

# Data Management

## 1 Introduction

**Data Management** is the development, execution and supervision of plans, policies, programs and practices that deliver, control, protect and enhance the value of data and information assets throughout their lifecycles.

A **Data Management Professional** is any person who works in any facet of data management, from highly technical to strategic business, to meet strategic organisational goals.

Data Management requires technical and non-technical skills, and business and IT people must collaborate to share responsibility for managing data. Data and information are vital to the operations of most organisations.

### 1.1 Business Drivers

- Competitive advantage of better decisions based on reliable, high quality data
- Failure to manage data is the same as the failure to manage capital
- The primary driver is to enable organisations to get value from their data assets

### 1.2 Goals

Data management goals within an organisation:

- Understanding and supporting the information needs of the enterprise including all stakeholders
- Capturing, storing, protecting and ensuring the integrity of data assets
- Ensuring quality of data and information
- Ensuring privacy and confidentiality of stakeholder data
- Preventing unauthorised or inappropriate access, manipulation or use of data
- Ensuring data can be used effectively to add value to the organisation

## 2 Essential Concepts

### 2.1 Data

Definitions of data:

- IT: Information that has been stored in digital form
- Facts which can be aggregated, analysed and used to make a profit, improve health or influence public policy
- Data is a means of representation which stands for things other than itself (Chisholm, 2010)
- Data is both an interpretation of the objects it represents and an object that must be interpreted (Sebastian-Coleman, 2013)
- To interpret data, context or Metadata is needed

### 2.2 Data and Information

- Data does not simply exist; it has to be created.
- Knowledge is required to create data in the first place
- Data is a form of information and information is a form of data
- Organisations may differentiate between information and data for communication between different stakeholders

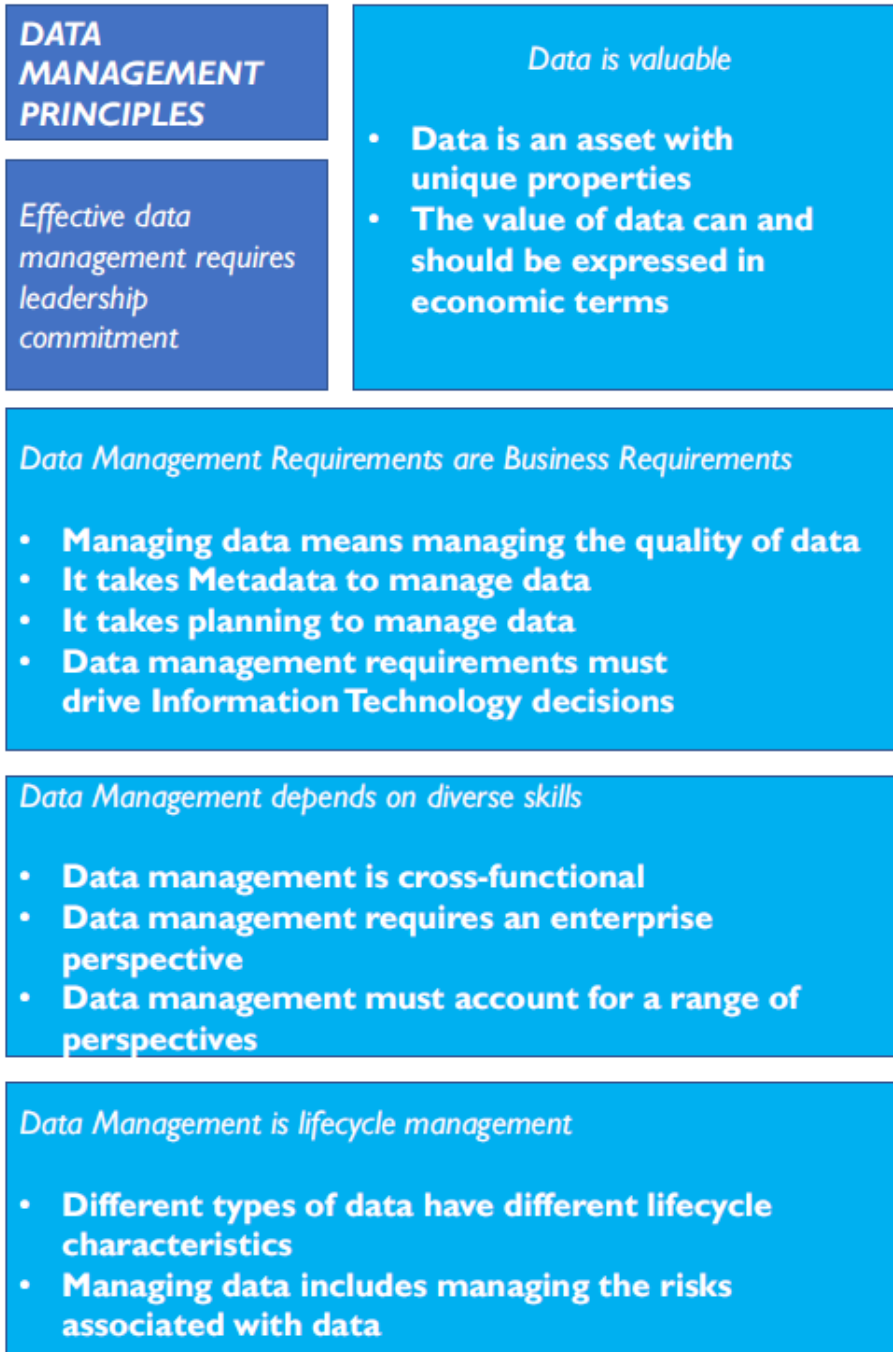
## Chapter 1

- The terms “data” and “information” are used interchangeably in the DMBOK

### 2.3 Data as an Organisational Asset

An asset is an economic resource, that can be owned or controlled, and that holds or produces value. As organisations become more reliant on data to make decisions the value of data assets is more clearly established.

### 2.4 Data Management Principles



- **Data is an asset with unique properties:** Not consumed as it is used unlike financial or physical assets.

## Chapter 1

- **The value of data can and should be expressed in economic terms:** No standards yet. Organisations should develop consistent ways to quantify value. Measure costs of low quality as well as benefits of high quality data.
- **Managing data means managing the quality of data:** Understand stakeholders' requirements for quality and ensure data is fit for their purpose.
- **It takes Metadata to manage data:** Metadata enables us to understand what intangible data is, and how to use it. Metadata originates from processes related to data creation, processing and use such as architecture, modelling, stewardship, governance, DQ management, SDLC, IT, business operations and analytics.
- **It takes planning to manage data:** Complex tech and business process landscapes, with moving data, coordinating work and keeping results aligned requires planning (architectural and business process)
- **Data management is cross functional; it requires a range of skills and expertise:** Collaboration between technical and business skills
- **Data management requires an enterprise perspective:** To be effective. That is why DM and DG are intertwined.
- **Data management must account for a range of perspectives:** DM must constantly evolve to keep up with the ways data is created and used
- **Data management is lifecycle management:** DM practices need to account for the data lifecycle
- **Different types of data have different lifecycle characteristics:** DM practices need to recognise this and be flexible enough to meet different kinds of lifecycle requirements
- **Managing data includes managing the risks associated with data:** Lost, stolen, misused and ethical implications. Managed as part of the lifecycle.
- **Data management requirements must drive Information Technology decisions:** Technology serves, rather than drives an organisation's strategic data needs.
- **Effective data management requires leadership commitment:** Complex processes of DM require the vision and purpose that comes from committed leadership.

## 2.5 Data Management Challenges

### 2.5.1 Data differs from other assets

Other assets exist in one place at one time. Data is different:

- Data is not tangible
- It is durable and does not wear out
- The value changes as it ages
- Easy to copy and transport
- Not easy to reproduce if lost or destroyed
- can be stolen but not gone
- Not consumed when used
- The same data can be used by multiple people at the same time

Data has a value but it is challenging to measure it. It must be managed with care.

### 2.5.2 Data valuation

**Value** is the difference between the cost of a thing and the benefit derived from that thing. The value of data is contextual and often temporal.

Sample cost categories:

## Chapter 1

- Obtaining and storing data
- Replacing lost data
- Impact of missing data to the organisation
- Risk mitigation and potential costs of risks associated with data
- cost of improving data
- Benefits of higher quality data
- What competitors would pay for data / what data could be sold for
- expected revenue from innovative uses of data.

### 2.5.3 Data Quality

Define business needs and the characteristics that make data high quality with business data consumers. People who use data need to assume data is reliable and trustworthy as they need to use the data to learn and create value, and make business decisions. Poor quality data is costly to the organisation.

Costs of poor quality data:

- Scrap and rework
- Work-arounds and hidden correction processes
- Organisational inefficiencies or low productivity
- Organisational conflict
- Low job satisfaction
- Opportunity costs, including inability to innovate
- Compliance costs or fines
- Reputational costs

Benefits of high quality data:

- Improved customer experience
- Higher productivity
- Reduced risk
- Ability to act on opportunities
- Increased revenue
- Competitive advantage gained from insights on customers, products, processes and opportunities

### 2.5.4 Planning for better data

View data as a product and plan for quality throughout its lifecycle. Planning is a collaboration between business and IT as it involves architecture, modelling and other design functions.

### 2.5.5 Metadata and Data Management

Includes:

- Business Metadata (Ch 12)
- Technical Metadata (Ch 12)
- Operational Metadata (Ch 12)
- Metadata embedded in: (Ch 4-11)
  - data architecture
  - data models
  - Data Security requirements

## Chapter 1

- Data integration standards
- Data operational processes

### 2.5.6 Data Management is Cross Functional

Data is managed in different places within the organisation by teams that have responsibility for different phases of the lifecycle.

### 2.5.7 Establishing an Enterprise Perspective

Data is a 'horizontal' as it moves across all the 'verticals' (sales, marketing, operations) of the organisation. Stakeholders assume the organisation's data should be coherent, and the goal of managing data is to make the disparate data originating from different places fit together in common sense ways so that it is usable by a wide range of consumers.

### 2.5.8 Accounting for other perspectives

People who create data must be aware others will use it later. Plan around other potential uses of the data to account for legal and compliance regulations and prevent future misuse.

### 2.5.9 The Data Lifecycle

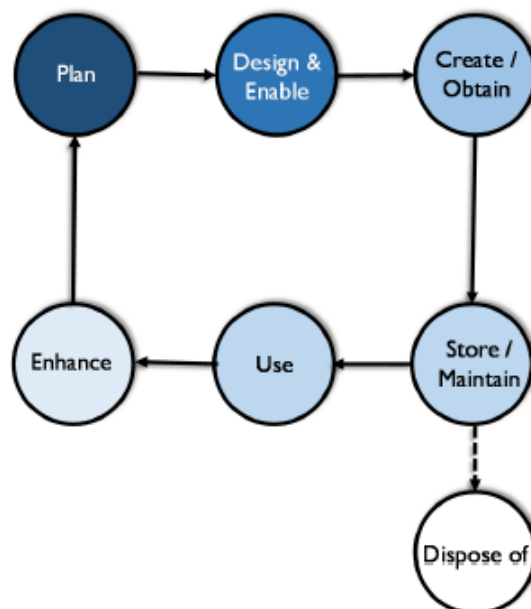


Figure 2 Data Lifecycle Key Activities

The data lifecycle is based on the product lifecycle. Throughout its lifecycle data may be cleansed, transformed, merged, enhanced or aggregate, and new data may be created. Managing data involves interconnected processes aligned with the lifecycle.

Data also has lineage, from point of origin to point of usage, which must be documented.

Implications of data lineage on the lifecycle:

- **Creation and usage are the most critical points in the data lifecycle:** it costs money to produce data and it only has value when it is used
- **Data Quality must be managed throughout the lifecycle** (see Chapter 13)
- **Metadata Quality must be managed through the data lifecycle** in the same way as the quality of other data (See Chapter 12)
- **Data Security must be managed throughout the data lifecycle:** Protection from creation to disposal (see Chapter 7)

## Chapter 1

- **Data management efforts should focus on the most critical data**

### 2.5.10 Different types of data

Classify the data to be managed as different types require different processes. Tools of data management focus on classification and control.

- By type:
  - Transactional data
  - Reference Data
  - Master Data
  - Metadata
  - Category data
  - Resource data
  - Detailed transaction data
- By content:
  - Data domains
  - Subject areas
- By format
- By the level of protection which the data requires

### 2.5.11 Data and Risk

- Low quality data represents risk as it is not right
- Data may be misunderstood and misused
- Information gaps
- Data privacy regulations
- Stakeholders requiring privacy may be broader than previously thought

### 2.5.12 Data Management and technology

Data requirements aligned with business should drive decisions about technology.

### 2.5.13 Effective data management requires Leadership and Commitment

A Chief Data Officer (CDO) leads data management initiatives and ensures data management is business driven instead of IT driven. The CDO leads cultural change required for the organisation to have a more strategic approach to its data.

## 2.6 Data Management strategy

A strategic plan is a high-level course of action to achieve high level goals. Should address all knowledge areas of the DAMA-DMBOK.

Components of a data management strategy:

- A compelling vision for data management
- A summary business case for data management with examples
- Guiding principles, values and management perspectives
- The mission and long term directional goals of data management
- Proposed measures of data management success
- Short term (12-24 months) Data Management program objectives that are SMART (Specific, Measurable, Actionable, Realistic, Time-bound)
- Description of Data Management roles, organisations and their responsibilities
- Descriptions of the Data Management program components and initiatives

## Chapter 1

- Prioritised program of work with scope boundaries
- Draft implementation roadmap with projects and action items

Deliverables:

- **Data Management Charter:** Overall vision, business case, goals, principles, measures of success, critical success factors, risks, operating model
- **Data Management Scope Statement:** Goals and objectives for usually 3 years, roles, organisations and individual leaders accountable
- **Data Management Implementation Roadmap:** Specific programs, projects, tasks and milestones

## 3 Data Management Frameworks

A framework helps to understand the data management comprehensively and see the relationships between component pieces. Frameworks are developed at different levels of abstraction and provide a range of perspectives.

Five models are presented in the DMBOK:

- The **Strategic Alignment Model** and the **Amsterdam Information Model** show high level relationships that influence how the organisation manages data
- The **DAMA DMBOK Framework** (DAMA Wheel, Hexagon and Context Diagram) describes Data Management Knowledge Areas and explains their visual representation within the DMBOK.
- The final two are rearrangements of the DAMA Wheel.

### 3.1 Strategic Alignment Model (Henderson and Venkatraman, 1999)

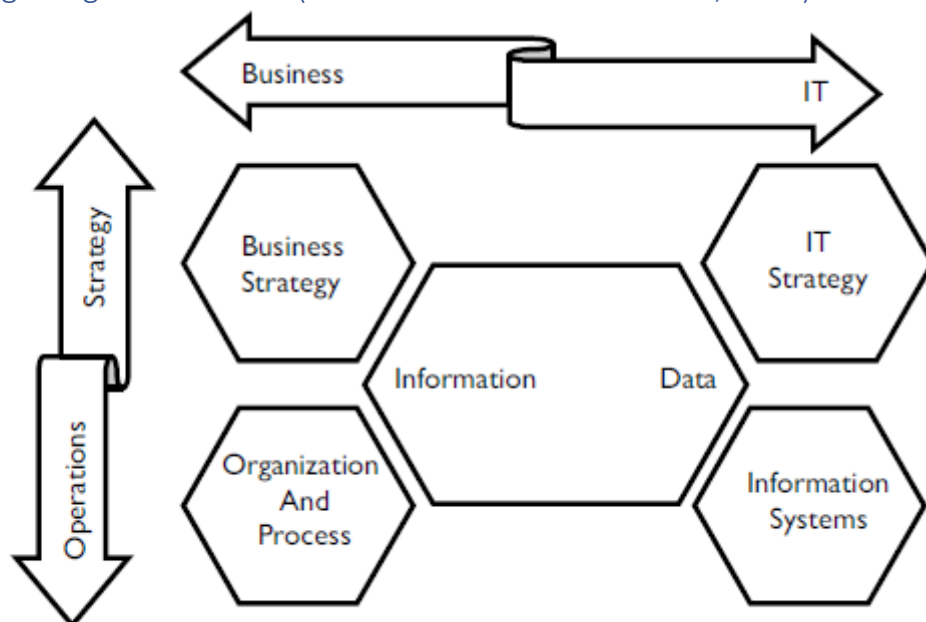
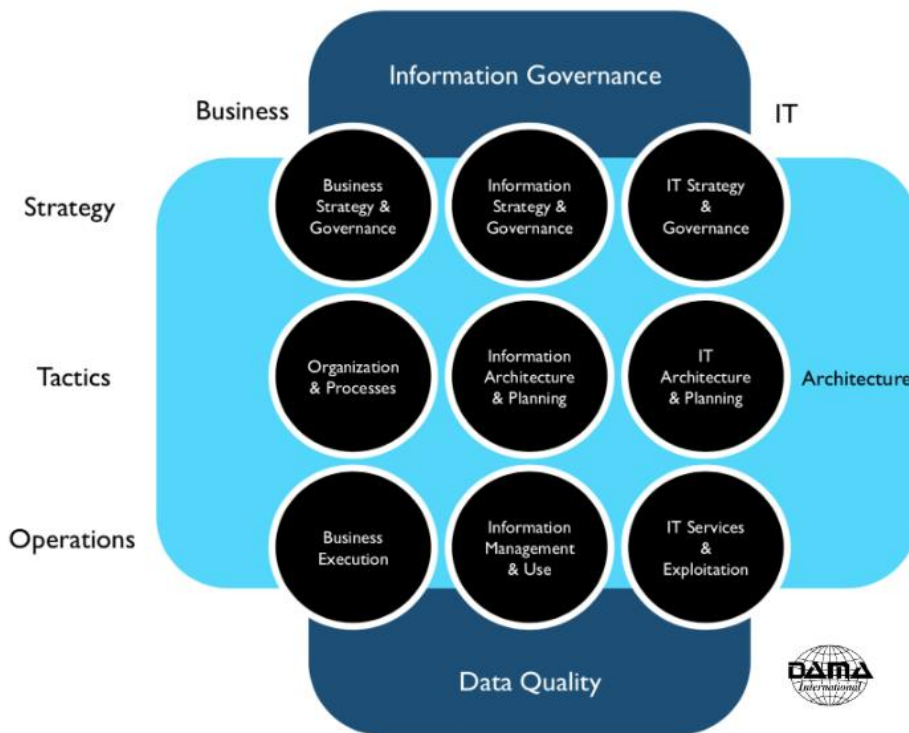


Figure 3 Strategic Alignment Model<sup>12</sup>



## Chapter 1

## 3.2 Amsterdam Information Model (Abcouwer, Maes and Truijens, 1997)



## 3.3 The DAMA-DMBOK Framework

Consists of three visuals:

- **The DAMA Wheel:** Governance in the centre for consistency within and balance between the Knowledge Areas

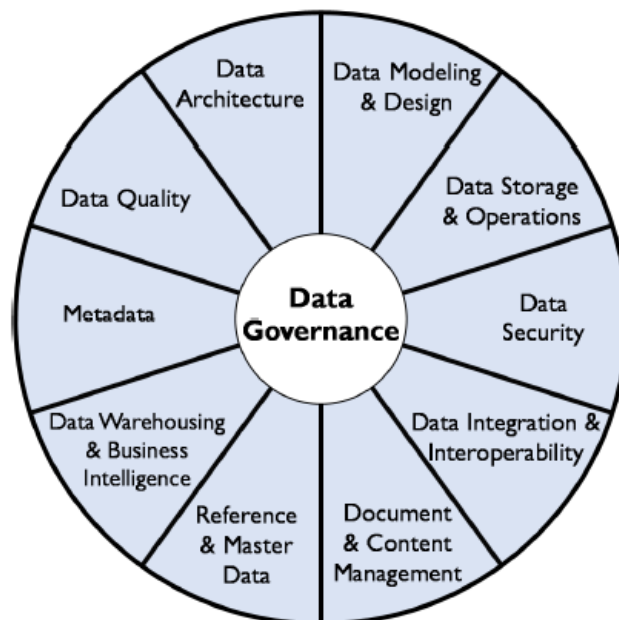


Figure 5 The DAMA-DMBOK2 Data Management Framework (The DAMA Wheel)

- **The Environmental Factors Hexagon:** Relationships between people, process and technology. Goals and principles in the centre.





Figure 6 DAMA Environmental Factors Hexagon

**The Basic Environmental Elements are:**

1. *Goals and Principles: The directional business goals of each function and the fundamental principles that guide performance of each function.*
2. *Activities: Each function is composed of lower level activities. Some activities are grouped into sub-activities. Activities are further decomposed into tasks and steps.*
3. *Deliverables: The information and physical databases and documents created as interim and final outputs of each function. Some deliverables are essential, some are generally recommended, and others are optional depending on circumstances.*
4. *Roles and Responsibilities: The business and IT roles involved in performing and supervising the function, and the specific responsibilities of each role in that function. Many roles will participate in multiple functions.*

**The supporting Environmental Elements are:**

5. *Techniques: Common and popular methods and procedures used to perform the processes and produce the deliverables. Practices and Techniques may also include common conventions, best practice recommendations, and alternative approaches without elaboration.*
6. *Tools: Categories of supporting technology (primarily software tools), standards and protocols, product selection criteria and common learning curves.*
7. *Organization and Culture: These issues might include:*
  - *Management Metrics measures of size, effort, time, cost, quality, effectiveness, productivity, success, and business value.*
  - *Critical Success Factors.*
  - *Reporting Structures.*
  - *Contracting Strategies.*
  - *Budgeting and Related Resource Allocation Issues.*
  - *Teamwork and Group Dynamics.*
  - *Authority and Empowerment.*
  - *Shared Values and Beliefs.*
  - *Expectations and Attitudes.*
  - *Personal Style and Preference Differences.*
  - *Cultural Rites, Rituals and Symbols.*

- *Organizational Heritage.*
- *Change Management Recommendations.*

| Data Management Functions                             | Goals and Principles | Activities | Primary Deliverables | Roles and Responsibilities | Technology | Practices and Techniques | Organization and Culture |
|---|----------------------|------------|----------------------|----------------------------|------------|--------------------------|--------------------------|
| Data Governance                                       |                      |            |                      |                            |            |                          |                          |
| Data Architecture Management                          |                      |            |                      |                            |            |                          |                          |
| Data Development                                      |                      |            |                      |                            |            |                          |                          |
| Data Operations Management                            |                      |            |                      |                            |            |                          |                          |
| Data Security Management                              |                      |            |                      |                            |            |                          |                          |
| Reference and Master Data Management                  |                      |            |                      |                            |            |                          |                          |
| Data Warehousing and Business Intelligence Management |                      |            |                      |                            |            |                          |                          |
| Document and Content Management                       |                      |            |                      |                            |            |                          |                          |
| Meta-data Management                                  |                      |            |                      |                            |            |                          |                          |
| Data Quality Management                               |                      |            |                      |                            |            |                          |                          |

- **The Knowledge Area Context Diagram:**

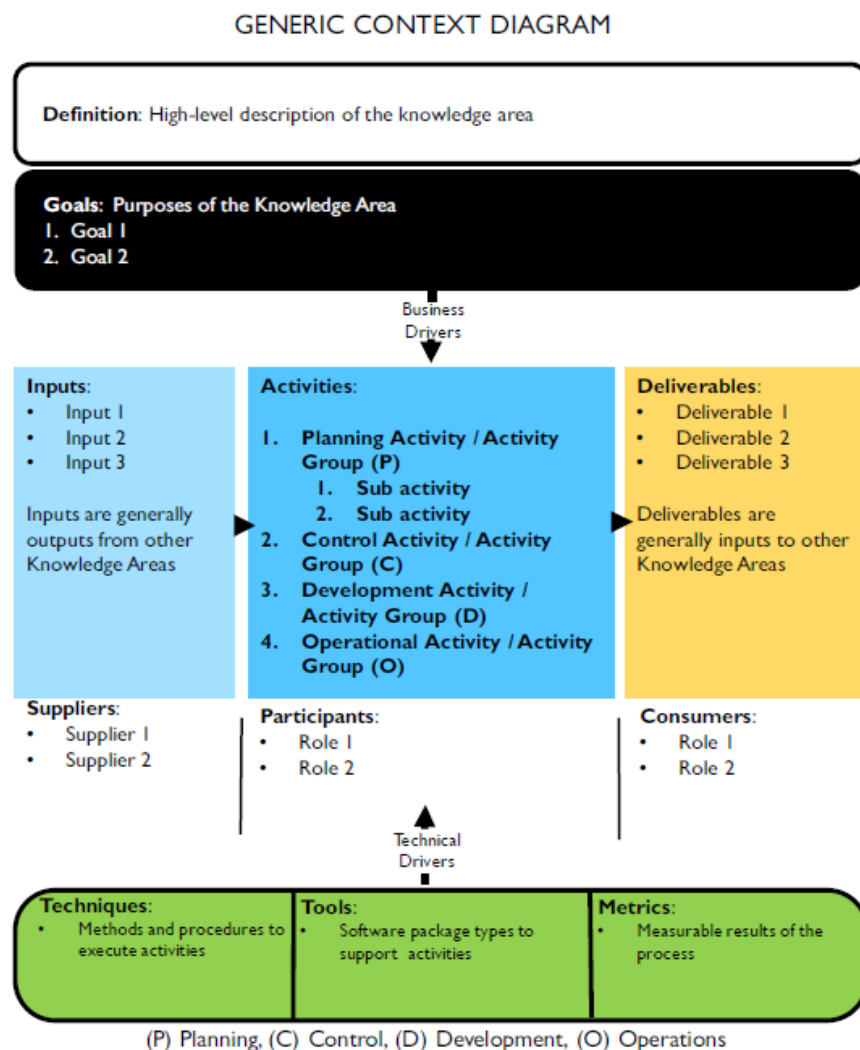


Figure 7 Knowledge Area Context Diagram

## Chapter 1

- Definition
- Goals
- Activities that drive goals classified in 4 phases:
  - (P) Planning – Strategic and tactical
  - (D) Development activities – organised around system lifecycle
  - (C) Control Activities – ensure quality, integrity, reliability, security
  - (O) Operational Activities – support systems and processes
- Inputs
- Deliverables
- Roles and Responsibilities
- Suppliers
- Consumers
- Participants
- Tools
- Techniques
- Metrics

### 3.4 DMBOK Pyramid (Aiken)

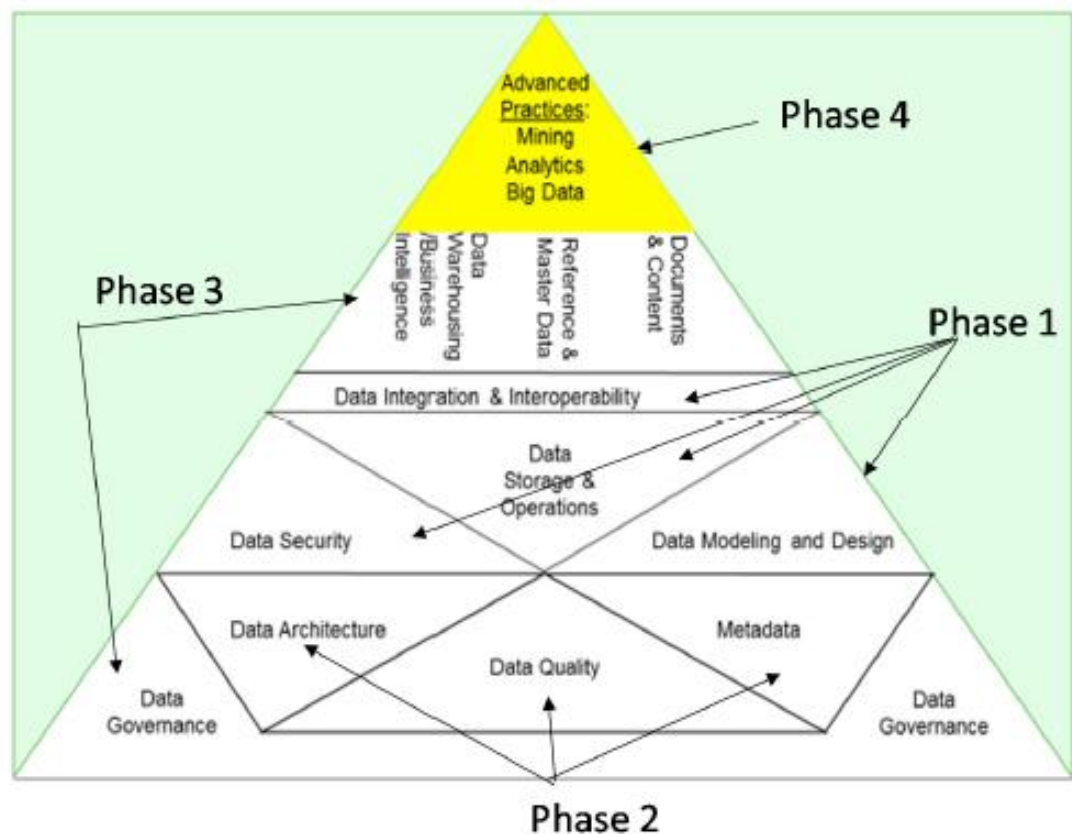


Figure 8 Purchased or Built Database Capability<sup>17</sup>

Describes how an organisation progresses to Data Management maturity:

### 3.5 DAMA Data Management Framework Evolved

Sue Geuens explored the dependencies between the knowledge areas, and developed fig 9, DAMA Functional Area Dependencies.

- BI/Analytics depend on all other functions
- DW and Master data depend on feeding systems/applications
- Depend on reliable data quality, design and integration
- Governance is the foundation on which all functions are dependent

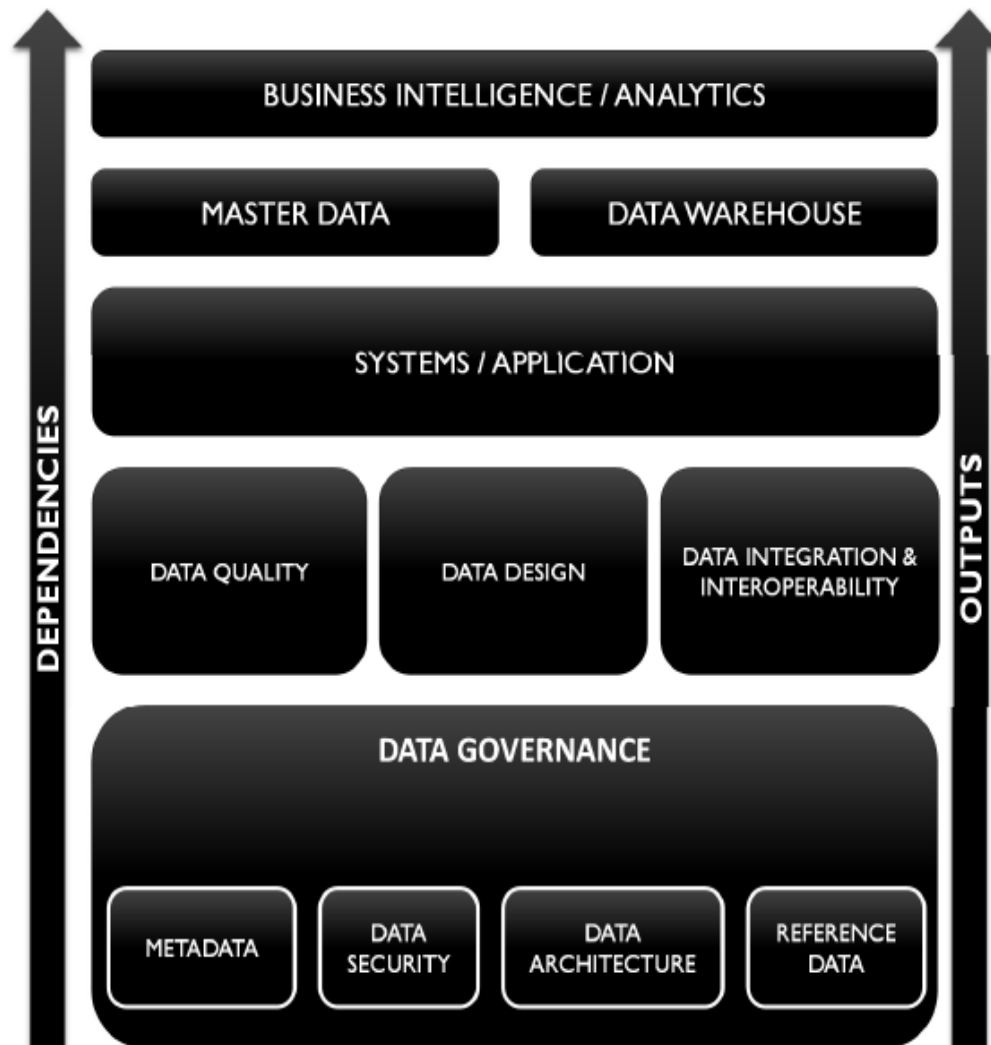


Figure 9 DAMA Functional Area Dependencies

DAMA Data Management Function Framework:

- Guiding purpose of Data Management
- Deriving value requires lifecycle management – centre of diagram
- Foundational activities support the lifecycle
- A data governance program enables the organisation to be data driven

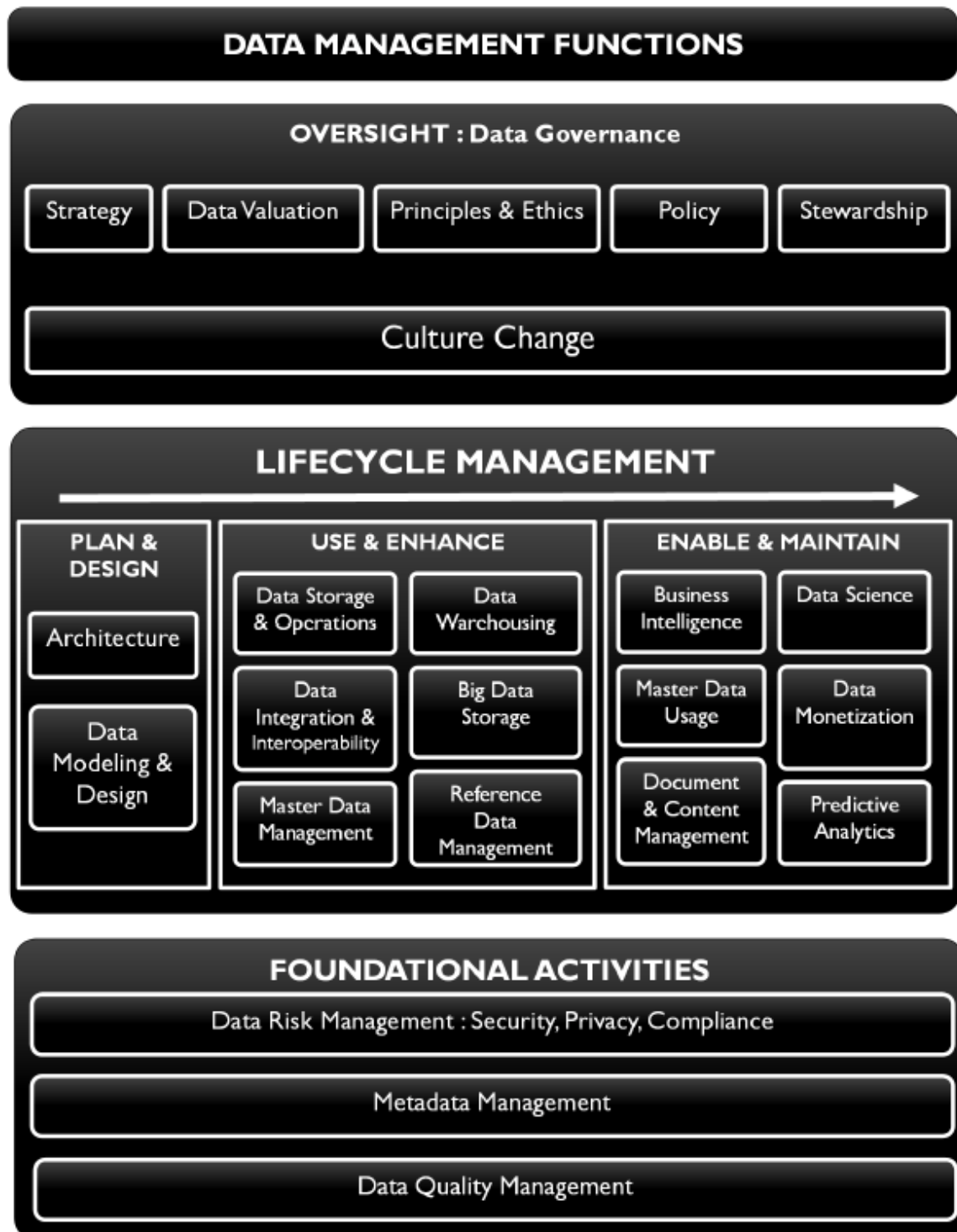


Figure 10 DAMA Data Management Function Framework

### Evolution of the DAMA Wheel:

Core activities surrounded by lifecycle and usage activities, within the strictures of governance:

- Core activities in centre
- Surrounded by lifecycle and usage activities
- Lifecycle management: Plan and Enable
- Uses come from lifecycle activities
- Data Governance provides oversight and containment

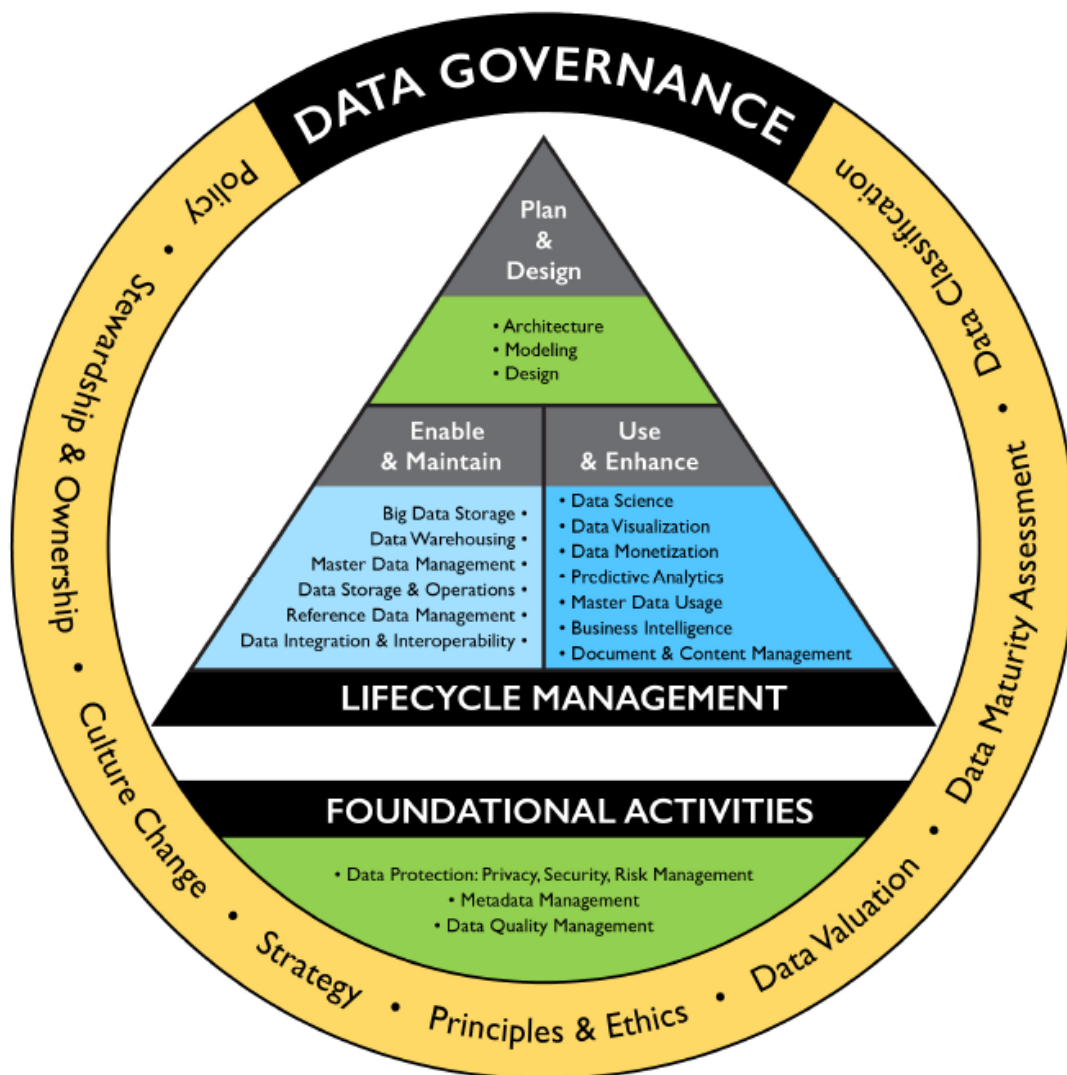


Figure 11 DAMA Wheel Evolved

### DM Functions and Data Lifecycle Management

Relationships with additional content of the knowledge areas. Data management enables organisations to get value from their data. This requires data lifecycle management, and these activities are in the centre of the diagram.



## Chapter 1

### 4 DAMA and the DMBOK

DAMA was founded to address the need for reliable data management practices.

The DMBOK supports DAMA's mission by:

- Providing a functional framework for the implementation of enterprise data management practices
- Enabling a common vocabulary for data management practices
- Serving as a fundamental guide for the CDMP exams

Knowledge Areas describe the scope and context of data management activities. They intersect with each other as data moves horizontally within an organisation.

1. **Data Governance** provides direction and oversight for data management by establishing a system of decision rights over data that accounts for the needs of the enterprise. (Ch 3 – 11%)
2. **Data Architecture** defines the blueprint for managing data assets by aligning with organisational strategy to establish strategic data requirements and designs to meet these requirements. (Ch 4 – 6%)
3. **Data Modelling and Design** is the process of discovering, analysing, representing and communicating data requirements in a precise form called a Data Model. (Ch 5 – 11%)
4. **Data Storage and Operations** includes the design, implementation and support of stored data to maximise its value. Operations provide support throughout the data lifecycle from planning to disposal of data. (Ch 6 – 6%)
5. **Data Security** ensures that data privacy and confidentiality are maintained, that data is not breached, and that data is accessed appropriately. (Ch 7 – 6%)
6. **Data Integration and Interoperability** includes processes related to the movement and consolidation of data within and between data stores, applications and organisations. (Ch 8 – 6%)
7. **Document and Content Management** includes planning, implementation and control activities used to manage the lifecycle of data and information found in a range of unstructured media, especially documents needed to support legal and regulatory compliance requirements. (Ch 9 – 6%)
8. **Reference and Master Data** includes ongoing reconciliation and maintenance of core shared data to enable consistent use across systems of the most accurate, timely and relevant version of the truth about essential business entities. (Ch 10 – 10%)
9. **Data Warehousing and Business Intelligence** includes the planning, implementation and control processes to manage decision support data to enable knowledge workers to get value from data via analysis and reporting. (Ch 11 – 10%)
10. **Metadata** includes planning, implementation and control activities to enable access to high quality, integrated Metadata, including definitions, models, data flows and other information critical to understanding data and the systems through which it is created, maintained and accessed. (Ch 12 – 11%)
11. **Data Quality** includes planning and implementation of quality management techniques to measure, assess and improve the fitness of data for use within the organisation. (Ch 13 – 11%)
12. **Data Handling Ethics** describes the central role that data ethics plays in making informed, socially responsible decisions about data and its uses. Awareness of the ethics of data collection, analysis and use should guide all data management professionals. (Ch 2 – 2%)



13. **Big Data and Data Science** describes the technologies and business processes that emerge as our ability to collect and analyse large and diverse data sets increases. (Ch 14 – 2%)
14. **Data Management Maturity Assessment** outlines an approach to evaluating and improving an organisation's data management capabilities. (Ch 15)
15. **Data Management Organisation and Role Expectations** provide best practices and considerations for organising data management teams and enabling successful data management practices. (Ch 16)
16. **Data Management and Organisational Change Management** describes how to plan for and successfully move through the cultural changes that are necessary to embed effective data management practices within an organisation. (Ch 17)

## Data Governance provides oversight and containment

### DMBOK Quote

- Data governance activities provide oversight and containment, through strategy, principles, policy, and stewardship.
- They enable consistency through data classification and data valuation

### Why is this important from a Governance point of view?

- To ensure that Data remains a strategic imperative for the business
- The Data Risks are identified & managed

### How will Data Governance do this?

- Implement a Data Governance function and supporting deliverables