Digital Technologies

Data governance framework in an organization

Giulio Nicelli

giulio.nicelli@polimi.it

Agenda

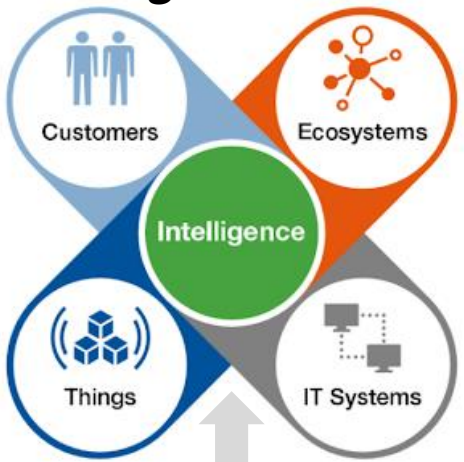
Managing Data and Data strategy adoption

Elements of Data Governance: logical and physical models

Setup of a Data Framework: an example

Digital Business requires to manage, analyze and respond to a huge amount of data

Digitalisation



Volume

Speed

Variety

Scale
How Much, How Fast and How Different?

Distribution
How Far and How Widely?

Pervasive data

Distributed processing

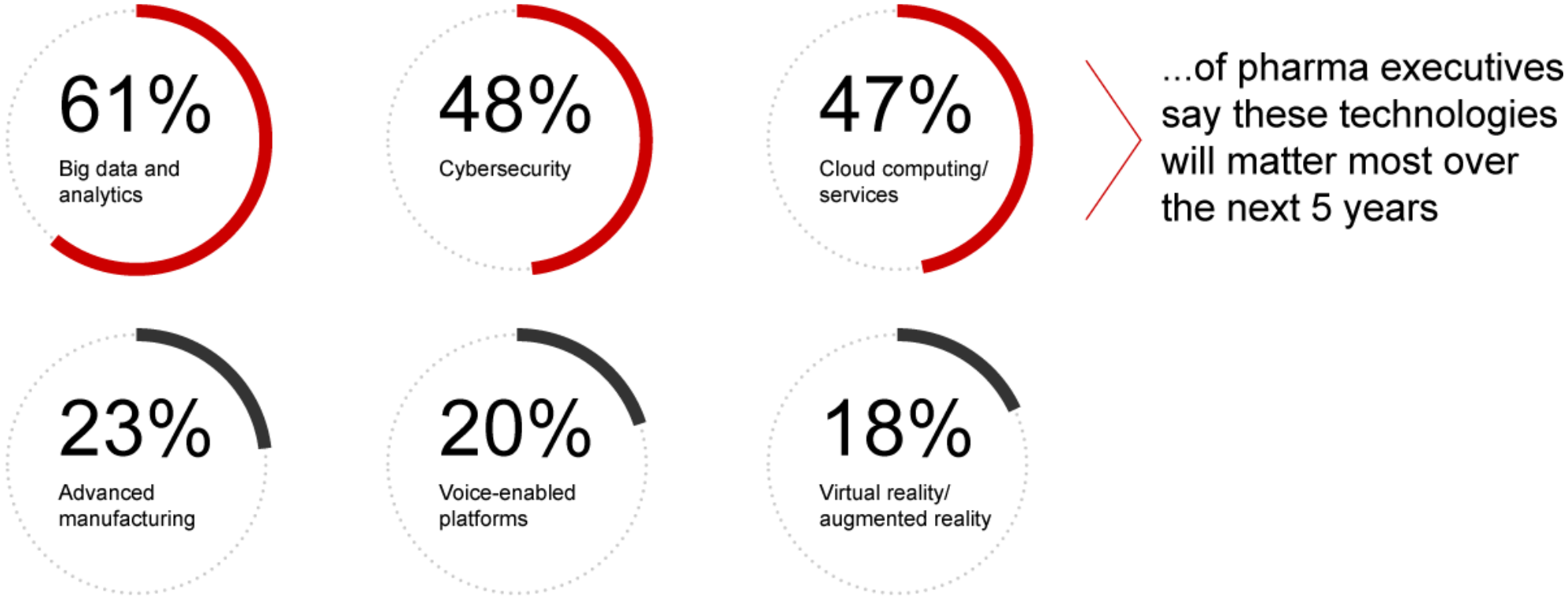
Hyperconnected

… low Data Quality will have a profound impact on Business Value

Source: Gartner

The use of data will be the first source of technological impact

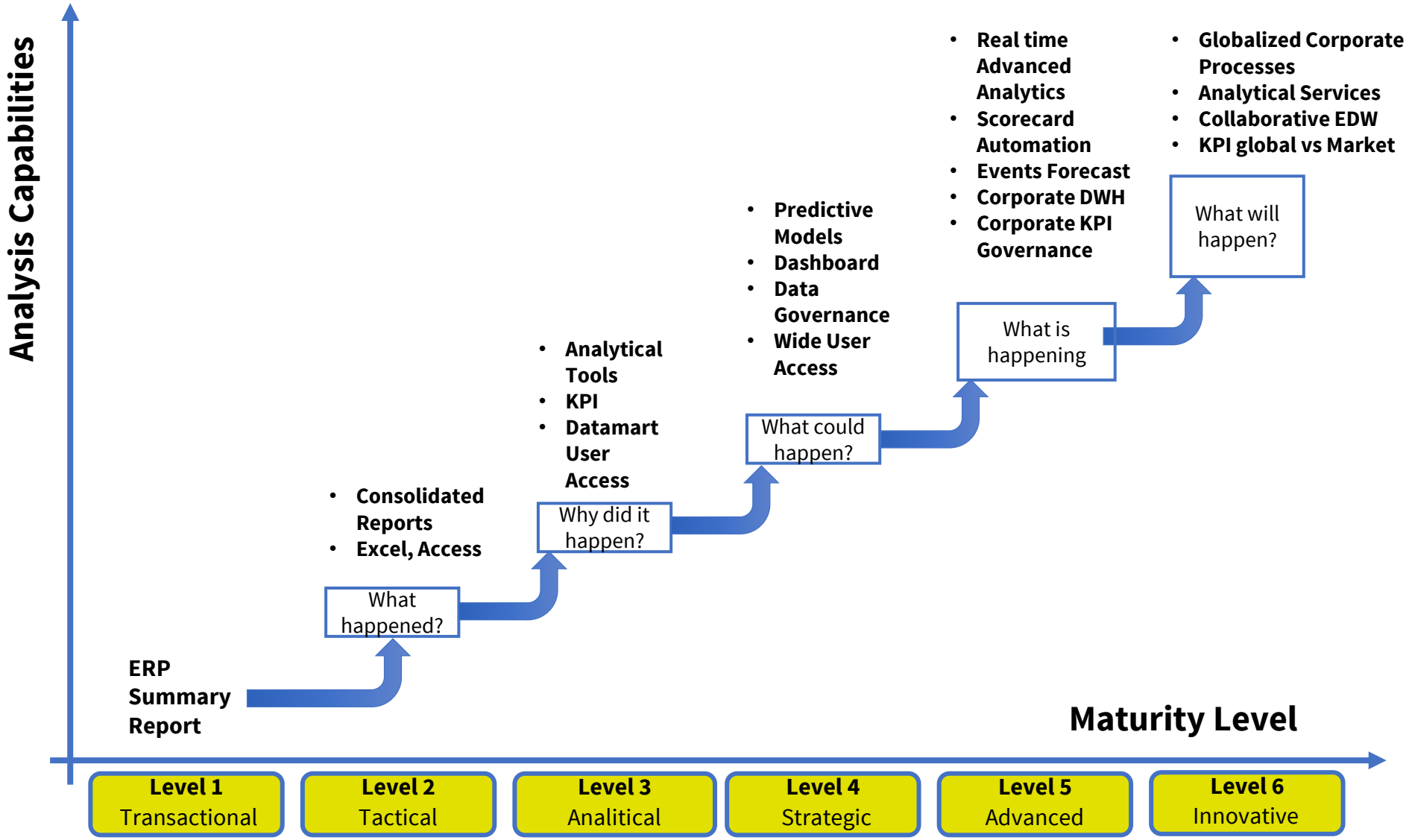
Pharma executives say big data, cybersecurity and cloud services will have the biggest impact on the industry



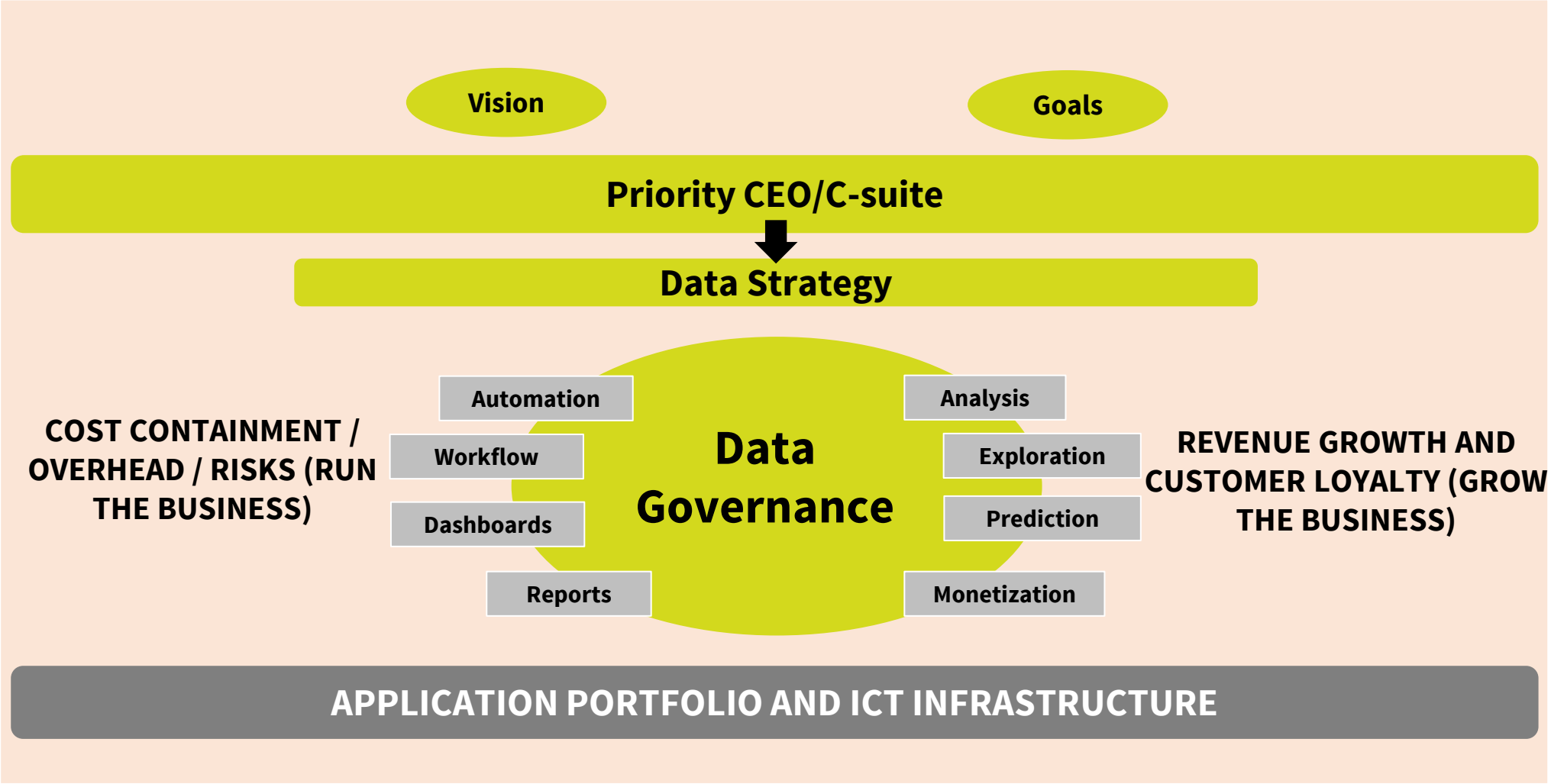
Notes: Respondents were asked to select up to 5 digital trends that would have the most impact on their firm's business activities over the next 5 years; respondents were also asked for most important trends of the last 5 years; trend acceleration is calculated as (next 5 average) – (last 5 average) with a margin 1 percentage point
Source: Bain Digital Insights Survey,

BAIN & COMPANY

The evolution in the use of data



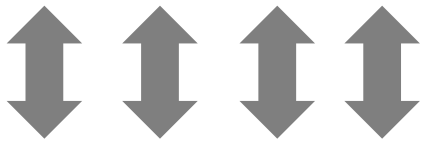
Manage Data in innovative companies



Intervention Areas

Data Strategy. **Define a clear vision** on the **use of data** in the company and integrate it into the definition of the innovation projects portfolio, starting with the priorities of the **CEO and C-Suites**

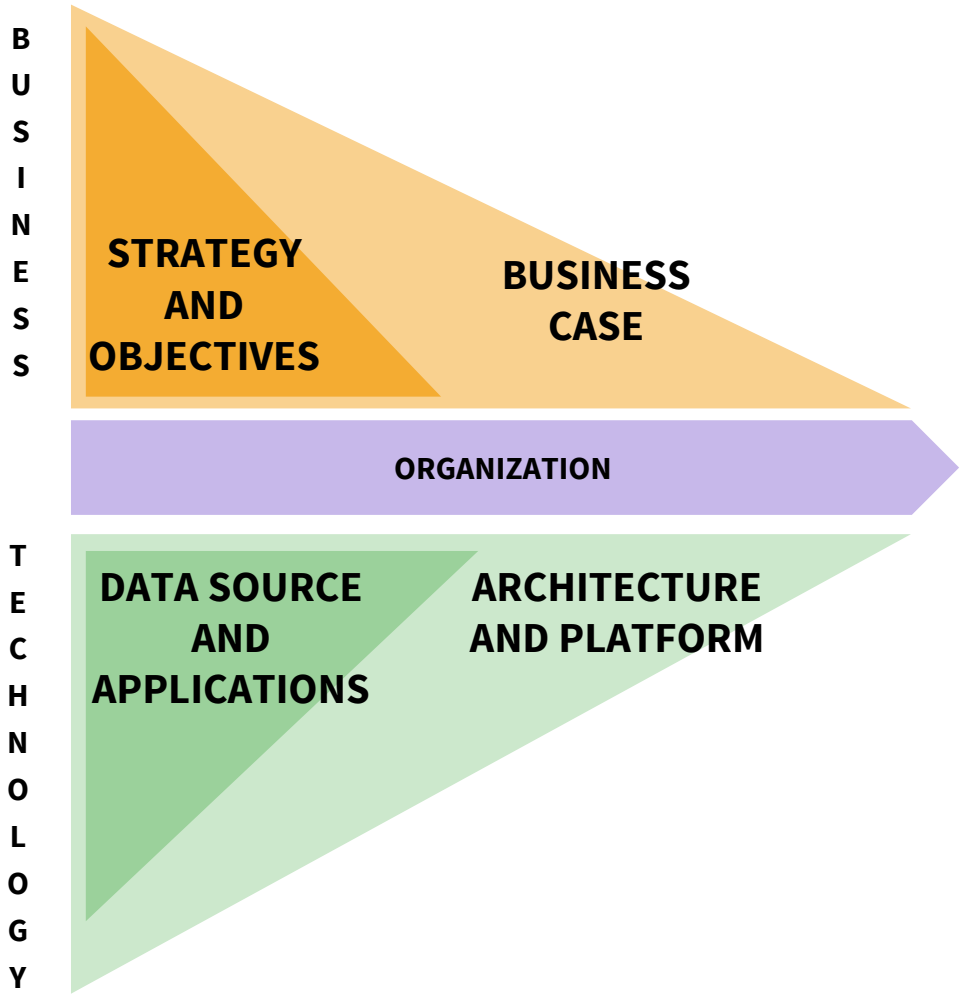
- Target the **data domain of value** for the implementation of the business strategy
- Define and develop the **business case**
- Define **goals** and metrics to measure results
- Orienting long-term **architectural and technological choices**



Data Governance. Define a **data management** system that guarantees **quality** and correct use, throughout the entire **life cycle**, from acquisition, archiving to cleaning and sharing

- Guarantee the **integrity, consistency and quality** of the data assets
- Identify the **roles and skills** needed to manage the data assets
- Organize the internal team with adequate **processes and tools**

Data Strategy: main focus



Strategy and Goals

- Business objectives definition and prioritization
- Identification of insights for building business cases
- Detection of the level of corporate maturity in the use of data

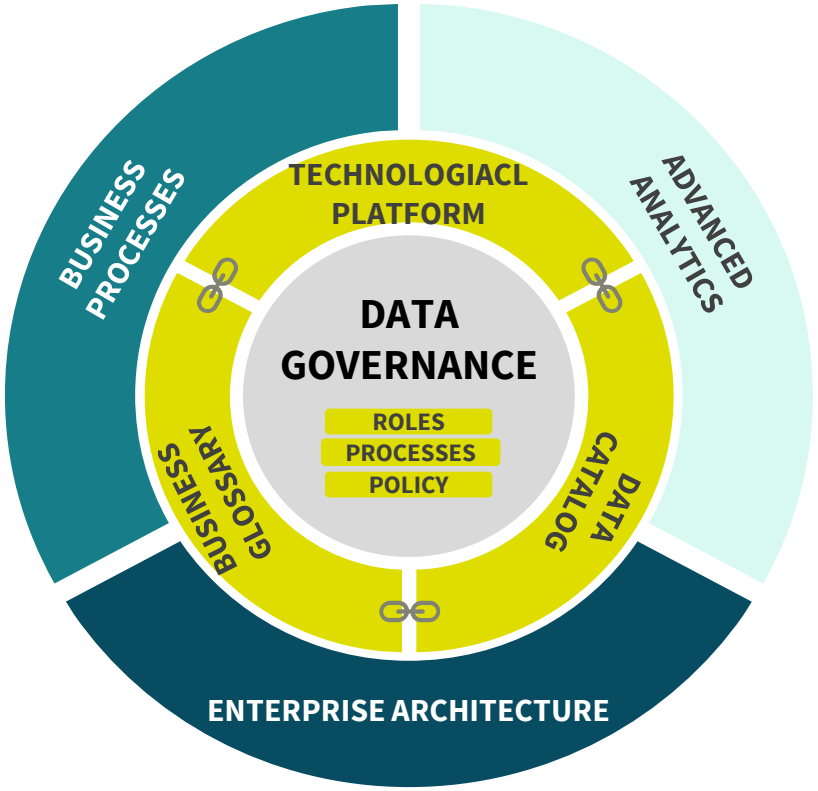
Feasibility Study

- Accessibility and quality of company data
- Data management technology architecture
- How to view and access data

Implementation Plan

- Priority Business Case to be built
- Technological evolution interventions necessary
- Creation of a Data Governance structure

Data Governance: main focus



Entire organization	Business Area/Domain
Data Platform/DataHub	Specific data driven project

Organization, processes, policies and standards

- Definition of roles and responsibilities of individuals and teams
- Definition of data management processes and its life cycle
- Definition of data management rules and standards

Data Architecture

- Alignment with corporate strategies
- Coordination and direction of overall architecture
- Supervision, direction, monitoring of processes, projects and initiatives

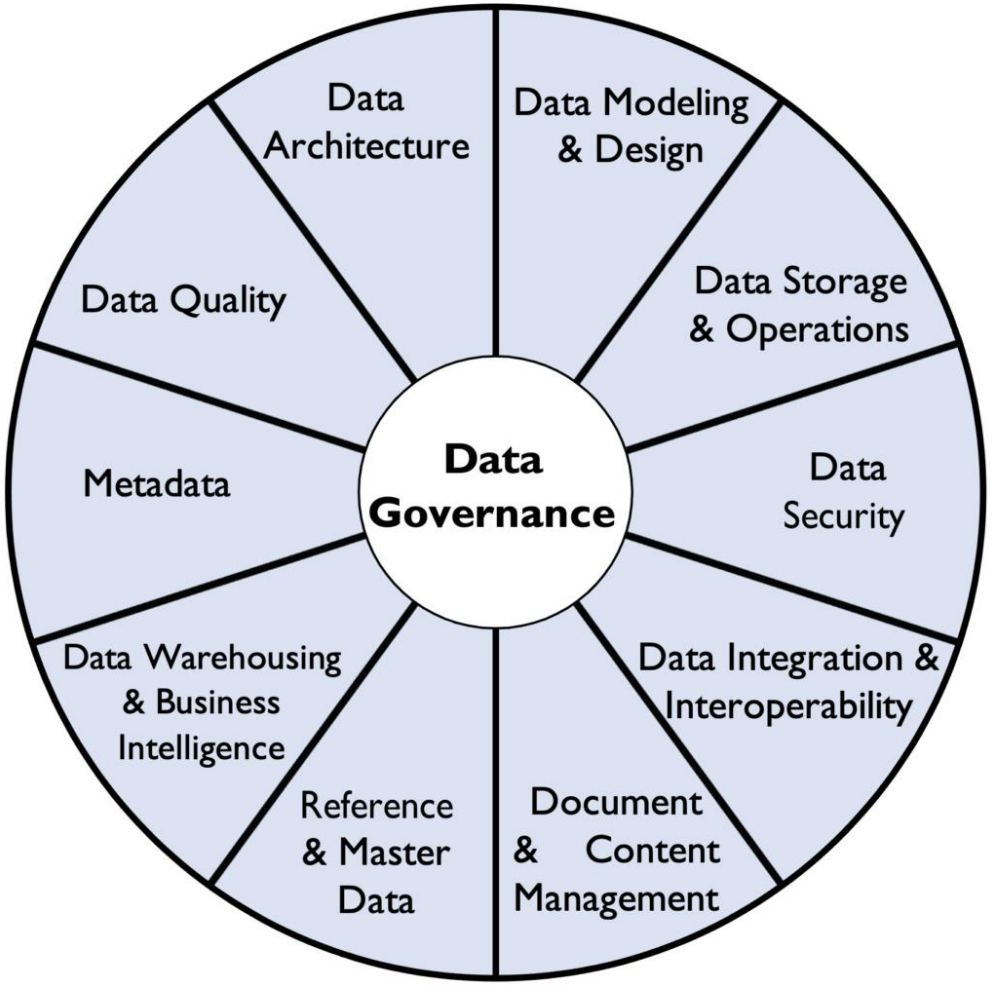
Technological architecture and Data Integration

- Data Lake technology stack presidium
- Integrations technology stack protection

Tools

- Applications and systems to support data governance processes:
  - Business Glossary / Data Catalog / Data Governance Platform
  - Data platform / Data integration / API Management Tools

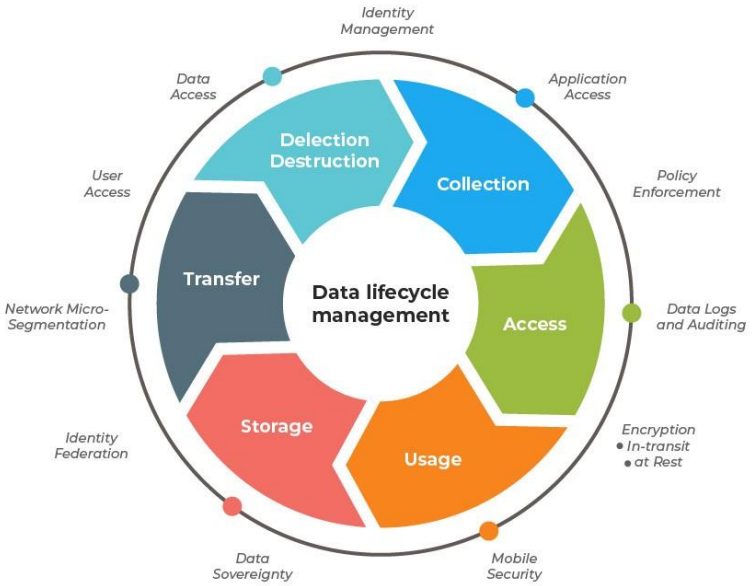
DAMA Reference Framework



The **DAMA Framework**, short for Data Management Body of Knowledge, serves as a comprehensive guideline for data management professionals. Developed by DAMA International, it provides a **structured approach to managing an organization's data assets**, encompassing a wide range of disciplines such as data governance, data architecture, and data quality management. In today's lecture, we will explore the ten core knowledge areas of the DAMA Framework, which aim to ensure that **data is managed as a valuable resource.**

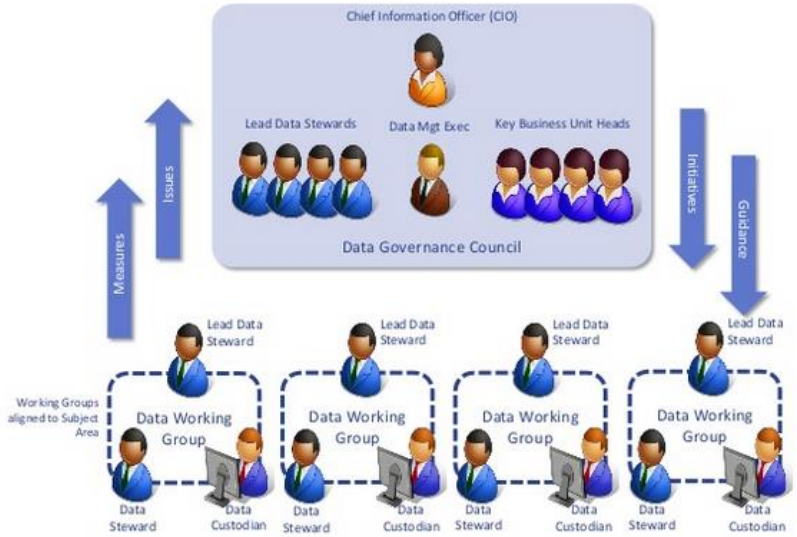
Data management is transversal within the company

Data Life Cycle

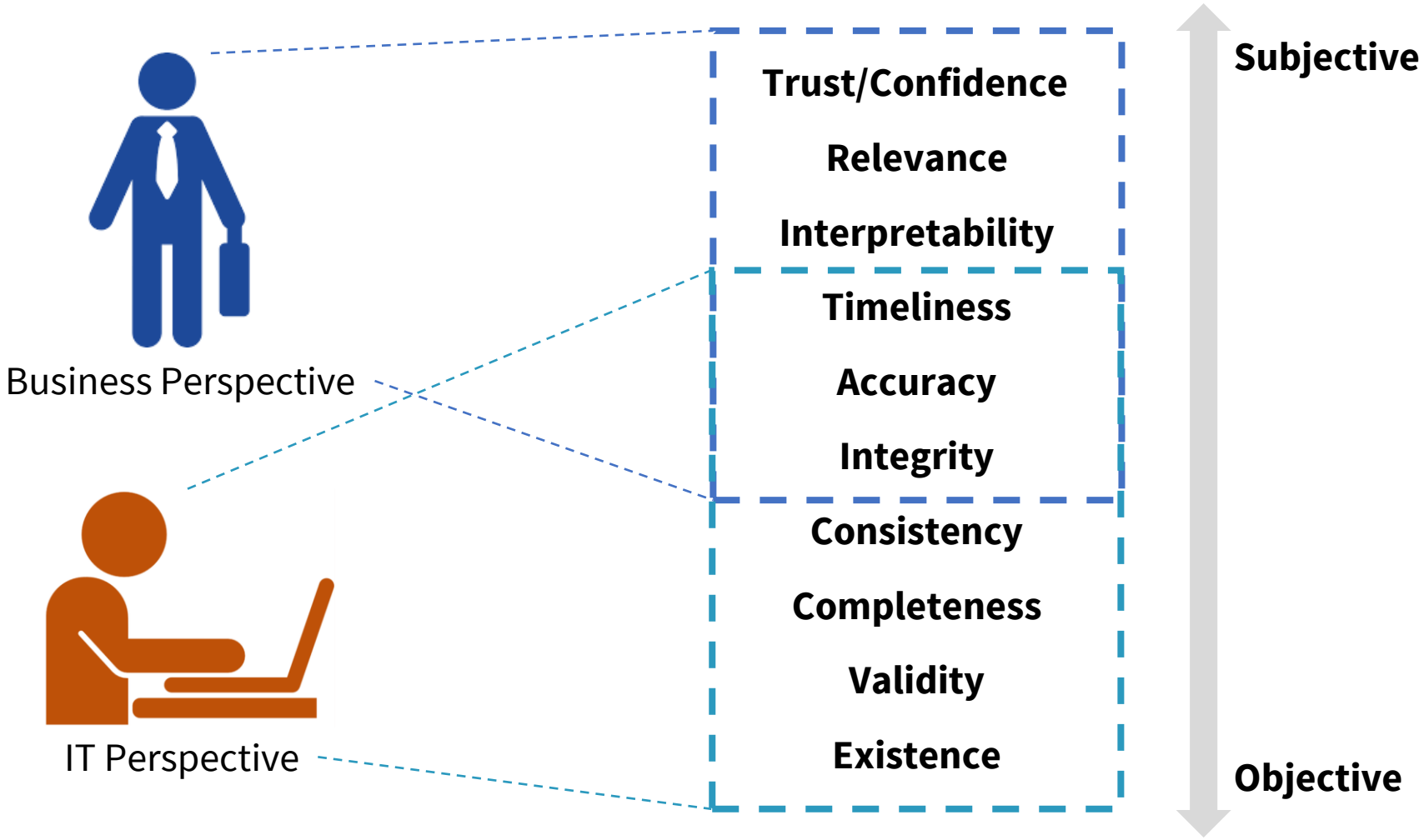


Data Types

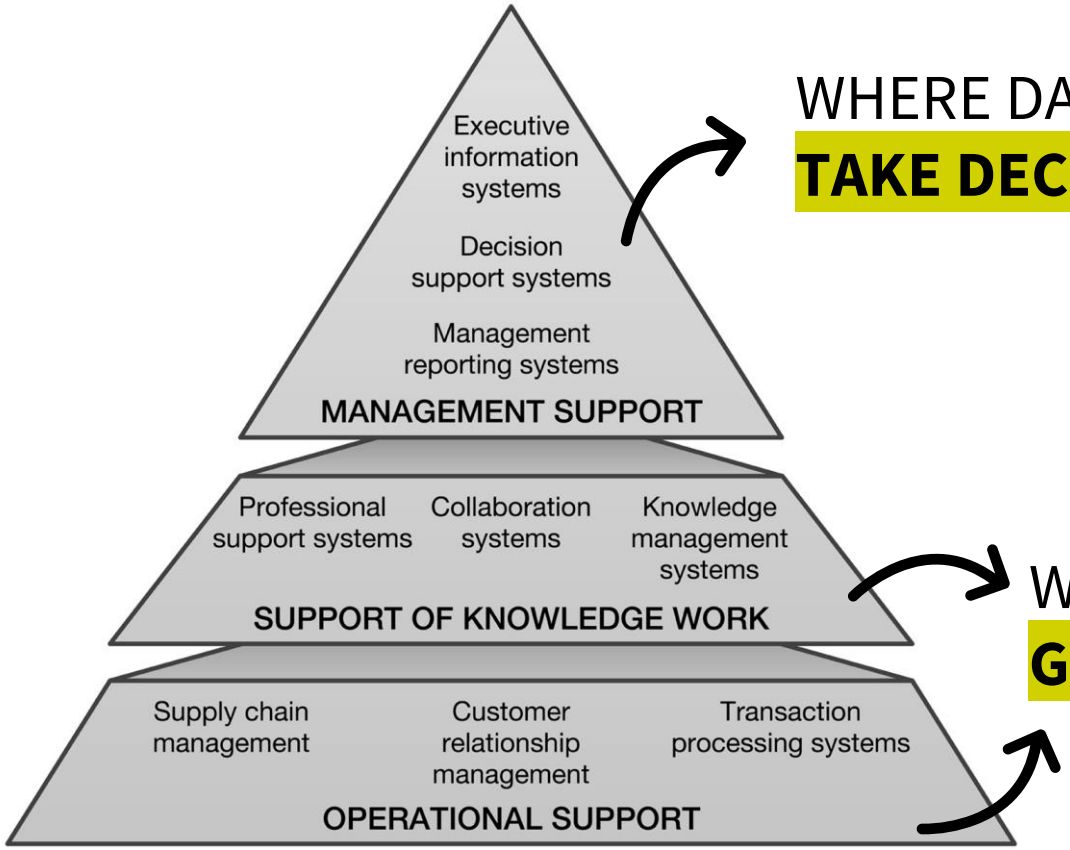
Data Governace Structure



Different points of view on the same data



Data flow



WHERE DATA ARE USED TO **TAKE DECISIONS**

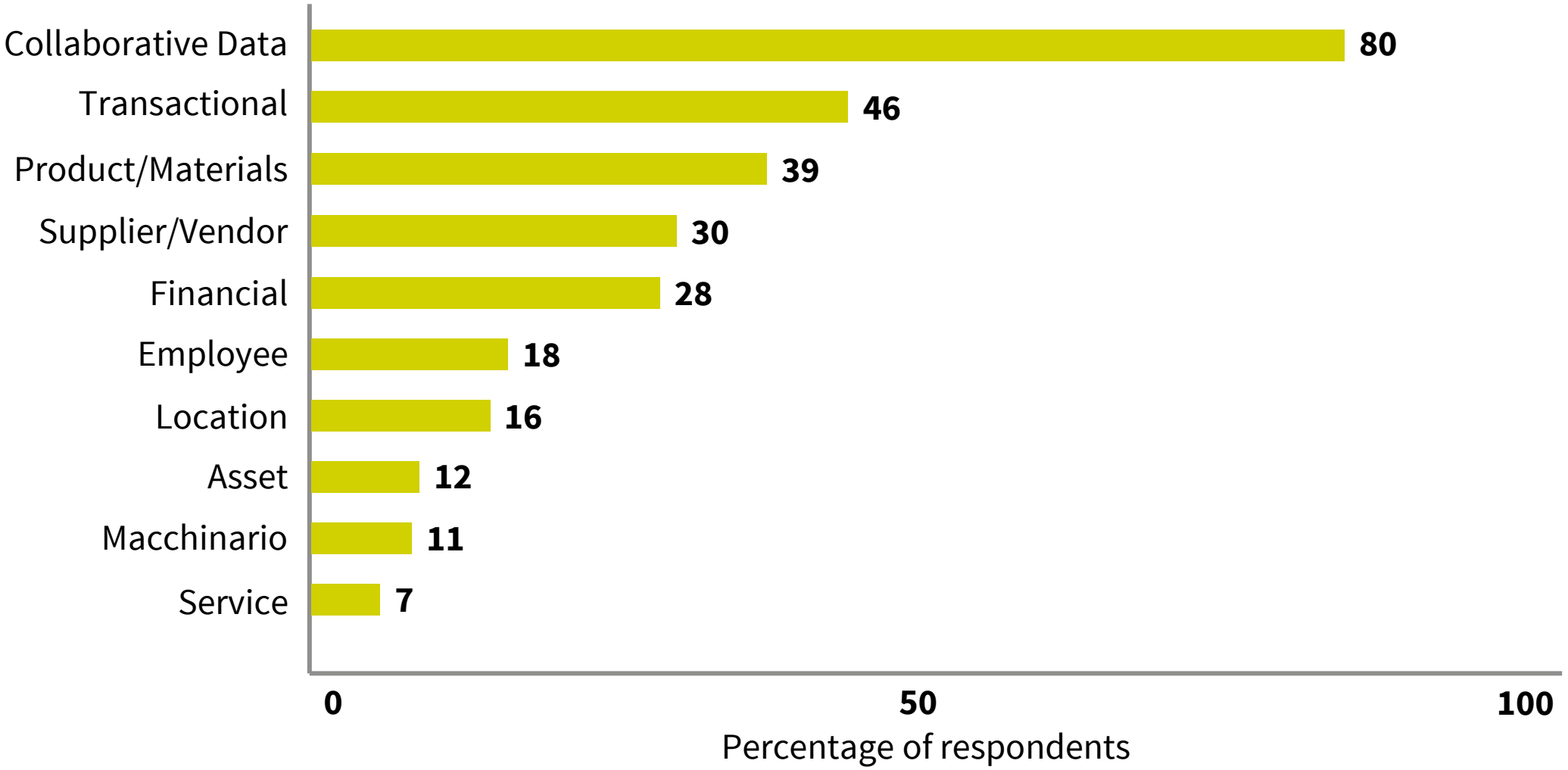
WHERE DATA ARE **GENERATED**

Data flow from Operational IS to Management IS, is crucial **guarantee quality through the entire Data Lifecycle** because **"Garbage-in Garbage out"**

On two occasions I have been asked (*by members of the British Parliament*), "Excuse me, Mr. Babbage, **if you put wrong digits in the machine, will they come up with the right answers?"** ... I am not actually able to understand the kind of confusion that could provoke such a question.

Charles Babbage, Passages from the Life of a Philosopher, 1864

What do most of your colleagues put their time and effort into on data quality?



Source: Gartner

Good quality data are the foundation of digital business

Compliance
Reduce the operational risk

Operational Excellence
Increase efficiencies and cut costs

Innovation
Accelerate growth

Insights
Take the right decisions

Fonte: Gartner

Agenda

Managing data and Data strategy adoption

Elements of Data Governance: logical and physical models

Setup of a Data Framework: : an example

Logical Model and Physical Model

LOGICAL LEVEL

It’s the level where Business Entities are located on, which are the information objects represented in terms of concept, used within the business processes

The Logical level responds to the need to provide a unique description to the elements here contained, defining information assets uniquely for the whole organization

PHYSICAL LEVEL

It’s the level on which everything existing at the physical level inside a company is mapped, the Physical Entities themselves, which are based on the business concepts defined at the Logical Level

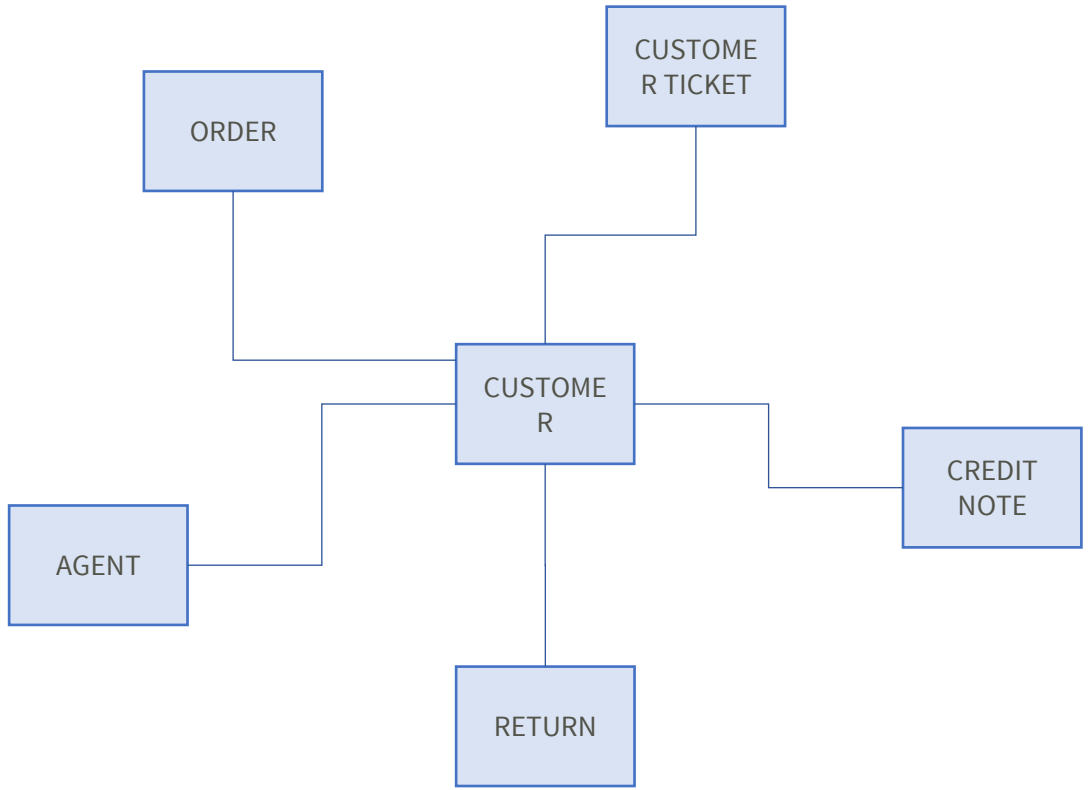
The Physical level translates the logical scheme in terms of tables and relationships, to constitute the real physical structure of the information assets, i.e. what is physically located within the company systems

Logical Model: Conceptual Data Model

The *Conceptual Data Model* (CDM) is the conceptual model for data providing a graphical representation of the various *Business Entities* (BEs) existing within an organization.

Different BEs can be grouped together, showing the links existing between them and relating one to each other.

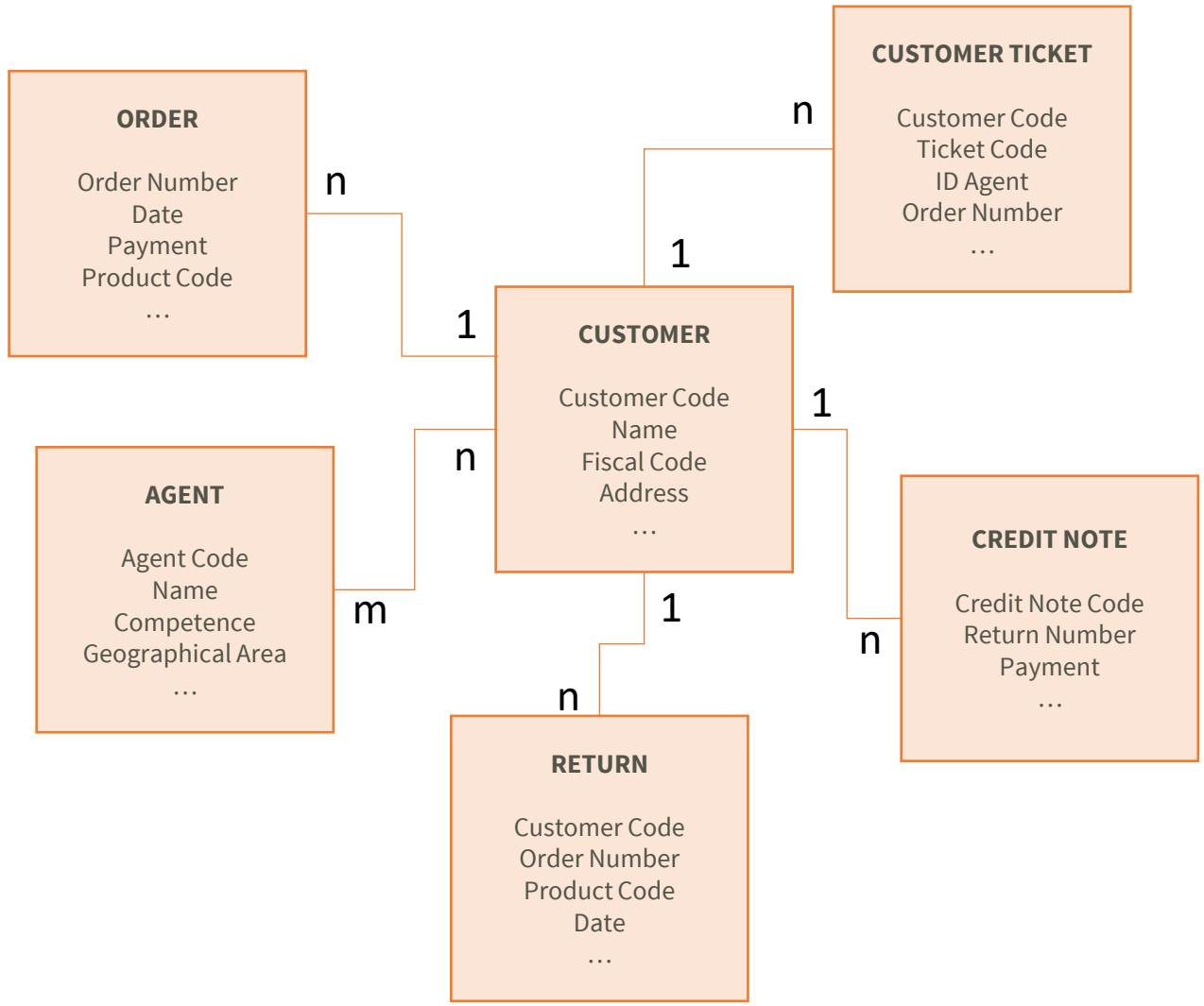
Each element within the CDM represents a business entity (eg. Customer, Order, ...), or rather a logical information element proper of that business context. For example, the EB Customer in the B2C field will represent the consumers to whom the company sells its products / services.



Physical Model: Entity-Relationship Model

The *Entity-Relationship Model* (or E-R Model) is a theoretical model for the conceptual and graphic representation of data, equivalent to the Conceptual Data Model for the Logical part.

Unlike the previous one, the E-R Model shows the relationships between the physical objects present in the system, thus tables, providing an indication of which ones are related and what type of relationship persists between them.

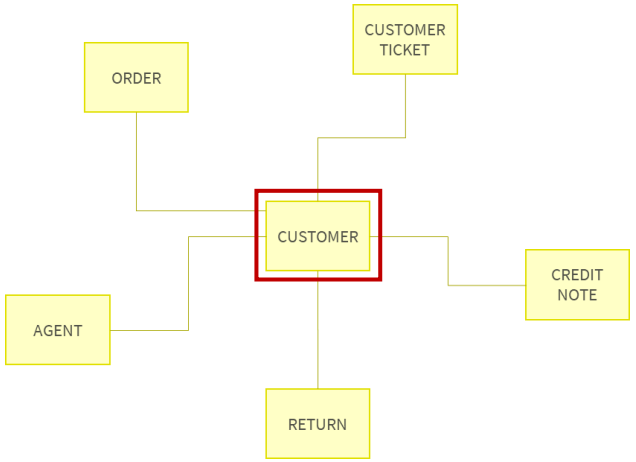


Two Levels - Business Entity and Physical Entity

LOGICAL LEVEL

The **Business Entity** is an information object represented in conceptual terms, or rather a package of information

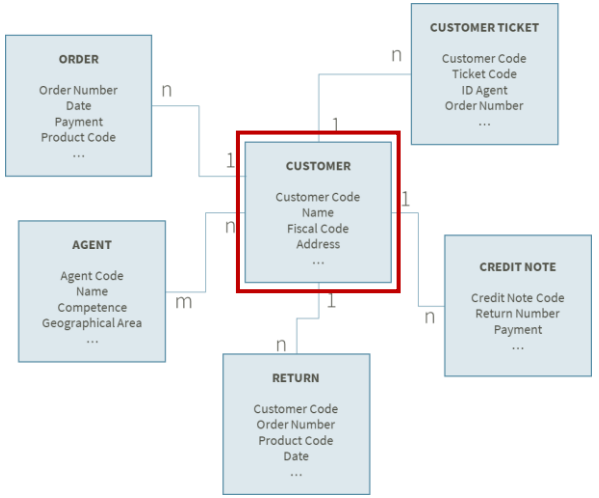
BE CUSTOMER



PHYSICAL LEVEL

Physical Entities represent the digital information objects that reside within the databases of the different systems used by the organization

CUSTOMER PHYSICAL ENTITY



Two Levels - Business Entity and Physical Entity

LOGICAL LEVEL

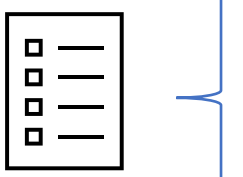
BEs can be defined and identified regardless of whether they are digitally managed on an organization's system / application

BE CUSTOMER

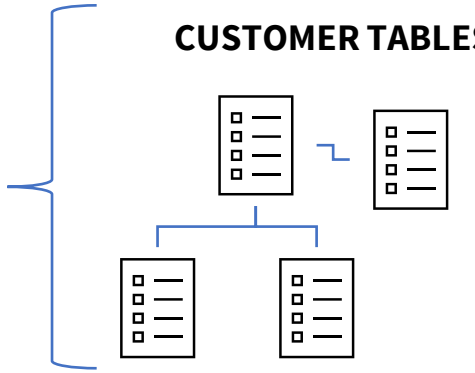
PHYSICAL LEVEL

Each Physical Entity are made up of aggregates of information coming from one or more source tables

CUSTOMER PHYSICAL ENTITY



CUSTOMER TABLES



Data Attributes and Physical Fields

LOGICAL LEVEL

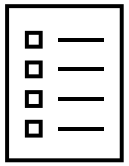
A Business Entity can be associated with one or more **Data Attributes** that allow the characterization of that specific business concept (e.g. the B2C customer could have as attributes Name, Surname, Address, ...)

BE CUSTOMER

PHYSICAL LEVEL

At the physical level, for each Data Attribute is defined a **Physical Field** within the table(s) which constitute the physical entity

CUSTOMER PHYSICAL ENTITY



Methodological Tools

LOGICAL LEVEL

Business Glossary:

tool used to give each business term a unique meaning and trace the elements contained on this level

BE CUSTOMER

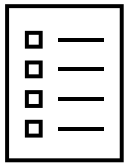
It is useful for creating and sharing definitions and conventions used in terms of the meaning of the data, at the semantic level

PHYSICAL LEVEL

Data Catalog:

methodological tool useful for have indication about data, how the data is made, how it is aggregated, and where it resides. What is mapped in the Data Catalog finds a counterpart in the Business Glossary

CUSTOMER PHYSICAL ENTITY



It is useful for defining and sharing the meaning of the Data, in terms of syntax

Metadata

LOGICAL LEVEL

Metadata can be added to the Data Attributes, being useful for improving the understanding and usability of what resides in the system

BE CUSTOMER

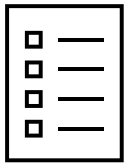
At the Logical level, Metadata can be: :

- **Description:** offers a textual description to define the semantics of the data
- **Taxonomy**: it offers an indication of the characteristic of the field, about how data should be indicated
- **Validity/quality rules:** indicates the validity of the data, for example indicating a range of validity
- **Usage rules:** they define how that data is to be used

PHYSICAL LEVEL

Metadata are information that describe various aspects of an information asset, and have value both for the Logical and for the Physical level

CUSTOMER PHYSICAL ENTITY

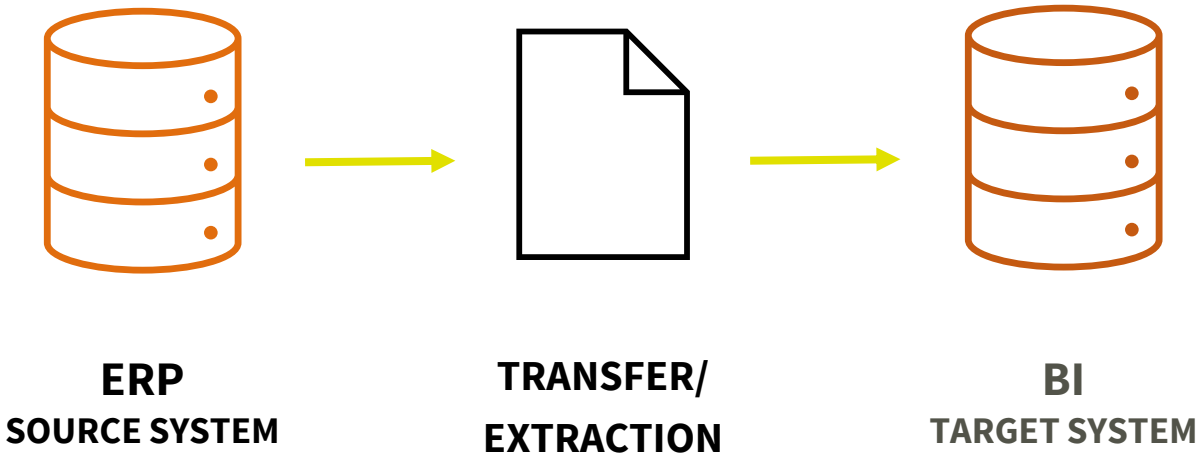


At the Physical level, an example of Metadata is provided by:

- **App**: indicates which application is associated with the Information Object in question
- **Tables**: provide information on which tables compete to define the Physical Entity in question
- **Source Database**: provides information on the origin of Data, in terms of source db

Data Lineage

The Data Lineage indicates the path that the data follows when transferred from one system to another, for example during the analysis operation of the data within the source system.



Agenda

Managing data and Data strategy adoption

Elements of Data Governance: logical and physical models

Setup of a Data Framework: an example

Data strategy: prioritizing the use cases

IDENTIFICATION OF "MISSION CRITICAL" USE CASES

Evaluated use cases with respect to:

- ✓ Shareholder value
- ✓ Business value
- ✓ Information value
- ✓ Data Asset
- ✓ Analytical potential

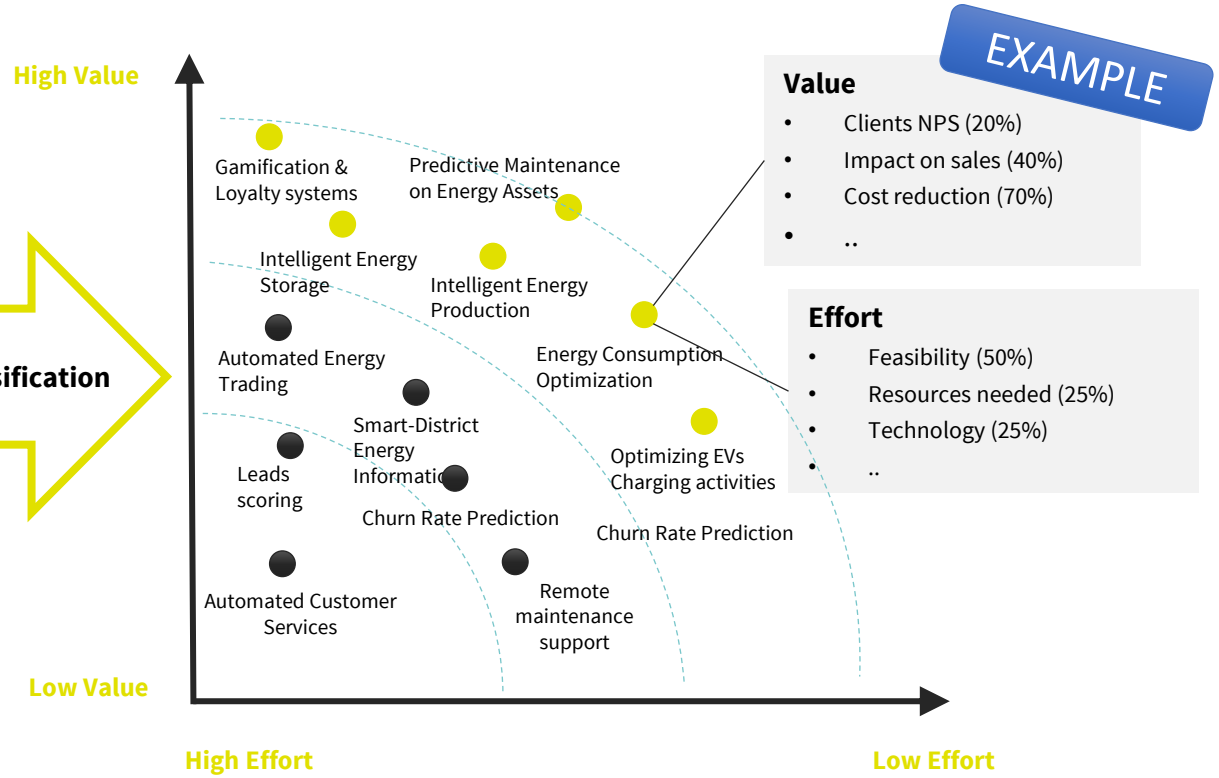
IMPACT ON VALUE FOR THE CUSTOMER

IMPACT ON BUSINESS GROWTH

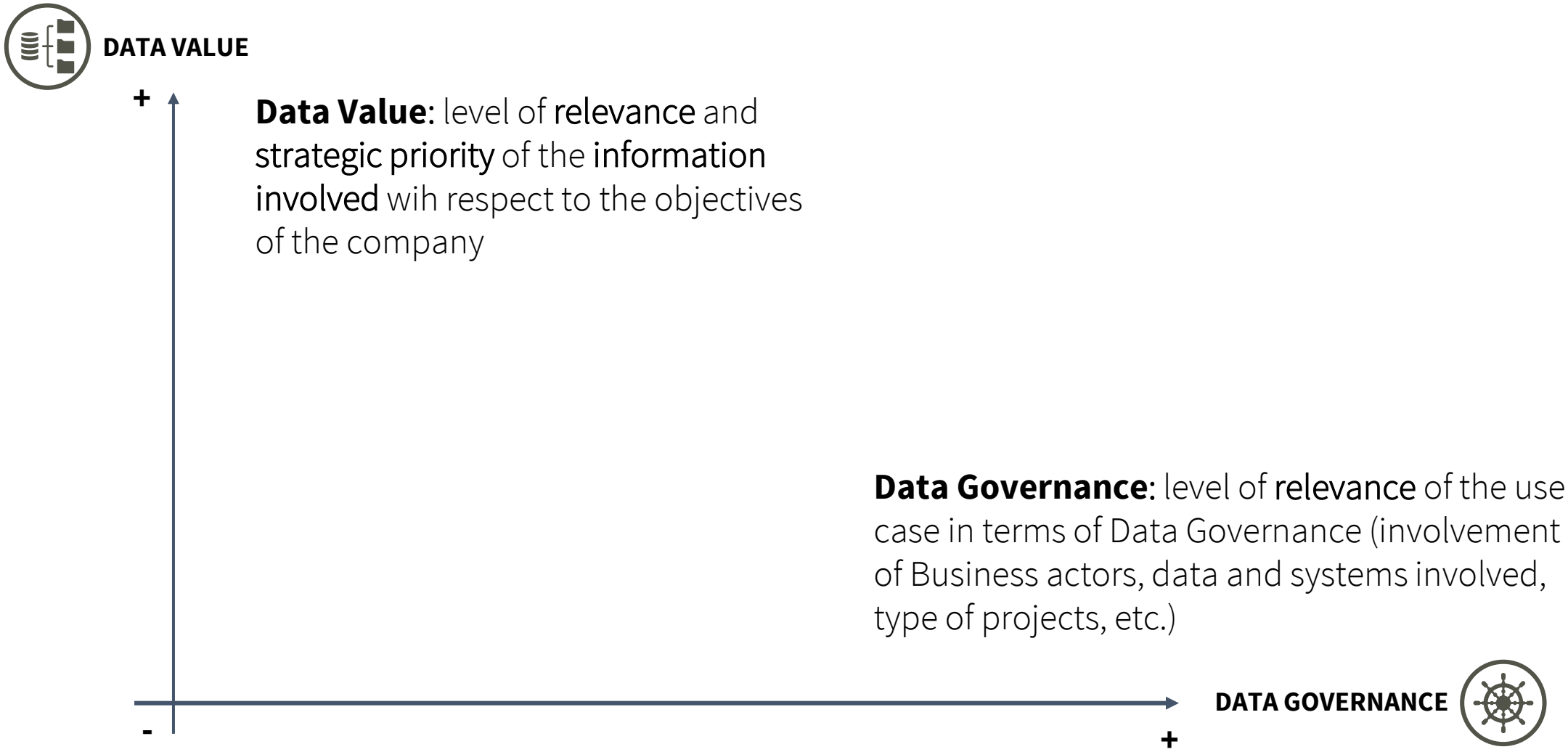
IMPACT ON COST EFFICIENCY

Classification

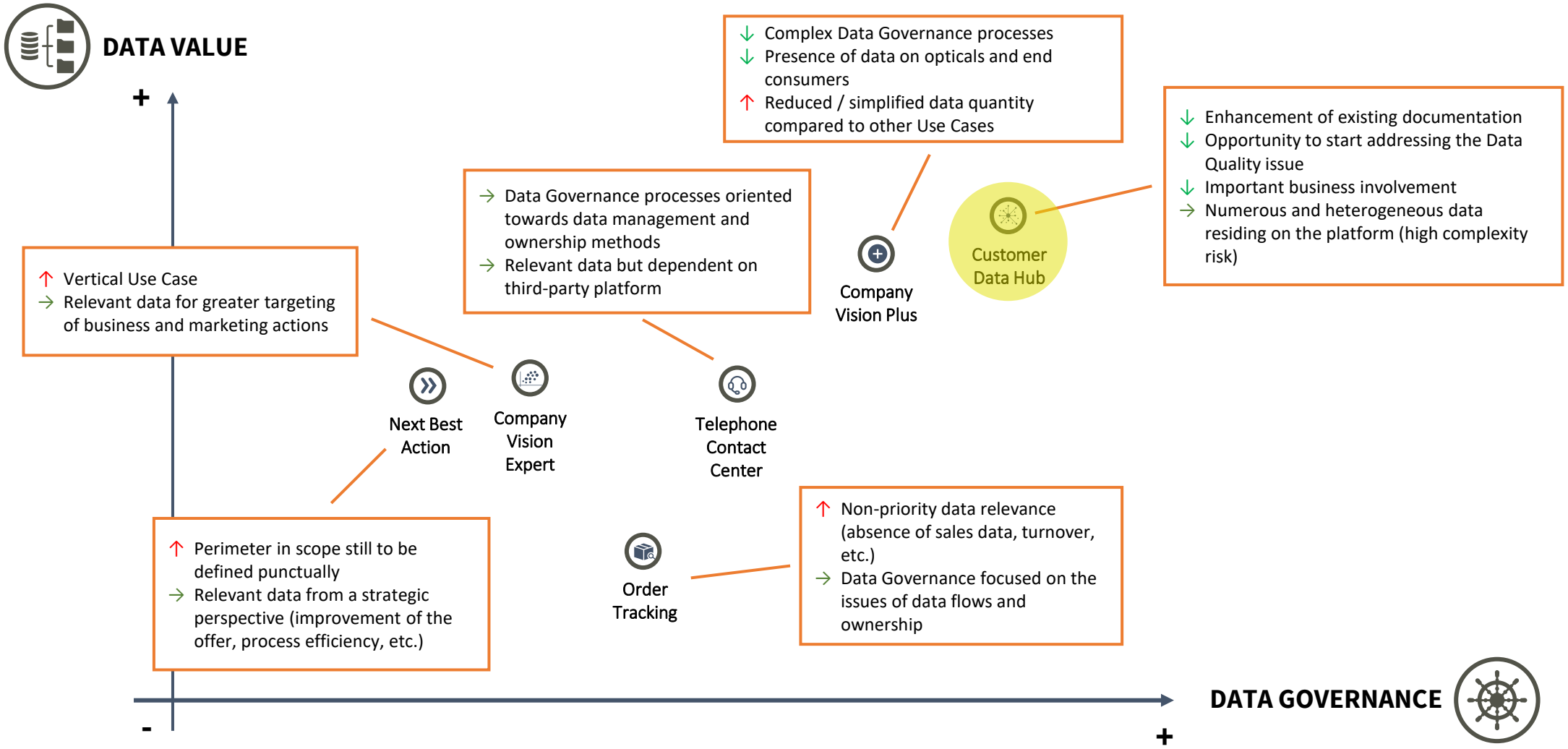
CLASSIFICATION OF IDENTIFIED USE CASES



Data strategy: prioritizing the use cases for a data governance project

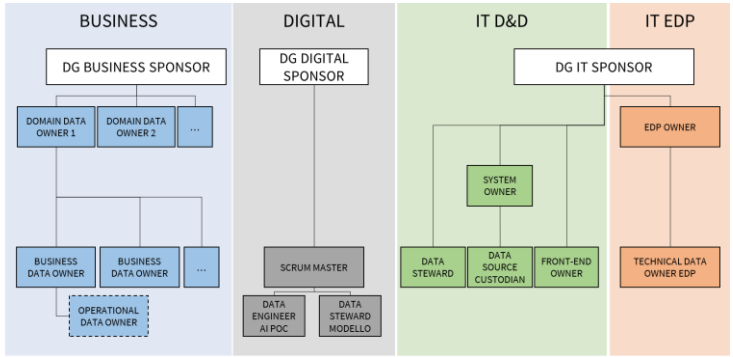


Synthetic evaluation of the initiatives

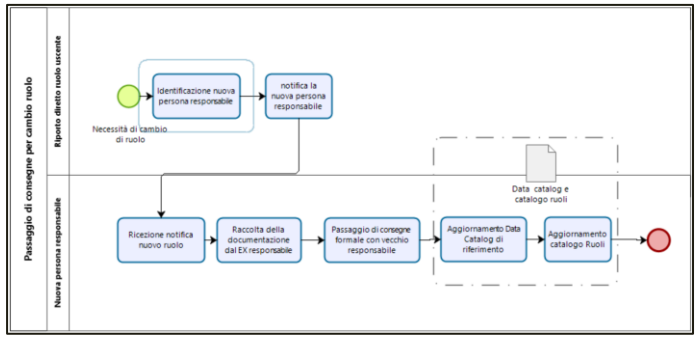


Characterizing elements of the Data Governance framework

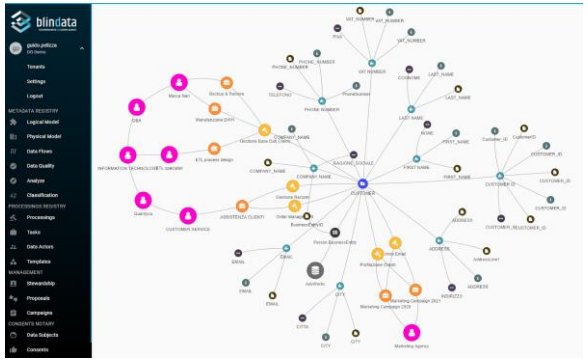
Roles and responsibilities



Phase governance process



Supporting tools



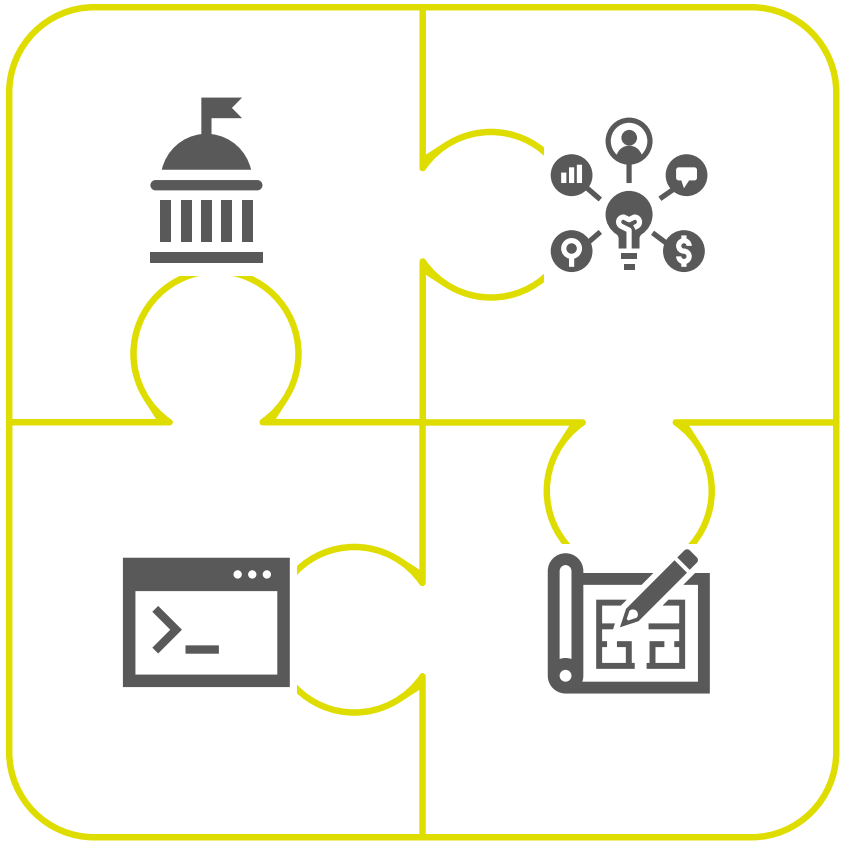
Data governance: responsibility domains

GOVERNMENT

- Role Assignment
- Check adoption of data governance processes Instruments

BUSINESS

- Semantic responsibility and guarantee of alignment with business processes
- Business documentation supervision
- DQ evaluation



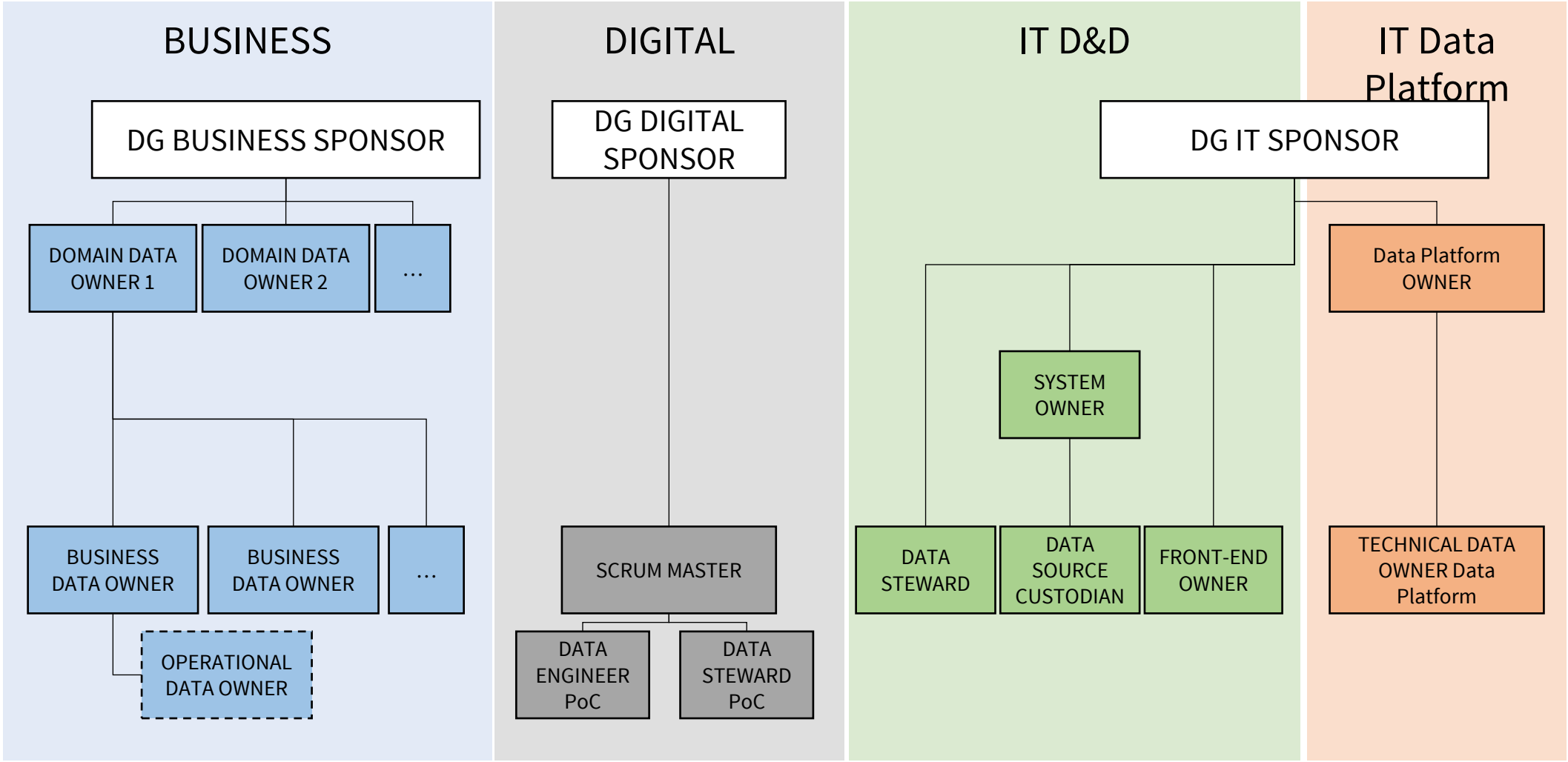
SYSTEMS

- Data Platform technical supervision, source systems and external data sources
- Technical documentation supervision

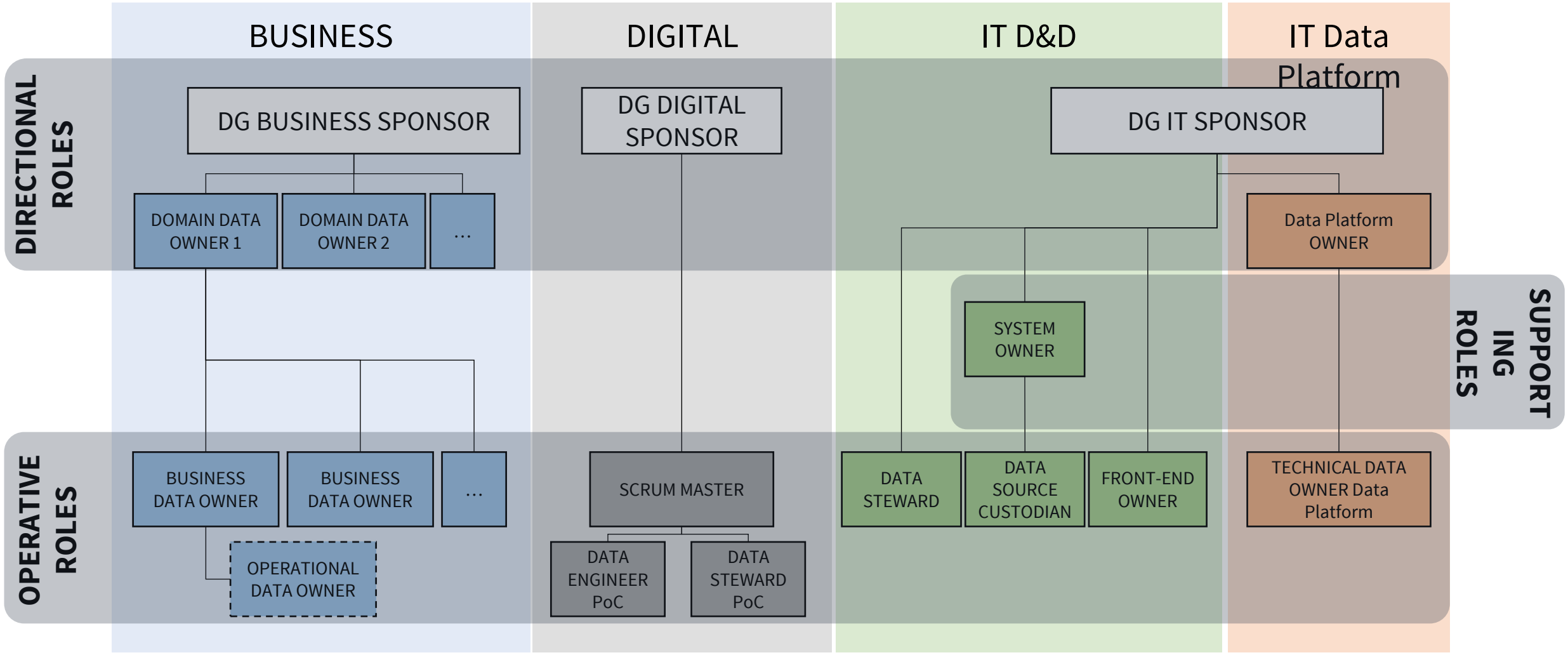
PROJECTS

Activation of DG business and technical roles within the project contexts

Data governance in an organization: example



Data governance in an organization: example



Data governance in an organization: example