IF3140 Database System

Data Management

SEMESTER I 2024/2025

KNOWLEDGE & SOFTWARE ENGINEERING

Source

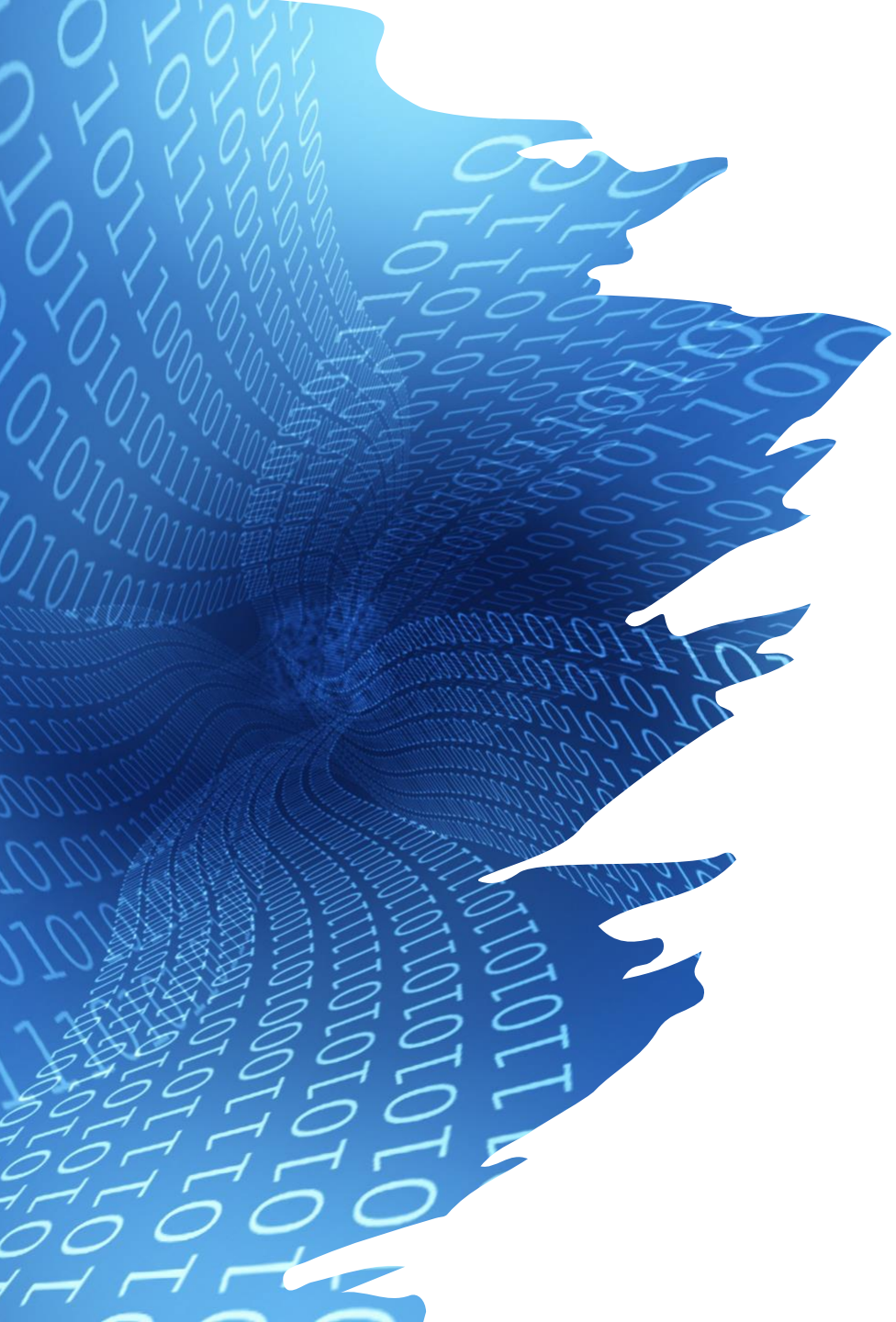
DAMA DMBOK: Data Management Body of Knowledge, 2nd Edition, Technics Publications, 2017

- Chapter 1 Data Management
- Chapter 2 Data Handling Ethics

Content

- Introduction
- Data Management
- A Data Management Professional
- Data as Organizational Asset
- Data Management Goals
- Data Management Principles and Challenges
- Data Management Frameworks
- Data Handling Ethics

KNOWLEDGE & SOFTWARE ENGINEERING



Introduction

- Many organizations recognize that their data is a **vital enterprise asset**
  - Data and information can give them insight about their customers, products, and services
  - Information and knowledge hold the key to competitive advantage
  - Organizations that have reliable and high data quality can make better decisions
  - Failure to manage data results in waste and lost opportunity

Data Management

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 lifecycle

A Data Management Professional

- Any person who works in any facet of data management to meet strategic organizational goals
  - From technical management of data throughout its lifecycle to ensuring that data is properly utilized and leveraged
  - Fill numerous roles, from the highly technical (e.g., database administrators, network administrators, programmers) to strategic business (e.g., Data Stewards, Data Strategists, Chief Data Officers)

KNOWLEDGE & SOFTWARE ENGINEERING

Data as Organizational Asset

Data is widely recognized as an **enterprise asset**

- Today's organizations rely on their data assets to make more effective decisions and to operate more efficiently

The **primary business driver** for data management is to enable organizations to get value from their data assets

- just as effective management of financial and physical assets enables organizations to get value from those assets.

Data Management Goals

- ✓Understanding and supporting the **information needs** of the enterprise and its stakeholders, including customers, employees, and business partners
- ✓Capturing, storing, protecting, and ensuring the **integrity** of data assets
- ✓Ensuring the **quality** of data and information
- ✓Ensuring the **privacy** and **confidentiality** of stakeholder data
- ✓**Preventing unauthorized or inappropriate** access, manipulation, or use of data and information
- ✓Ensuring data can be used effectively to **add value** to the enterprise

Data Management Principles and Challenges

- Data management must balance strategic and operational needs as in other management processes and can best be struck by a following **set of principles**.
- The distinct characteristics derived from the properties of data presents **challenges** following these principles.

DATA MANAGEMENT PRINCIPLES

Effective data management requires leadership commitment

Data is valuable
- Data is an asset with unique properties
- The value of data can and should be expressed in economic terms

Data management requirements are business requirements
- Managing data means managing the quality of data
- It takes metadata to manage data
- It takes planning to manage data
- Data management requirements must drive information technology decisions

Data Management depends on diverse skills
- Data management is cross-functional
- Data management requires an enterprise perspective
- Data management must account for a range of perspectives

Data management is lifecycle management
- Different types of data have different lifecycle characteristics
- Managing data includes managing the risks associated with data

Data is valuable

Data is **an asset with unique properties**

- Data is an asset, but it differs from other assets, makes it challenging to put a monetary value on data and therefore, it is difficult to measure how data contributes to organizational success
- Raise other issues in data management, such as data ownership, data inventorying, etc.

The **value of data** can and should be expressed in economic terms

- Organizations that want to make better decisions about their data should develop consistent ways to quantify that value.
- A primary challenge to data asset valuation is that the value of data is contextual and often temporal.

Data management requirements are business requirements

Managing data means managing **the quality of data**

- To ensure data meets business needs, they must work with data consumers to define these needs, including characteristics that make data of high quality
- Producing high quality data requires planning, commitment, and a mindset that builds quality into processes and systems

It takes **metadata** to manage data

- Managing any asset requires having data about that asset. The data used to manage and use data is called metadata.
- The challenge is that metadata is a form of data and needs to be managed as such.

Data management requirements are business requirements

It takes **planning** to manage data

- Deriving value from data does not happen by accident. It requires planning in many forms
- The challenge is that there are usually organizational pressures, as well as the perennial pressures of time and money, that get in the way of better planning.

Data management requirements **must drive information technology decisions**

- Successful data management requires sound decisions about technology.
- Organizations need to understand the impact of technology on data, in order to prevent technological temptation from driving their decisions about data.

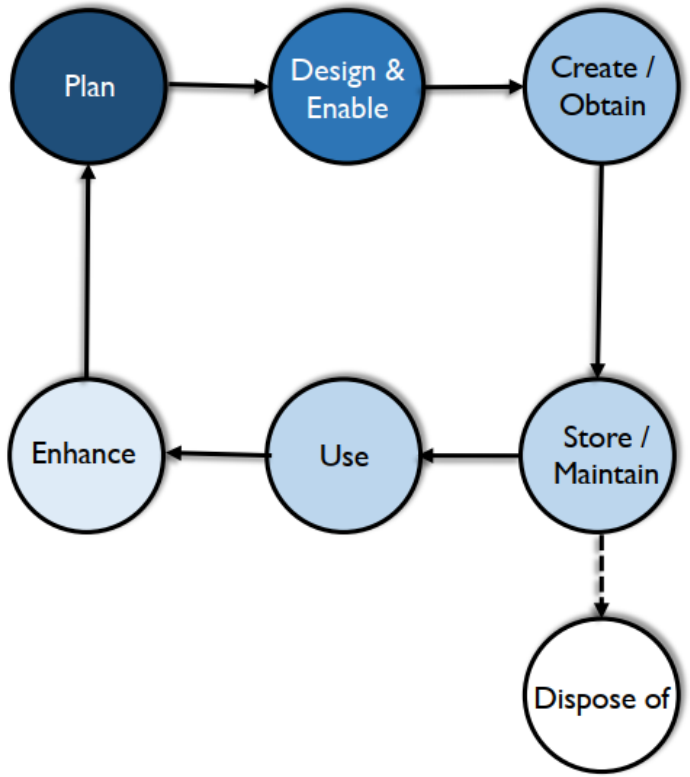
Data management depends on diverse skills

- Data management is **cross-functional**
  - Data management is a complex process. The challenge is getting people with this range of skills and perspectives to recognize how the pieces fit together so that they collaborate well as they work toward common goals.
- Data management requires an **enterprise perspective**
  - Managing data requires understanding the scope and range of data within an organization.
- Data management must **account for a range of perspectives**
  - Data management has local applications, but it must be applied across the enterprise to be as effective as possible

KNOWLEDGE & SOFTWARE ENGINEERING

Data management is lifecycle management

- **Different types of data** have different lifecycle characteristics
  - Data management practices have to recognize these differences and be flexible enough to meet different kinds of data lifecycle requirements.
- Managing data includes managing the **risks** associated with data
  - Data can be lost, stolen, or misused.
  - Organizations must consider the ethical implications of their uses of data.



The data lifecycle key activities

KNOWLEDGE & SOFTWARE ENGINEERING

Effective Data Management Requires Leadership and Commitment

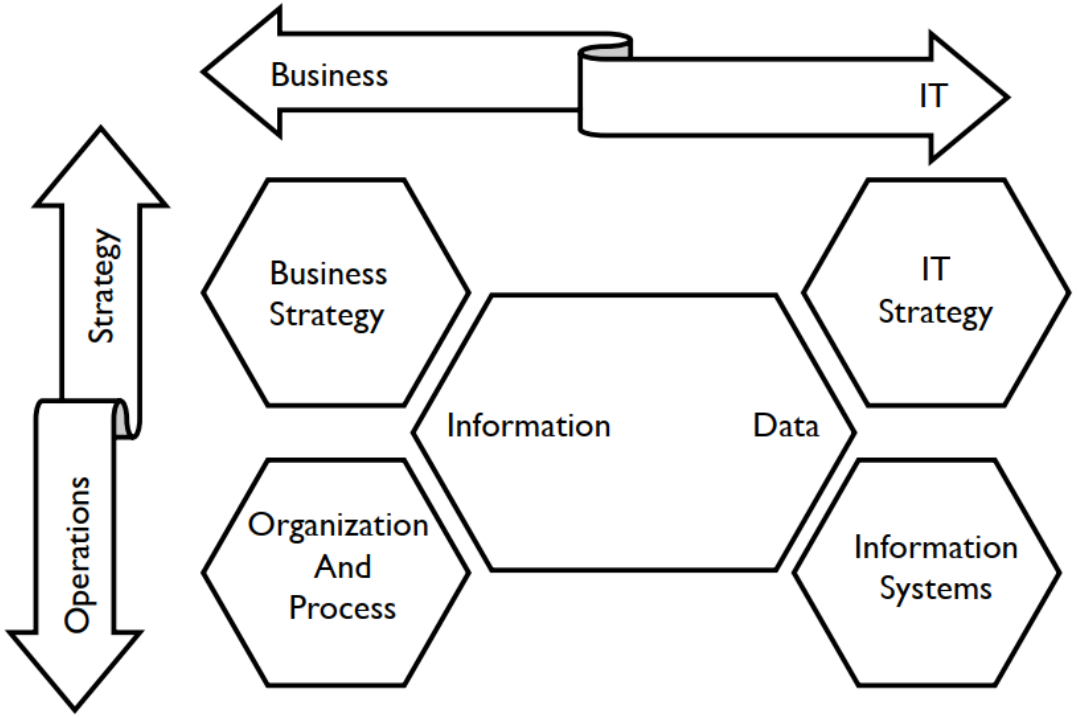
- Data management involves a complex set of processes
- To be effective, require coordination, collaboration, and commitment
- Requires not only management skills, but also the vision and purpose that come from committed leadership

Data Management Frameworks

- Frameworks developed at different levels of abstraction provide a range of perspectives on how to approach data management
  - These perspectives provide insight that can be used to clarify strategy, develop roadmaps, organize teams, and align functions
- Several frameworks to be discussed
  - Strategic Alignment Model
  - The Amsterdam Information Model
  - The DAMA-DMBOK Framework
  - DMBOK Pyramid (Aiken)
  - DAMA Data Management Framework Evolved

KNOWLEDGE & SOFTWARE ENGINEERING

Strategic Alignment Model

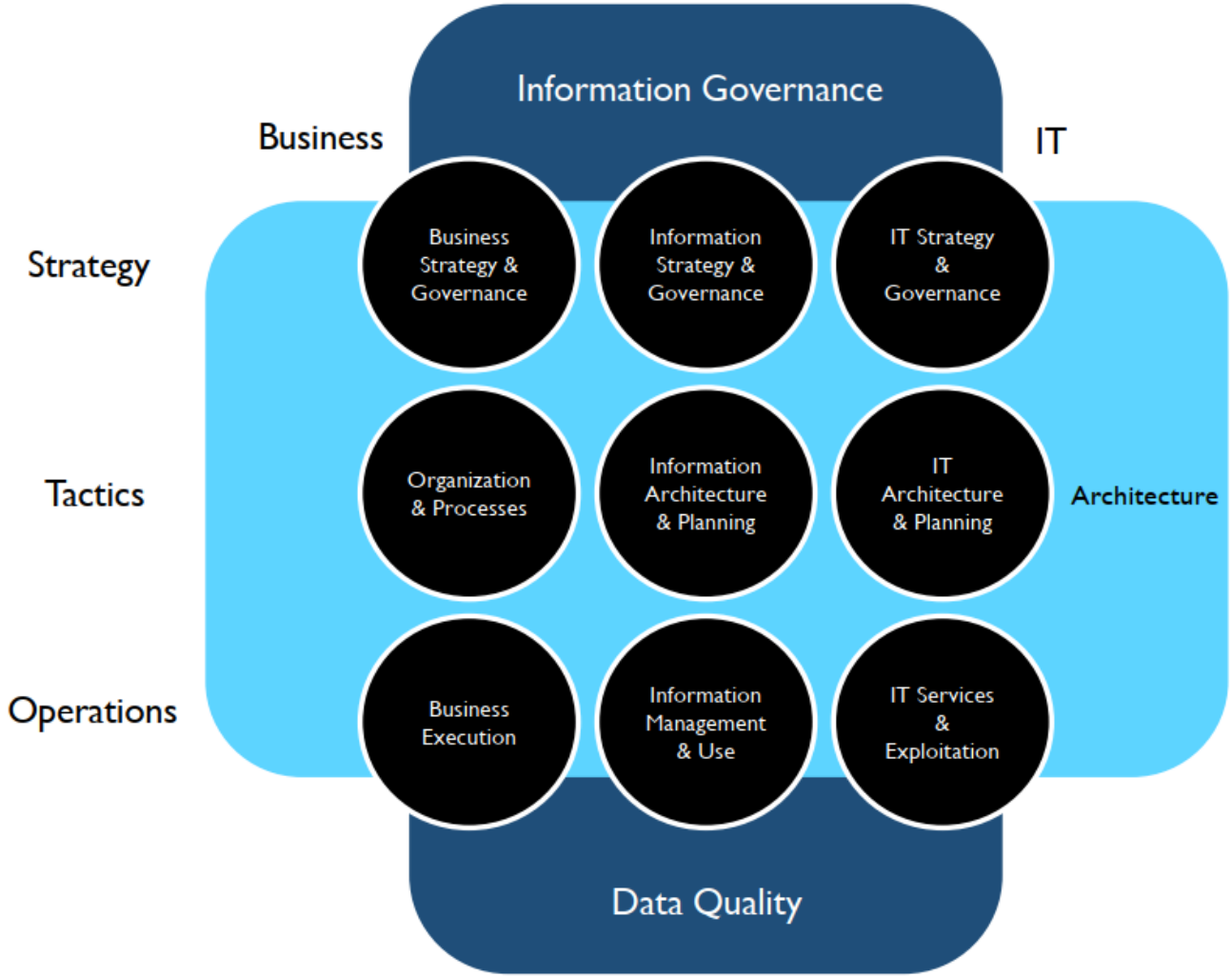


Abstracts the fundamental drivers for any approach to data management (Henderson and Venkatraman, 1999)

KNOWLEDGE & SOFTWARE ENGINEERING

The Amsterdam Information Model

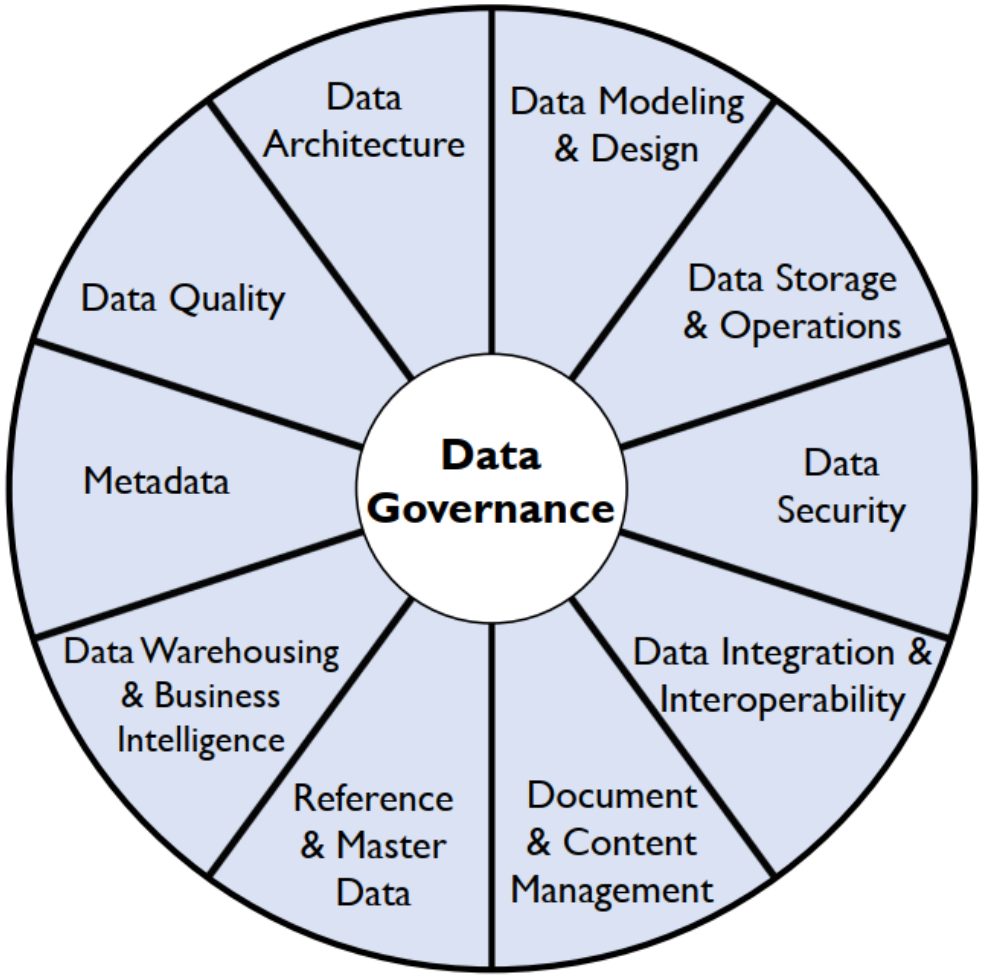
Takes a strategic perspective on business and IT alignment (Abcouwer, Maes, and Truijens, 1997)



KNOWLEDGE & SOFTWARE ENGINEERING

The DAMA DMBOK Framework

- DAMA DMBOK (Data Management Body of Knowledge) which includes DAMA DMBOK Framework was published by DAMA (Data Management Association), a global non-profit organization which focuses on information and data management.
  - 1st edition in November 2009, 2nd edition in July 2017
- Provides the context for the work carried out by data management professionals within various data management knowledge areas



The DAMA DMBOK Framework

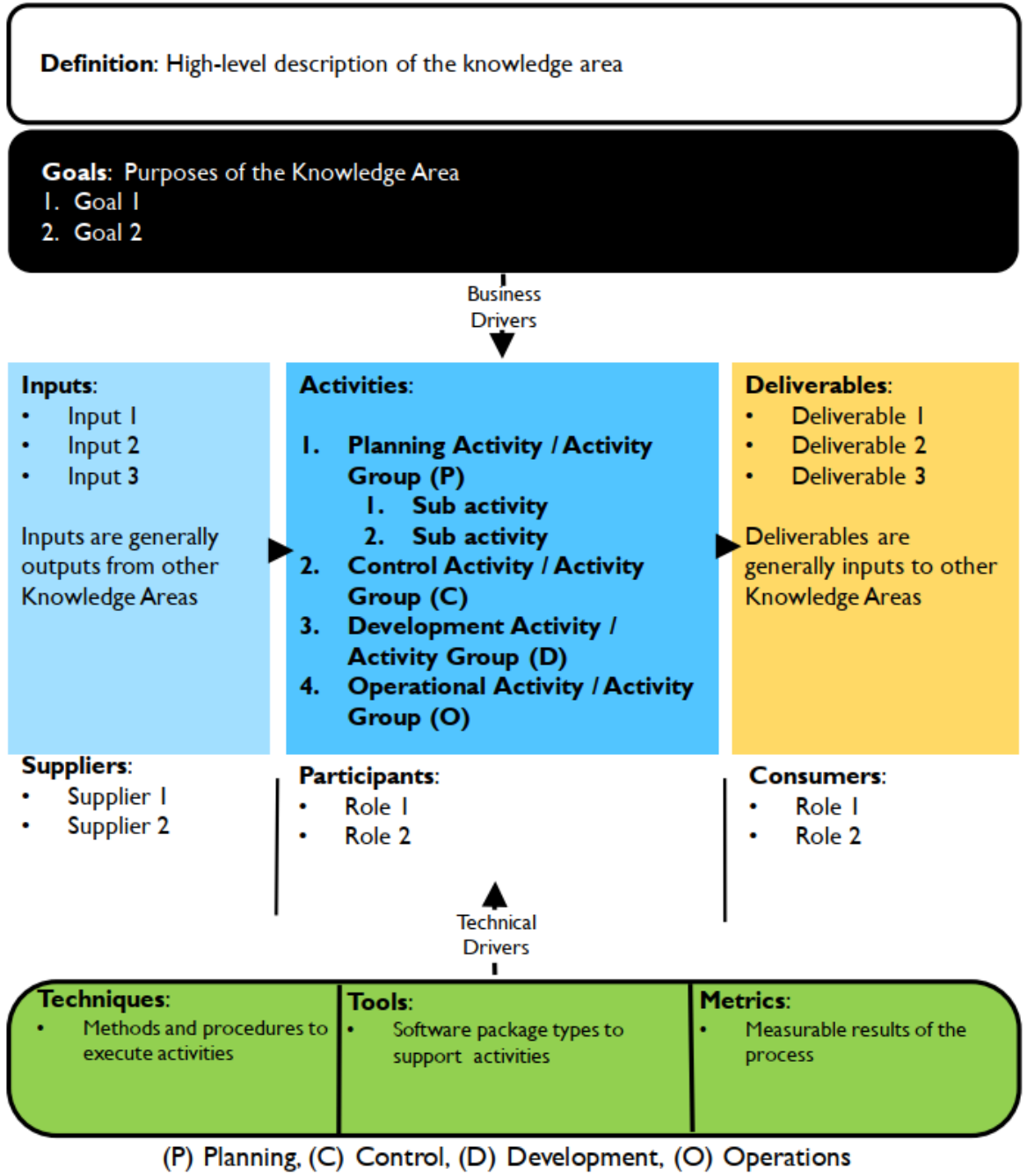
The **DAMA Wheels** defines data management knowledge areas that make up the overall scope of data management

The DAMA DMBOK Framework

The **DAMA Environmental Factors hexagon** shows the relationship between people, process, and technology

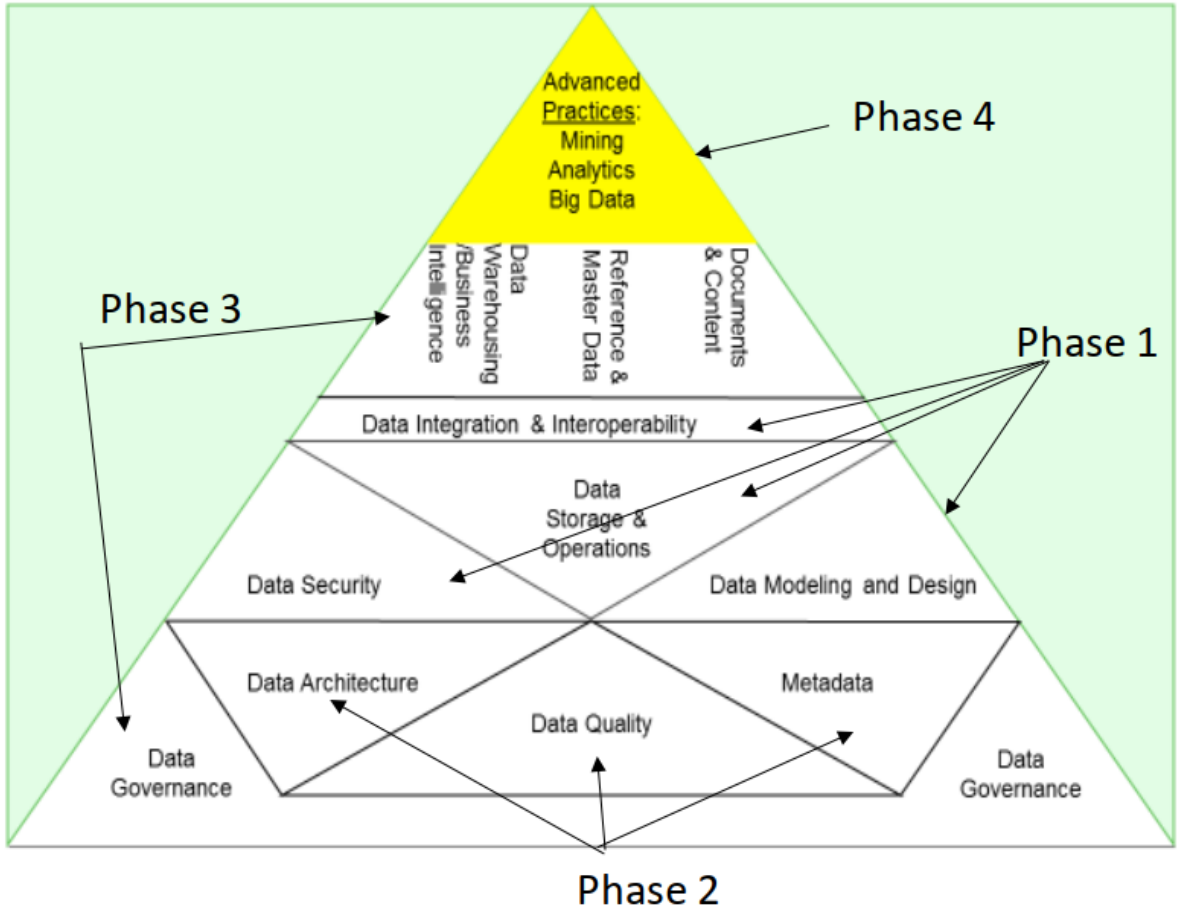
The DAMA DMBOK Framework

The DAMA Knowledge Area Context Diagrams describe the detail of the Knowledge Areas, including detail related to people, processes and technology



DAMA Pyramid (Aiken)

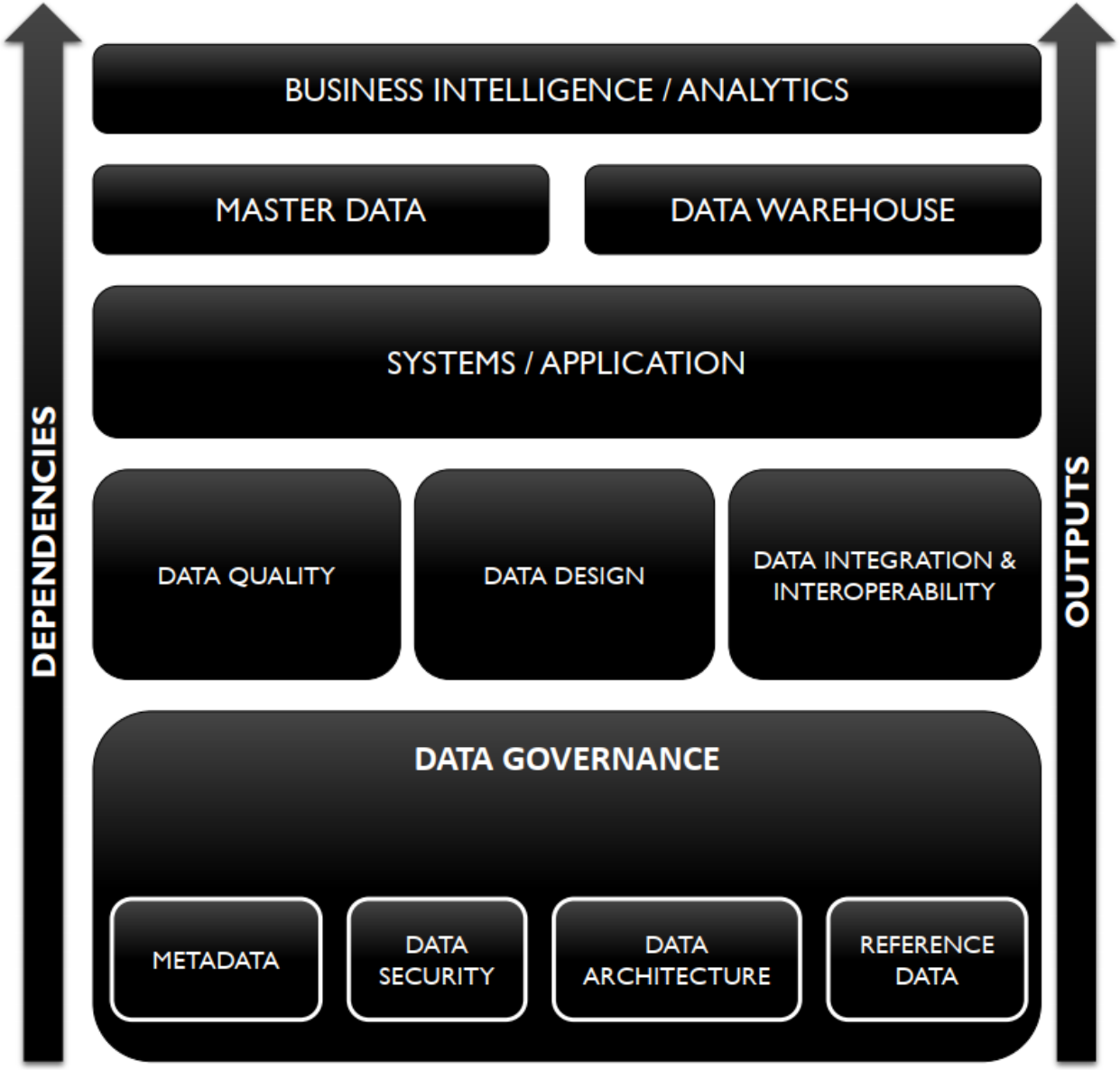
- Peter Aiken's framework uses the DMBOK **functional areas** to describe the situation in which many organizations find themselves
- Aiken's pyramid draws from the **DAMA Wheel**, but also informs it by showing the relation between the Knowledge Areas



KNOWLEDGE & SOFTWARE ENGINEERING

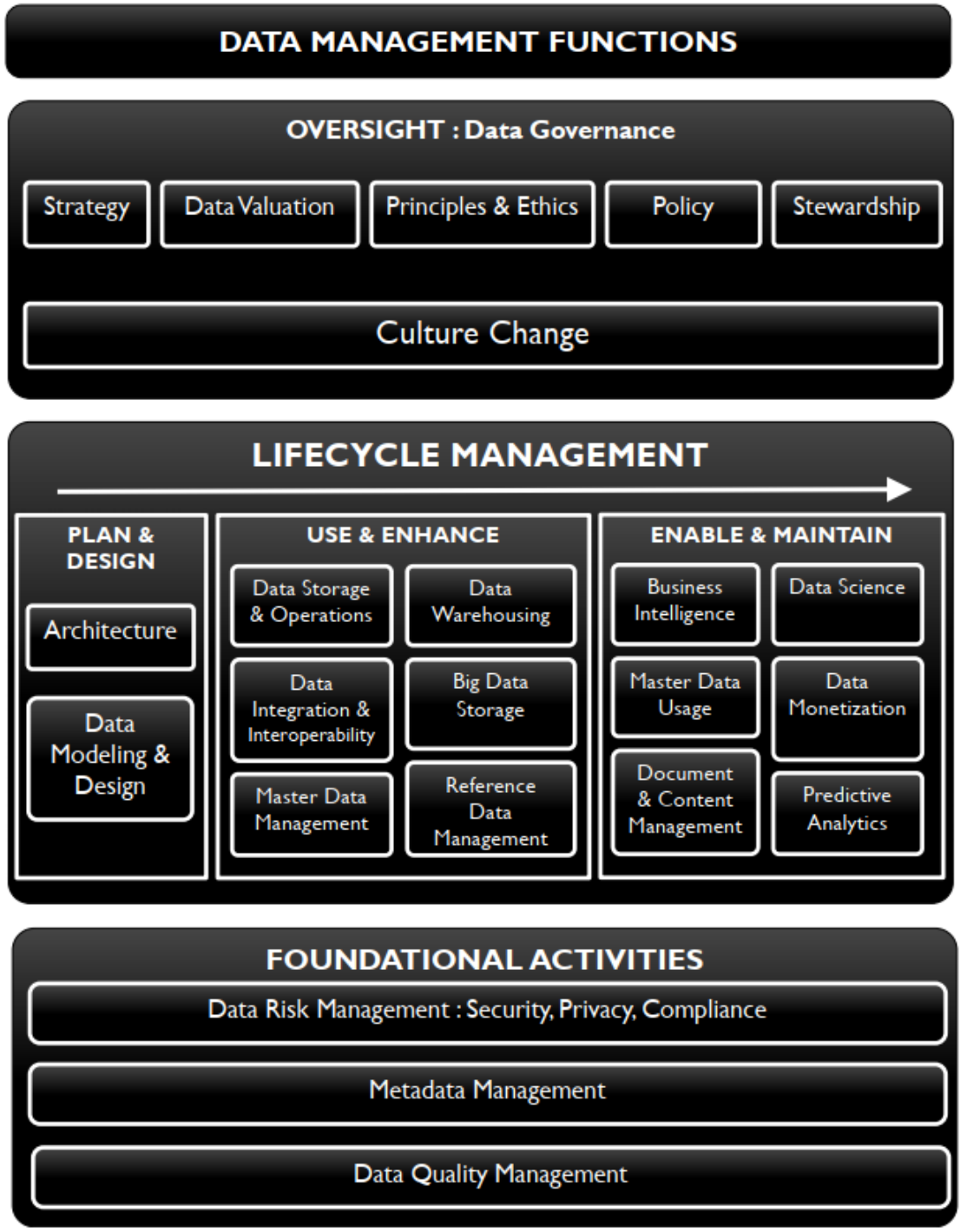
DAMA Data Management Framework Evolved

- Another way to look at the DAMA Knowledge Areas is to explore the dependencies between them.
- Developed by Sue Geuens, the framework recognizes that Business Intelligence and Analytic functions have dependencies on all other data management functions



DAMA Data Management Framework Evolved

- A third alternative to DAMA Wheel draws on architectural concepts to propose a set of relationships between the DAMA Knowledge Areas.
- It provides additional detail about the content of some Knowledge Areas in order to clarify these relationships.

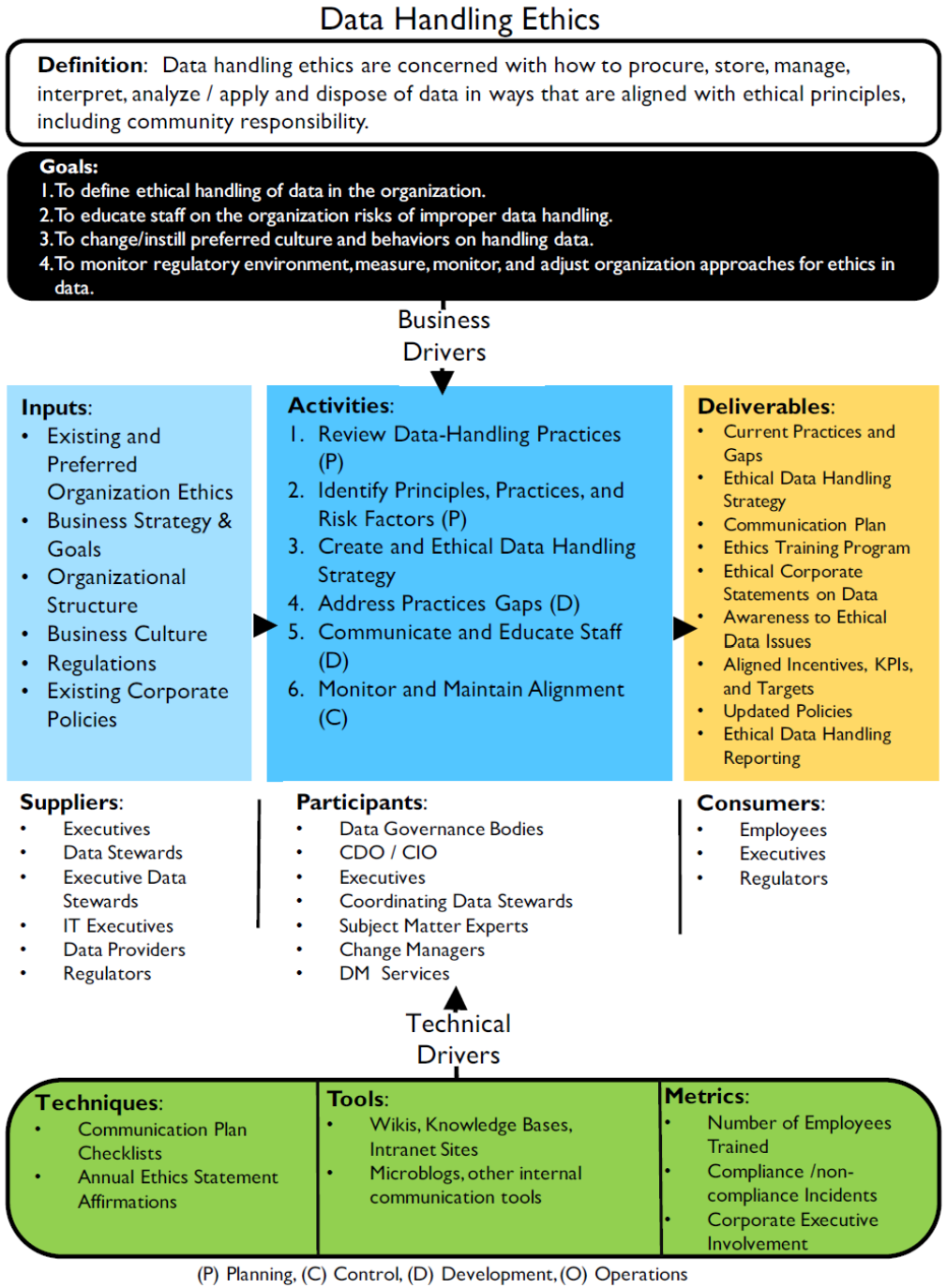


Data Handling Ethics

- Data handling ethics are concerned with how to procure, store, manage, use, and dispose of data in ways that are aligned with ethical principles.
- Center on several core concepts:
  - Impact on people
  - Potential for misuse
  - Economic value of data

KNOWLEDGE & SOFTWARE ENGINEERING

DAMA DMBOK Context Diagram: Data Handling Ethics



Ethical Principles for Data

- **Respect for persons**:
  - This principle reflects the fundamental ethical requirement that people be treated in a way that respects their dignity and autonomy as human individuals. It also requires that in cases where people have 'diminished autonomy', extra care be taken to protect their dignity and rights
- **Beneficence**:
  - This principle has two elements: first, do not harm; second, maximize possible benefits and minimize possible harms.
- **Justice**:
  - This principle considers the fair and equitable treatment of people
- Possible fourth principle: **Respect for law and public interest**

KNOWLEDGE & SOFTWARE ENGINEERING

Some risks of unethical data handling

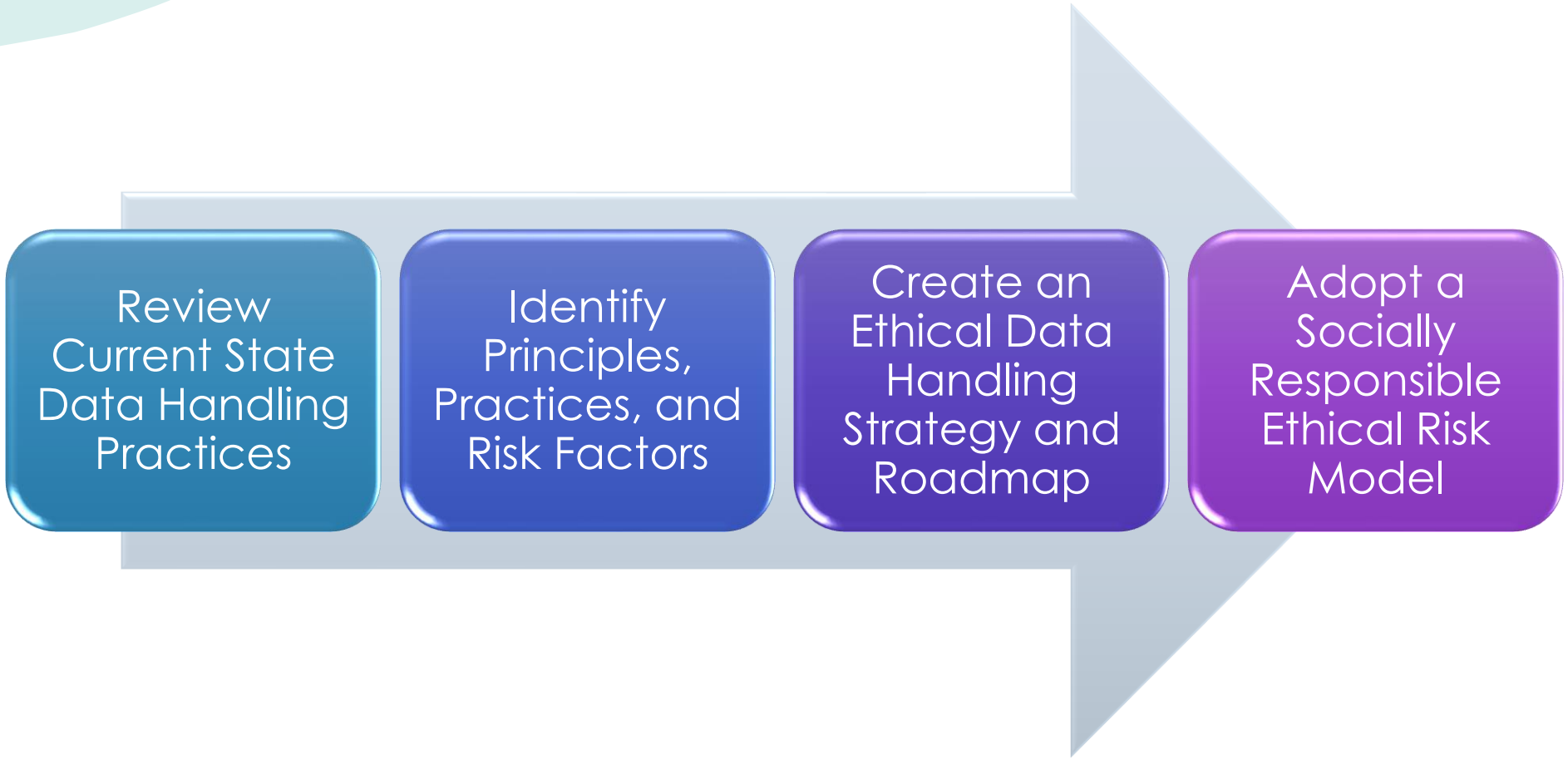
- People can lie through omission or inclusion of certain data points in a report or activity based on timing.
- Charts and graphs can be used to present data in a misleading manner. For instance, changing scale can make a trend line look better or worse
- Unclear definitions or invalid comparisons in presenting information may imply meaning that the data does not support.
- Bias can be introduced at different points in the data lifecycle: when data is collected or created, when it is selected for inclusion in analysis, through the methods by which it is analyzed, and in how the results of analysis are presented
- If data is not transformed and integrated with care, it presents risk for unethical or even illegal data handling.
- Obfuscating or redacting data is the practice of making information anonymous, or removing sensitive information. It may not be sufficient to protect data if a downstream activity can expose the data.

KNOWLEDGE & SOFTWARE ENGINEERING

GDPR Principle	Description of Principle
Fairness, Lawfulness, Transparency	Personal data shall be processed lawfully, fairly, and in a transparent manner in relation to the data subject.
Purpose Limitation	Personal data must be collected for specified, explicit, and legitimate purposes, and not processed in a manner that is incompatible with those purposes.
Data Minimization	Personal data must be adequate, relevant, and limited to what is necessary in relation to the purposes for which they are processed.
Accuracy	Personal data must be accurate, and where necessary, kept up-to-date. Every reasonable step must be taken to ensure that personal data that are inaccurate, having regard to the purpose for which they are processed, are erased or rectified without delay.
Storage Limitation	Data must be kept in a form that permits identification of data subjects [individuals] for no longer than is necessary for the purposes for which the personal data are processed.
Integrity and Confidentiality	Data must be processed in a manner that ensures appropriate security of the personal data, including protection against unauthorized or unlawful processing and against accidental loss, destruction or damage, using appropriate technical or organizational measures.
Accountability	Data Controllers shall be responsible for, and be able to demonstrate compliance with [these principles].

General Data Protection Regulation of the EU

Establishing an ethical data culture



KNOWLEDGE & SOFTWARE ENGINEERING

Data Ethics and Governance

- Data governance is a vital tool for ensuring the data ethics principles are considered
  - in deciding who can do what with which data and under what circumstances processing is appropriate or necessary.
- The ethical impacts and risks of data processing on all stakeholders must be considered by practitioners.
- Data Governance must set standards and policies for and provide oversight of data handling practices.

KNOWLEDGE & SOFTWARE ENGINEERING