data usability

Once the data architecture and modeling have been established, then the focus should be on providing applications that make current data usable and accessible throughout the organisation. Then ensuring that data is maintained so that it continues to meet organisational requirements.

Ensuring data usability and accessibility is often the role of Database Administrators and Network Storage Administrators. Their roles involve defining storage and access requirements, developing the database instances, managing the physical storage environment, tracking usage, planning for continuity, managing data migration, tracking, through to audits and validation.

Data integration and interoperability enable the data to be moved and consolidated across the business into consistent forms across the asset lifecycle. The objective is to make data available when and where it is needed, in the right format. This should lower the cost and complexity of data sharing, and automatically trigger alters and actions. It should also support the business intelligence, analytics and master data management. The key principles for effective data integration and interoperability are:

- Take an enterprise perspective to ensure the ability to accommodate new capabilities or functionality in an incremental or iterative manner.
- Balance local data needs with enterprise data needs, including support and maintenance.
- Ensure business engagement in the design and modification of data transformation rules.

Doing this effectively requires:

- Ensuring the data is correctly updated
- Managing and minimising latency between data entry and presentation
- Ensure users receive data in a format and manner that enhances business performance
- Organisation to ensure consistency and continuity.

Data types

Master Data and Reference data are two of the base data types required to ensure consistent and unambiguous operation of the business. Both require identification of domain owners, developing rules for their operation, establishing approaches to identify and restore incorrect data and have an approach to distribute trusted data to systems across the enterprise.

Reference data – data used to characterise other data used by the organisation. It allows other data to be meaningfully understood, reducing inconsistencies and ambiguity.

Master Data – data about business entities (employees, customers, products, vendors, financial structures, assets and locations) that provides context for business transactions and analysis.

Big Data – Big data is about the integration of various types of data including structured databases and unstructured documents along with audio, video and streaming data. It provides a means to mine the data, develop predictive models and operational intelligence to improve business efficiency and customer value. Key challenges for big data are managing the volume of data, velocity of data, variety of inputs, viscosity, volatility and veracity. Metadata is a critical aspect of managing big data.

Managing unstructured data - metadata

Unstructured data takes a number of forms, typically this is data outside of a traditional database structure and can include reports, figures, diagrams, video, audio and internet files. Traditionally metadata is attached to unstructured data files to enable searchability and indexing. Metadata are key tags or identifiers that provide an external approach to structuring the data. The management of this data requires:

- Policy for different types of access and handling.
- A defined information architecture and metadata required to support a content strategy.
- Enabling management of terminology to organize, store and retrieve content.
- Capturing versioning and ensuring security of content.

Data analysis

Data analysis is undertaken though the process of developing and assessing predictive models commonly known as data science. A data analyst will develop an initial hypothesis about an action or activity. They then test the validity of this hypothesis through statistical analysis of available data. Using this information to develop predictive models based on current trends.

There are several types of predictive models. The simplest is the forecast which uses historic data treads to anticipate what will happen in the future. The next level is a probabilistic estimate which seeks to use different algorithms to predict a range of possible future outcomes. The model triggers a reaction when a particular trigger level is activated.

The highest level is prescriptive analytics. This takes predictive analytics a step further to define actions that will affect outcomes, rather than just predicted the outcomes from actions that have occurred. It anticipates what will happen, when it will happen and implies why it will happen. It can therefore suggest how to take advantage of an opportunity or avoid a risk.

Data visualization

The process of interpreting concepts, ideas and facts using pictures or graphical representations often in the form of dashboards. It helps consumers, boards and senior management ask relevant questions and set appropriate goals. In mature analytics approaches these graphical representations are integrated with scalable maps and interactive displays, enabling rapid changes in view and data presentation including drill through to view the underlying data.

Data protection, privacy, security and risk management

The goal of data security goal is to protect information assets in manner that is aligned with privacy and confidentiality regulations, contractual agreements and business requirements. Planning for data security should be undertaken at the enterprise level based on the level of risk. It will require ongoing training and monitoring to ensure effective deployment and ongoing maintenance. The best way to avoid data security breaches is to build awareness and understanding of security requirements, policies and procedures.

The key principles for data security are:

- Security is everyone’s responsibility, coordinated by a central responsible group. Over 60% of all cyber security risks are due to social engineering – influencing humans.
- It requires an enterprise approach (to provide for efficiency, consistency and minimise risks from data breaches).
- Clear accountability
- Proactive management
- Minimise sensitive/confidential data proliferation.
- Clear definitions and data classification rules. Metadata can be of assistance in categorizing data confidentiality and sensitivity.

A key aspect of data protection is an effective data security architecture. This is a subsect of the enterprise data architectures that describes how data security is implemented with the business to satisfy the business rules and external regulations. It includers tools, standards, guidelines and protocols for access, encryption, documentation, remote access and security breaches. A good data security architecture effectively integrates across the organisation between business units and systems, along with external business partners and regulatory agencies.

The steps to building an effective data security architecture are:

- Identify and classify sensitive data assets

- Locate sensitive data – affects risk and controls required.
- Determine how each data asset needs to be protected – dependent on data content and type of technology.
- Identify how this information interacts with businesses processes – what access is required and under what conditions.

Data quality

High quality data should be the primary goal of all aspects of data management. To gain value from data it must be reliable, trustworthy and fit for purpose. Data quality is therefore dependent on the context and on the needs of the data consumers. This requires a whole of organisational commitment to ensure the quality of data. It also requires awareness and coordination. Formal management of data is essential to good data quality because data can deteriorate markedly over time and through manipulation.

One of the challenges in managing the quality of data is that expectations related to quality are not always known. Customers may not articulate them. Often, the people managing data do not even ask about these requirements. But for data to be reliable and trustworthy then data management professionals need to better understand the quality requirements of their customers along with how to measure and meet them. This conversation about expectations needs to be ongoing, because requirements change over time as business needs and external forces evolve.

Enablers for good data quality management are data governance and stewardship, requiring:

- Risk and security personnel who can help identify data-related organisational vulnerabilities.
- Business process engineering and training staff who can help teams implement process improvements that increase efficiency and result in data more suitable for downstream uses.
- Business and operational data stewards and data owners who can identify critical data, define standards and quality expectations and prioritise remediation of data issues.

Ensuring the quality of data is about firstly defining what is consider high-quality data against agreed criteria. Unfortunately, there is currently no international or national standard for assessing data quality. The International Data Management Association (DAMA) suggests the following as initial assessment criteria – completeness, uniqueness, timeliness, validity, accuracy, and consistency. In addition, the following factors also impact the quality of data: usability, timing issues, flexibility, confidence and value.

Data quality issues can emerge at any point in the lifecycle. Barriers to effective data quality management include:

- Lack of awareness on the part of leadership and staff
- Lack of business governance
- Lack of leadership and management
- Difficulty in justification of improvements
- Inappropriate or ineffective instruments to measure value.

Assessing the current state of data – a maturity assessment guide

Understanding the current state of data in an organisation is most readily achieved through a maturity assessment. The following are the maturity assessment criteria outlined in the DM BOK. Essentially the organisation assesses its performance against each of the 11 elements of the Data Management Framework (Figure 1) using the following 6-point scale.

- Level 0: no capability, no organized data management practices or formal enterprise processes.
- Level 1: Initial/Ad Hoc: Data handling high reliant on a few experts limited governance, roles and responsibilities are defined within silos, controls are applied inconsistently, solutions for managing data are limited. Data quality issues are pervasive and not addressed. Assessment criteria may include the presence of any process controls such as logging of data quality issues.
- Level 2: Repeatable – emergence of consistent tools and role definitions. Starting to use centralised tools. Roles are defined and not solely reliant on specific experts. Organisational awareness of data quality issues and concepts. Recognize concepts of master and reference data management. Assessment criteria might include formal role definition in artifacts like job descriptions, the existence of process documentation and the capacity to leverage tool sets.
- Level 3: Defined – introduction of scalable data management processes and a view of data management as an organisational enabler. Replication of data across an organisation with some controls in place. A general increase in overall data quality along with coordinated policy definition and management. Reduction in manual intervention and centralised design processes making process outcomes more predictable. Assessment criteria might include the existence of data management policies, the use of scalable processes, and the consistency of data models and system controls.
- Level 4: Managed - institutional knowledge enables the organisation to predict results when approaching new projects and tasks and to begin to manage risks related to data. Data management includes performance metrics. Standardised tools for data management from desktop to infrastructure, coupled with well-formed centralised planning and governance functions. Measurable increase in data quality and enterprise capabilities, including end-to-end data audits. Assessment criteria might include metrics related to project success, operational metrics for systems and data quality metrics.
- Level 5: Optimisation – highly predictable and automated data management practices. A focus on continuous improvement. Tools enable a review of data across all processes. Data is controlled to prevent needless duplication. Well-understood metrics are used to manage and measure data quality and processes. Assessment criteria might include change management artifacts and metrics on process improvement.

References:

- Laura Sebastian-Coleman, (2018) *Navigating the Labyrinth, An Executive Guide to Data Management* for DAMA International, Technics Publications
- Laura Sebastian-Coleman (2018) *Data Management Body of Knowledge,* Edition 2, for DAMA International, Technics Publications.