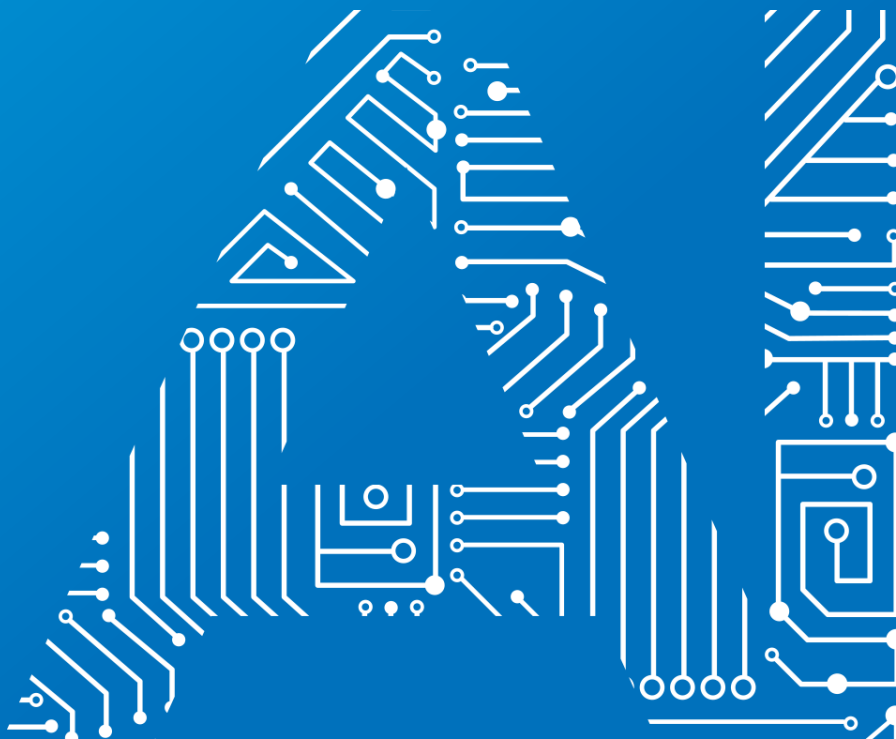
Strategy, Governance and Data Management Guide

BASQUE ARTIFICIAL INTELLIGENCE CENTER



Contents

1. Introduction, context and purpose of this document 5
2. Data at the heart of digital culture1 8
Value of data 8
Digital culture 9
3. Operational versus informational data 13
Data processing for analytical purposes 16
Data flow 17
4. Keys to understanding data strategy, governance and management 20
Reference Initiatives 20
Data strategy 23
Data governance and management 24
5. The data office 29
6. Key profiles of the data ecosystem 32
7. Enabling technologies 35
Data Catalogue solutions or tools 37
Solutions or tools to Centralise Data – Data Hub 37
Data exploitation solutions or tools – Data Fabric 38

Illustrations

Illustration 1: Use cases of governed data 6
Illustration 2: Data as the new oil 8
Illustration 3: Expected growth of data in different areas – European Commission - 2022 .. 9
Illustration 4: Pillars of digital culture - World Economic Forum - 2021 10
Illustration 5: Evolution of industrial revolutions - World Bank - 2018 11
Illustration 6: Traditional data ecosystem 13
Illustration 7: New data ecosystem 14
Illustration 8: Vision of the data ecosystem 15
Illustration 9: Data flow and environments in the organisation 15
Illustration 10: Architecture and data conversion process within the organisation 17

Illustration 11: Different frameworks and standards for data governance and management 21

Illustration 12: Relationship between ISO standards and UNE specifications on data – Spanish Ministry for Digital Transformation and the Public Function 22

Illustration 13: UNE 0080 - Data Governance, Management and Quality Management Assessment Guide - Maturity levels of organisations 22

Illustration 14: Benefits of having a Data Strategy 23

Illustration 15: Data Silos to Data Governance through the implementation of the Data Strategy 25

Illustration 16: Data Management Areas - DAMA 26

Illustration 17: Data Governance Office 29

Illustration 18: New data ecosystem 35

Illustration 19: Technologies for data integration and governance 37

1. Introduction, context and purpose of this document

1. Introduction, context and purpose of this document

The Basque Artificial Intelligence Center (BAIC) is an Association that is set up as a centre for leadership in the development of Artificial Intelligence (AI) in the Basque Country with the aim of **improving business competitiveness and the welfare of Basque society**. BAIC fulfils its mission through the development of different areas of activity, which, when combined, make it possible to build complete solutions to complex AI challenges. These strategic axes include several areas of activity such as:

- **Observatory**: To strengthen and develop the AI ecosystem in the Basque Country with a focus on vigilance, knowledge, connection and international positioning under a framework of development and ethical application of AI through adherence to the laws and regulations.
- **Talent and capabilities**: To enhance the development of AI talent in the Basque Country, as well as to support the cultural change implied by the data economy and the implementation of AI tools, strengthening infrastructures and enabling technological assets.
- **Data strategy**: To position the Basque Country as a benchmark in the implementation of advanced Data Strategy, driving initiatives for the development of data sovereignty and the data economy and promoting the development of federated and interoperable data spaces.
- **Applied AI**: To promote an applied AI ecosystem in the Basque Country, increasing specialisation and technological maturity, being leaders in practical and collaborative application, as well as working on the acceleration of challenges through open innovation tools.
- **Positioning**: To highlight the Basque Country's skills in AI, stimulate collaboration and project the AI ecosystem on the international scene. To provide valuable communication between the agents of the AI ecosystem.

The 2025 Basque Country Digital Transformation Strategy (ETDE as abbreviated in Spanish) identifies AI as a technological lever and is, therefore, a clear area of opportunity for the Basque Country's business fabric, as it is a tool capable of driving innovation and transforming the way in which organisations operate.

Although AI has already started to spread in the Basque business fabric, its implementation on a larger scale among our companies and mainly SMEs, which make up a large part of the Basque business fabric, requires a collective effort.

Given this scenario of clear opportunity, **one of the main challenges facing BAIC and the group of associated entities is to guide the different agents in the ecosystem towards a strategic and responsible implementation of AI, and to achieve this it is essential to have a data strategy that guarantees an appropriate data governance and management. Data governance, through policies, procedures and standards, ensures that the data is quality,**

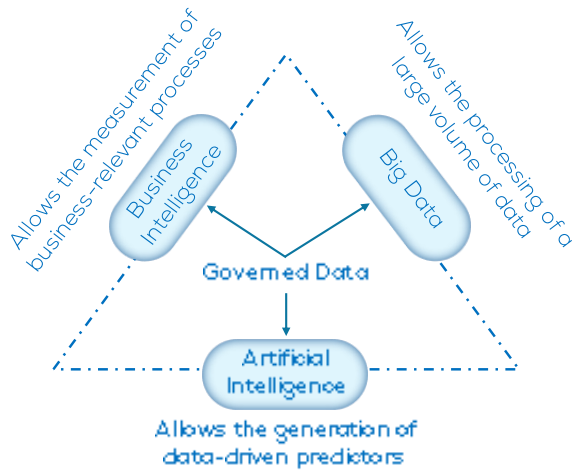
reliable, relevant and accessible, which is crucial for developing and maintaining effective AI models.

Governed data translates into data that has been "validated" by the organisation, has a clearly identified meaning and is understandable to all business units. The availability of this data will allow new use cases to be developed quickly, securely and efficiently, such as:

- Big Data: Ability to process and analyse large volumes of data, both structured and unstructured, at high speed and in real time or in batches.
- Business Intelligence (BI): Transformation of data into relevant and useful information for making business decisions.
- Artificial Intelligence (AI): Application of algorithms and machine learning techniques to solve specific problems from the data available.

In summary, data strategy and governance will result in reliable and exploitable data, both processed and governed, which will maximise the impact and outcomes of these three blocks of Use Cases.

Illustration 1: Use cases of governed data



For all these reasons, one of BAIC's strategic axes is the "Data Strategy" with the aim of "positioning the Basque Country as a benchmark in the implementation of Data Strategy, driving initiatives for the development of data sovereignty and the data economy and promoting the development of federated and interoperable data spaces".

In this context, from all the entities comprising BAIC, this manual is presented in order to inform, inspire and guide the ecosystem and the industrial fabric in general on their path towards the adoption of a data strategy.

2. Data at the heart of digital culture

2. Data at the heart of digital culture

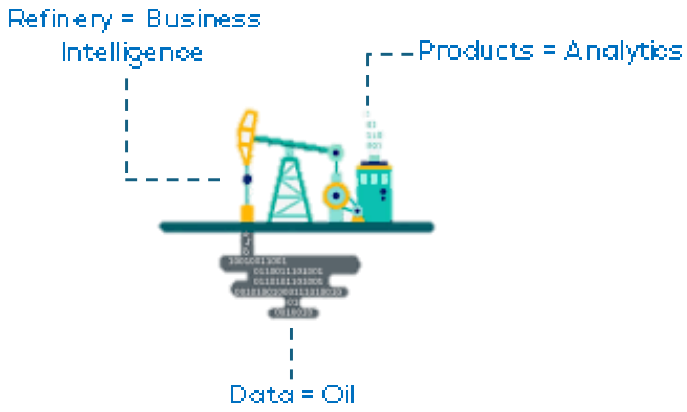
Value of data

Data, according to the definitions of the Royal Spanish Academy (RAE [Real Academia Española], consists of "appropriately codified information" and "that serves as a basis for obtaining exact knowledge of something", which means that it is a fundamental asset for all organisations, as data, when processed and analysed, is transformed into valuable knowledge for informed decision-making.

Its importance as an organisational asset lies in its ability to drive operational efficiency, improve decision-making and generate competitive advantages.

With accurate and up-to-date data, companies can identify trends, anticipate market changes and formulate effective strategies. Moreover, data is the fuel that powers the exponential advancement of artificial intelligence (AI) and its enabling technologies, such as machine learning and natural language processing. The ability of AI to learn and evolve is based, among other things, on access to large volumes of data, which allows for the development of innovative solutions.

Illustration 2: Data as the new oil



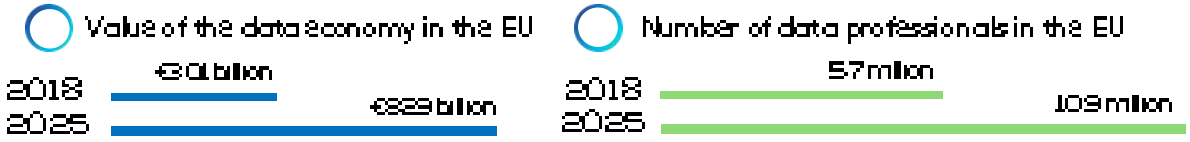
If we take as a reference the companies that have experienced the greatest growth in recent years, it has been those that have been able to be disruptive, have made decisions based on data (**data-driven**) and have detected the new needs that society was demanding.

Also in the Basque Country, important advances are being made in the adoption of data-driven strategies, with outstanding examples of public and private organisations that are transforming their management and decision-making thanks to the efficient use of data, such as, for example:

- The Basque Government's own "Data governance strategy for the public sector in the Basque Autonomous Community[1]", Ardatz 2030.
- Public entities like EUSTAT are making progress in data governance and management thanks to their "efficient and digitised e-Eustat organisational model". This is also the case with EITB, which is implementing its own quality, governance and data analytics system.
- At an academic level, both in formal and non-formal education, different training institutions offer training products in areas of knowledge related to data analysis, architecture, governance, engineering, security and visualisation.
- At the same time, numerous private organisations like CAF, EITB, Eroski, ITP Aero, Tubacex, etc., are evolving their data ecosystems and strategies to optimise decision-making, improve their competitiveness and personalise their products and services, among others.
- The impact of data on R&D in the Basque Country is also considerable, as it is revolutionising the way in which research is conducted, new technological solutions are created and key sectors are strengthened. For almost all BRTA (Basque Research & Technology Alliance) agents, data has become an essential strategic resource to drive innovation in areas such as advanced manufacturing, energy, biosciences, mobility and information technologies.

Organisations are increasingly aware of the value of data, which is reflected both in companies' own valuation systems and in the exponential growth of professionals dedicated to safeguarding and exploiting data.

Illustration 3: Expected growth of data in different areas – European Commission - 2022



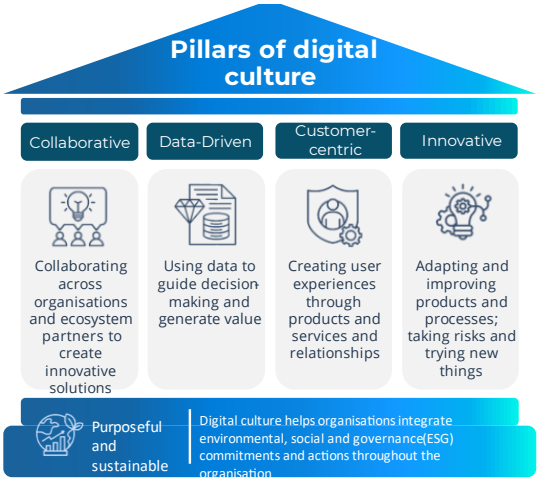
Digital culture

One of the main challenges that organisations have faced in recent times is adapting to the possibilities offered by digitisation to gain a competitive advantage. To achieve this, it is essential to adopt a digital culture within the organisation.

1 Data governance strategy for the public sector in the Basque Autonomous Community – Ardatz 2030 – Department of Governance, Digital Administration and Self-Government – Basque Government

One of the pillars of digital culture is precisely to transform organisations to become 'Data-driven' so that they are able to make decisions based on information derived from their data, rather than on opinions or intuitions, and for this to be possible, it is necessary to have relevant and quality data.

Illustration 4: Pillars of digital culture - World Economic Forum - 2021



When we talk about digital culture, it is essential to understand that it goes far beyond providing technology that helps to streamline processes, as the important thing is to achieve a change of mentality in the organisation **in order to bring about a digital transformation both internally (optimising operational processes) and externally (detecting new needs and developing new service and business models).**

In this change of mentality, data plays a fundamental role and, as such, must be taken care of. For technical (IT) teams, data has always been a valuable asset to manage, but it is the business profiles that must lead the exploitation of data so that they can serve as a guide for organisations in decision-making and in providing competitive advantages.

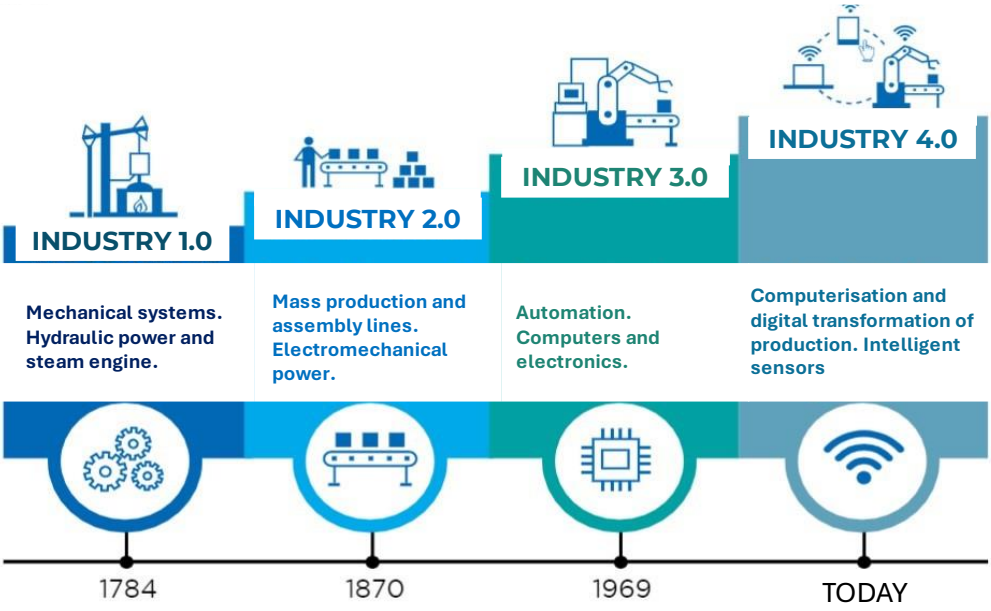
Therefore, every organisation must devote efforts to having a data ecosystem capable of meeting the needs of the organisation, understanding the **data ecosystem as the set of technologies and processes used by people within an organisation to obtain and exploit its data for operational purposes (for day-to-day operations) and informational purposes (for decision-making).**

Therefore, when we talk about a data ecosystem, we must take into account the 3 pillars on which it is based:

- **People**: It is essential to raise awareness among the staff of organisations of the need to properly manage data and for this it is necessary to have staff with the necessary technical knowledge and training, as well as to set up an organisational structure that governs data management.
- **Processes**: To govern and manage data properly, a data strategy must be in place to guide the organisation's actions around data and to establish policies and procedures that allow for a standardised and governed data management.
- **Technologies**: Once the organisation is aware of the relevance of data and once the procedures for data management have been established, it is time to rely on technology to carry out this data governance and management in a more efficient way.

If we look at the great revolutions that organisations have undergone throughout history, we will also see the relevance that these 3 pillars have had in all of them:

Illustration 5: Evolution of industrial revolutions - World Bank - 2018



On some occasions, it is technology that has driven these historical revolutions, such as the appearance of steam machinery or automation, but on other occasions it has been the specialisation of people and improvements in production processes, such as what happened with chain production and specialisation, where, thanks to the timekeeping (data-driven decisions), organisations realised that the more a certain operator repeated an operation, the more efficiently they performed their work.

We are currently in the midst of the big data revolution, with multiple characteristics that give us all the capabilities we need to extract the maximum value from our data.

On the one hand, thanks to the enormous amount of technological devices, sensors (IOT), machinery and applications that constantly collect and process data, we are experiencing a data boom. We are living in the era of Big Data, where all organisations have a huge amount of data, either because they are captured by themselves or because they obtain them from external sources (open data, data spaces, data economy, etc.). Data is therefore available to all organisations, and those organisations that manage to make a proper use of this data will obtain a great competitive advantage over the rest.

On the other hand, from a technological point of view, there are currently multiple solutions for data processing: Open Source solutions, proprietary solutions, on-premise solutions, cloud solutions, etc., and at a computational level, organisations do not need to continually expand their DPCs (Data Processing Centres) with vertical solutions as they did until now, but can have "resources on demand", thanks to the horizontal solutions offered, for example, by the large Cloud solutions, whereby organisations can contract technological resources on a variable basis as they need them.

Another aspect to consider is the regulations in force, which are responsible for protecting the confidentiality of information, such as the GDPR (General Data Protection Regulation). Therefore, those with an adequate data ecosystem will be best prepared for all present and future opportunities, for example, AI.

3. Operational versus informational data

3. Operational versus informational data

Nowadays, all organisations have been creating a data ecosystem to manage their daily operations (ERPs, CRM, DBs, etc.), so the usual ecosystem of organisations is as follows:

Illustration 6: Traditional data ecosystem

An ecosystem dominated by Relational Databases (which store data in tables structured into rows and columns), which are very consistent and solid when it comes to storing structured information. In this ecosystem, data custodians and exploiters are usually technical profiles, such as database administrators or BI analysts.

Therefore, when business profiles need access to data, they are dependent on these technical teams to prepare the data and provide them with the requested information, which also means that technicians have to dedicate their time to these tasks to the detriment of dedicating that time to optimising the performance of these operational systems.

Therefore, when we refer to **Operational data, we are referring to data that is used by organisations to carry out their day-to-day operations and which is designed for a very specific use and purpose. This data is dominated by technical staff and is managed through operational and functional tools such as a CRM to manage customers, an ERP to manage production processes, or an accounting system to manage the organisation's finances. In this environment, what is needed is consistency and performance, so systems must work quickly to meet the operational needs of the organisation.**

On the contrary, this management in the form of silos (an organisational structure in which the different departments or areas within an organisation operate in isolation) sometimes leads to the existence of duplicated information in different systems, which creates doubts, conflicts and quality problems, as well as difficulties in obtaining an overall view of the information.

It is therefore necessary to broaden the perspective of the legacy (the operational systems already in place in the organisation) and be able to obtain a 360º view of the data. In other words, the current (legacy) systems must continue to function and maintain their focus on the

efficiency and performance of the systems, and it is important to protect these systems from being overloaded by new uses, such as using these environments for new purposes such as analytical, data science, BI, etc.

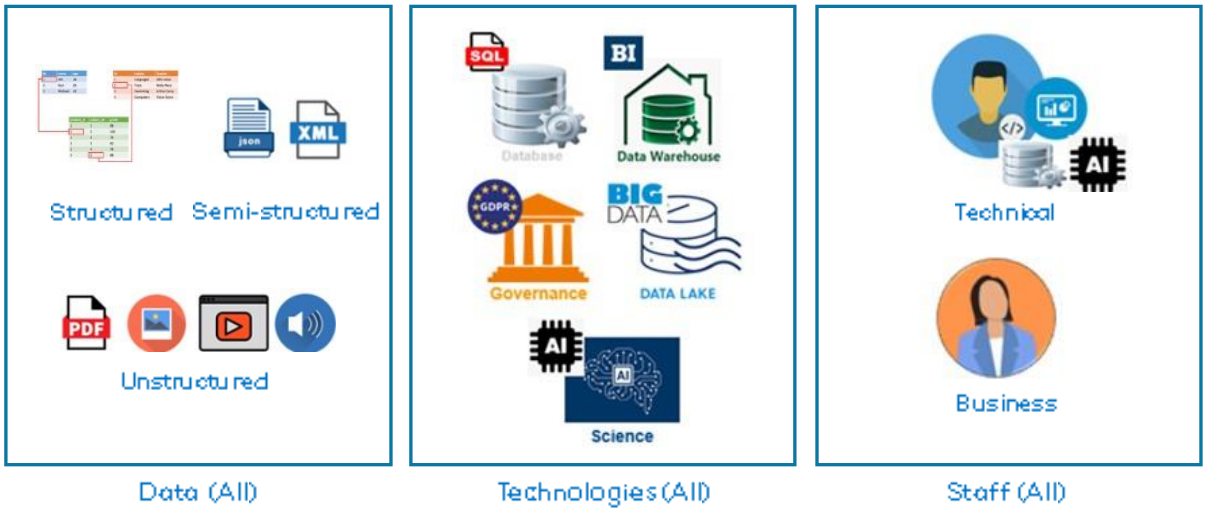
The important thing is that the organisation is able to have a 360º view of the data, so that the data is accessible, understandable, usable and of high quality, independent of the source system in which the data was generated.

Business profiles are the ones who best understand the needs of organisations, their context and who must make the decisions that help them to evolve. In order to move towards a Data-Driven organisation, it is essential that business profiles have access to the information they need, but in a format that is understandable to them and abstracting from the technical characteristics, format or systems in which this information is originally found.

To this end, it is essential to build this informational environment through the incorporation of modern systems that complement the organisation's existing legacy systems. **Informational data is data that must be the only source of truth, because it contains only quality data, data that has been "validated" by the organisation, that has a clearly identified and understandable meaning for everyone and therefore, it is the data that must be used to be exploited in the way the organisation needs (BI, projects, analytics, decision-making, etc.).**

Therefore, the new data ecosystem must have this informational approach, so that it contains the necessary data, regardless of its original structure (structured or unstructured data), so that it can be accessed by both technical and business staff and that it is accompanied by technologies that can be accessible and appropriate to the profile that will use it.

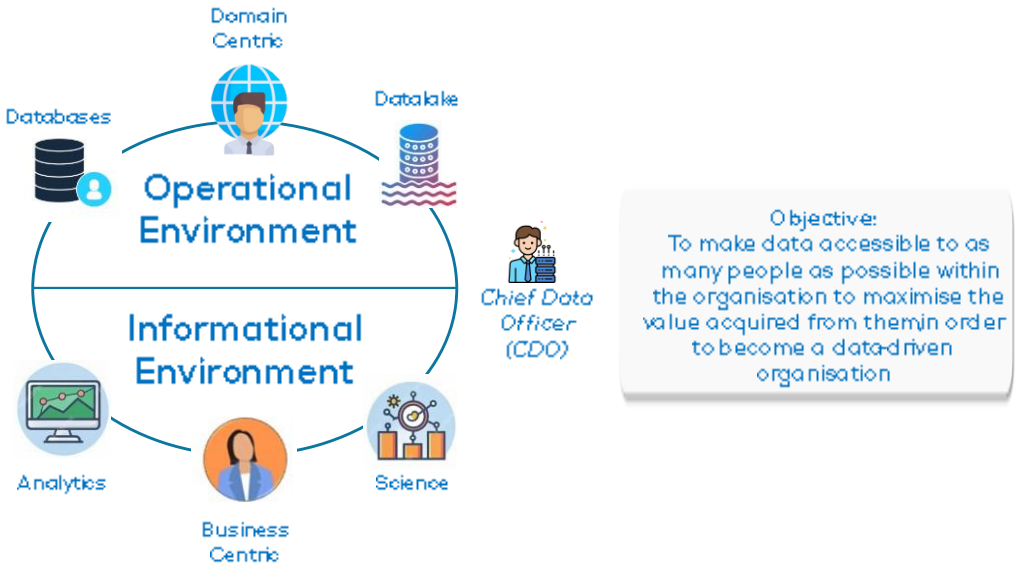
Illustration 7: New data ecosystem



In short, the challenge facing organisations is to meet the needs of these two environments. An operational environment, focused on transactional data, day-to-day operations and where the important thing is consistency and performance, and an informational environment, where the important thing is to have the source of truth, with quality data,

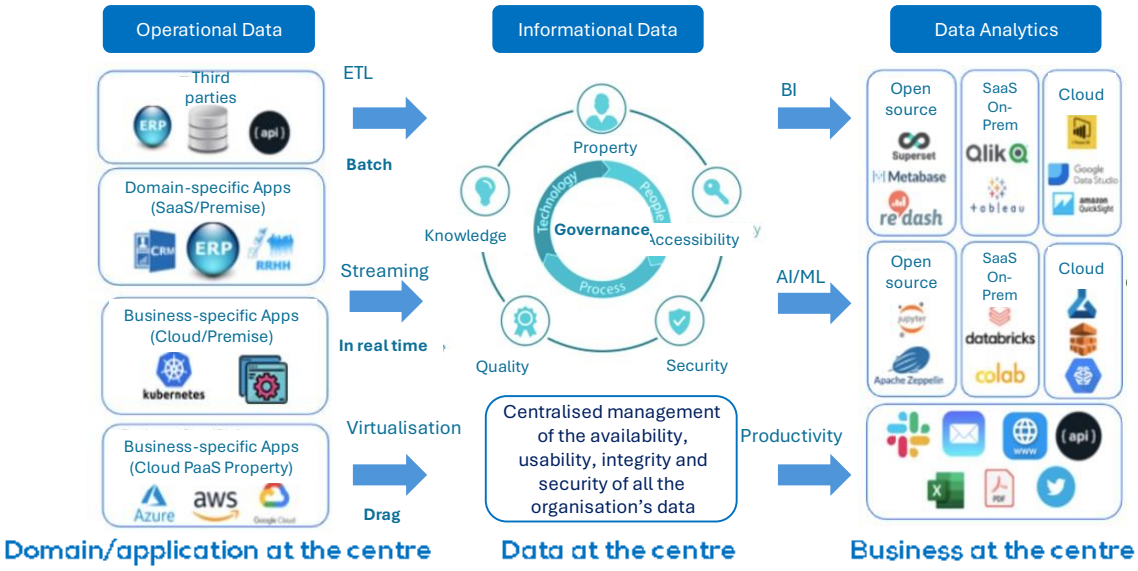
unequivocally identified, accessible and understandable for the entire organisation. Therefore, the need or desirability of one environment or the other will be determined by the intended use of the data.

Illustration 8: Vision of the data ecosystem



Being aware of these two realities, the important thing is to implement the necessary mechanisms so that, starting from the operational data, this informational data can be used by the organisation, respecting all internal and external regulations in terms of security, accessibility, quality, etc., so that it can be used for the purposes it is needed for.

Illustration 9: Data flow and environments in the organisation



In order to convert operational data into informational data, it is necessary to first analyse the characteristics and needs of the organisation with respect to the data. When it comes to "collecting" data from its operational origin, there are different approaches:

- ETL/ELT: This mainly consists of making **a copy of the source data to be further processed with an informational approach**. In this approach, you first extract the data from the source system, then transform it (cleaning, organising or adapting it according to what you need), and finally load it into an informational system. This process is not in real time; it is usually done in batches, i.e., periodically (daily, weekly, etc.). It is useful when you do not need the data instantly, but rather well organised and structured for later analysis.
- **Streaming: This consists of capturing operational data in real time for informational use**. It is as if the data were a continuous flow, and each time it is generated in the operational environment, you "catch" it to analyse it or use it immediately in the informational environment.
- **Virtualisation: This consists of directly allowing access to the source where the data is located.** You do not copy the data to another system; instead, you allow other systems to directly access the source where the data is located, as if you were giving them a window to look at the data from where they are. It is accessed on demand when needed.

Data processing for analytical purposes

As explained above, there is a problem when it comes to obtaining the maximum value from the data: Business wants to experiment and implement new analytical use cases, but the existing data in the organisations are in an operational environment designed to support day-to-day operations, and therefore the typology and formats of these data sources were not designed to be analysed en masse.

To do this, it is necessary to process the data and make it available in a properly governed information environment so that it can be used for the necessary purposes.

The need for Data Governance for the creation of intelligent systems is based on a simple idea:

- Garbage In – Garbage Out: - The central idea is that the quality of the results obtained from a system or process is directly dependent on the quality of the input data. If the input data is incorrect, inaccurate or of poor quality, the results will also be incorrect, inaccurate or useless. It is precisely here that governance plays a key role in understanding the fields and the quality of the data.

In short, in the world of Artificial Intelligence and Machine Learning, **data scientists spend more than 70% of their time cleaning, preparing and understanding the meaning of data**. While this is iterative and ongoing work, there are stages such as Data Exploration and Data Wrangling where understanding the data and validating its reliability are crucial. Some of the tasks directly related to understanding the data are as follows: Understanding Data, Outlier Detection, Normalisation and Standardisation, Data Type Conversion, Creation of New Variables, Establishment of Consistent Formats, Column Renaming, etc.

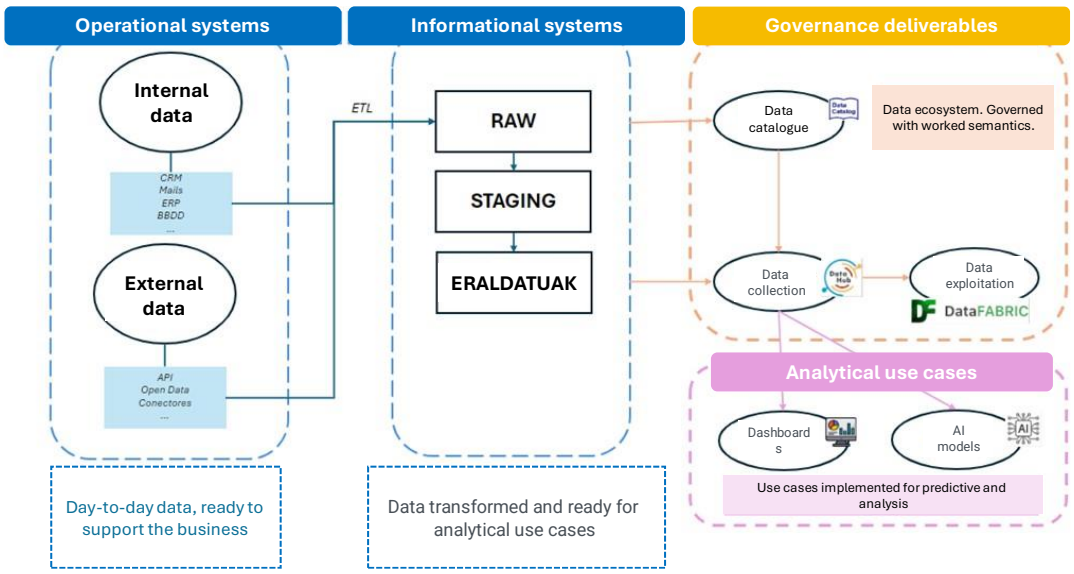
Having governed data speeds up analytics projects, because, for example, thanks to the data catalogue, the available data and their descriptions are known, and thanks to the centralisation platform, these data can be accessed with quality guarantees.

Data flow

In the next block, the aim is to **visualise the process of converting operational data (designed for day-to-day use) into informational data that is governed and made available for analytical use**. This process is vital for any organisation that wants to use its data to make better decisions. The main focus is on understanding that the design of the architecture must allow converting raw data from different sources into useful and governed data for analysis, reporting, forecasting and other applications.

Next, a high-level architecture is presented, with the aim of showing how, **starting from an operational environment made up of different origins (legacy), it is possible to achieve an information environment that provides the necessary data to deal with the different use cases.**

Illustration 10: Architecture and data conversion process within the organisation



Although these concepts will be explained in more detail in section 7. Enabling Technologies, it should be noted that Data Catalogue solutions are mainly used to visualise and discover data, providing information on where it is located, how it is structured and who is responsible for it, thanks to metadata management and search tools, whereas a Data Hub focuses on accessing and extracting data in a centralised way. It acts as an integration point where data from multiple sources is connected and distributed for real-time or on-demand use. Finally, Data Fabric solutions offer functionalities for data exploitation.

ETL (Extract, Transform, Load) or ELT (Extract, Load, Transform) processes, as we have already seen, are a set of procedures used to move data from multiple sources into a unified data warehousing system. This will produce the RAW, STAGING, if necessary, and TRANSFORMED layers. This process is **essential for moving information from the operational to the informational environment.**

This approach is suitable for both small and large projects. In simple projects, such as consuming data from an API, a basic script can be used without complex intermediate steps. In larger projects, where large volumes of data are handled, the ETL process is more comprehensive and requires intermediate stages, such as STAGING, to organise and clean up the data before final processing. For each layer, the following governance aspects can be worked on:

- The **RAW/BRONZE layer** (there are different denominations to explain these layers) contains the data as extracted from the original sources, without any transformation or cleaning. Here, the important thing will be to apply basic quality and access control rules to ensure that data is captured correctly and its integrity is maintained. On the other hand, more advanced data quality rules can also be implemented, such as de-duplication, correction of incorrect values and standardisation of formats. Security and privacy policies start to be applied to sensitive data.
- The **STAGING/SILVER layer** is a temporary zone where data is cleaned up and transformed before being moved to the next layer. Depending on the size of the data, if it is a small project, this layer disappears. At this stage, if applicable, you simply work with the data to quality and transform it.
- The **TRANSFORMED/TRUSTED/GOLD layer** contains data that has been completely cleaned, transformed and validated. These data are considered reliable and ready for analysis and decision-making, as well as for exploitation in Artificial Intelligence projects. These data comply with all rules and standards set by the organisation.

4. Keys to understanding data strategy, governance and management

4. Keys to understanding data strategy, governance and management

Reference Initiatives

There is an enormous concern in society to ensure that nobody is left behind in the race for competitiveness through data, and for this reason there are numerous initiatives, especially in the regulatory sphere, to revitalise and **accelerate** the definition and implementation of data initiatives **by organisations**.

Therefore, different **initiatives** of various types are being launched at all levels. For example, at European level:

- General Data Protection Regulation (GDPR): Key regulation on privacy and personal data protection.
- Data Governance Act (DGA): It facilitates the exchange of data between sectors, encouraging the re-use of data.
- European Data Strategy: Creation of common data spaces and their ethical use.

At state level, there are also regulations and initiatives in this regard:

- Data Office: It coordinates the national data strategy and promotes the efficient and ethical use of data in the public and private sector.
- Digital Spain Plan 2026: Initiative that promotes digitisation in all sectors.
- Spanish Law 18/2022, on Business Creation and Growth: It promotes the digitisation of SMEs, encouraging the use of data to improve their competitiveness.
- National Cybersecurity Act: Regulation guaranteeing security in the handling of data at national level.

At regional level, the Strategic Plan for Governance, Public Innovation and Digital Government, Ardatz 2030, has been published, which aims to "implement a strategy to promote new management models and a new culture of governance in public administration, make progress in the use of data to design services, and generate innovation in companies and citizens who can use them for entrepreneurial activities".

In order for an organisation to improve its data ecosystem, there are multiple frameworks and standards that can be taken as a reference, as they contain good practices in data governance and management. Some of these initiatives have been promoted by public institutions, others by software manufacturers and others by organisations set up by professionals specialised in data governance and management.

Illustration 11: Different frameworks and standards for data governance and management

One of the most active organisations is the international **professional association DAMA** (Data Management Association), which has various national and regional chapters and organises numerous events and working groups in which groups of professionals exchange impressions and best practices in order to evolve their methodology.

DAMA compiles all this knowledge in its **DMBoK: Data Management Body of Knowledge** and offers training courses that facilitate obtaining the **CDMP certification: Certified Data Management Professional.**

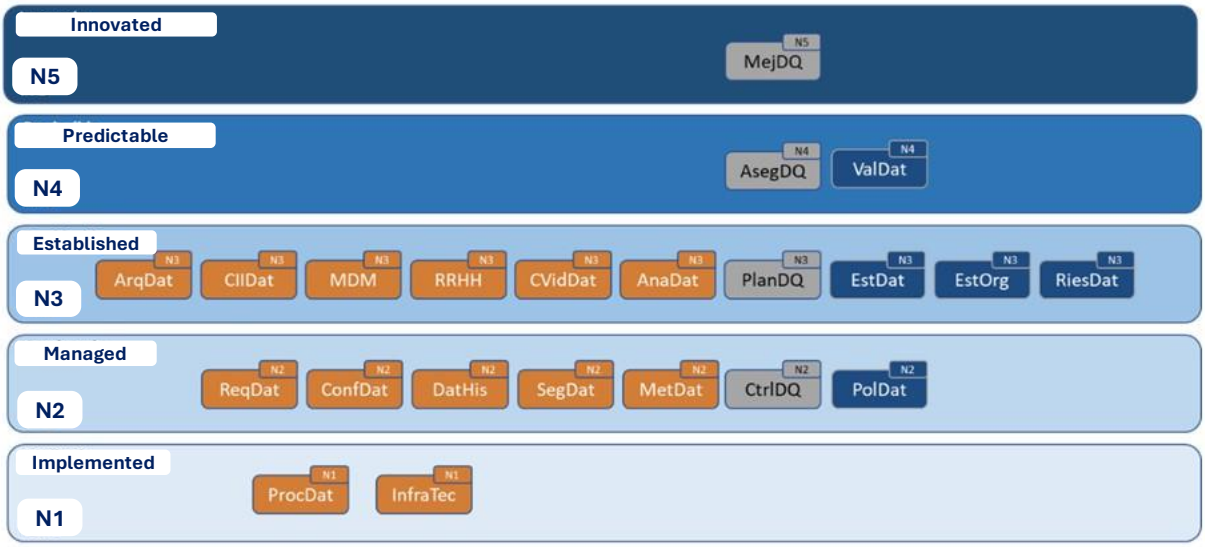
As an example of public initiatives, the Secretary of State for Digitisation and Artificial Intelligence of the Spanish Ministry of Economy, Trade and Enterprise, created the aforementioned **Data Office**, with the mission of boosting the management, sharing and use of data across the different productive sectors of the Spanish economy and society.

The Data Office, with the aim of bringing together the knowledge of the different existing frameworks, associations and specifications, has published a series of **UNE Standards** that cover the keys for the Governance, Management and Quality of data (e.g., the following image shows the ISO standards taken as a reference to develop the UNE standards).

Illustration 12: Relationship between ISO standards and UNE specifications on data – Spanish Ministry for Digital Transformation and the Public Function

Within this family of standards, the UNE 0080 offers a Data Governance, Management and Quality Management Assessment Guide that establishes different levels so that an organisation can certify the **maturity level** of its data ecosystem.

Illustration 13: UNE 0080 - Data Governance, Management and Quality Management Assessment Guide - Maturity levels of organisations



Data strategy

A data strategy is a long-term plan for managing and leveraging data within an organisation to drive decision-making, optimise operations and generate value. It involves the collection, storage, analysis, and secure use of data, aligning it with business objectives. This strategy encompasses everything from data governance, management and quality to the implementation of technologies and analytical tools, ensuring that data is accessible, reliable and useful to all levels of the organisation.

Having a data strategy is essential because it defines the scope and focus. The specific content of the strategy will be tailored to each organisation, but it should be defined holistically and articulated around the overall business strategy. It must be implemented iteratively as all phases of the strategy are developed and approved, and should therefore provide a roadmap that allows organisations to increase their level of maturity and thus be able to obtain increasing value from the data.

Having a strategy in place will help the organisation to increase its **level of data maturity**, as it will gradually establish an organisational structure for data governance, implement data management policies and procedures, incorporate the necessary technology to obtain the maximum performance from data, increase its competitive advantage and guide its actions based on the knowledge obtained from the data.

The strategy must be set out in a document that allows it to be made known to the organisation, monitored and continuously improved. One possibility is to set out the data strategy and its implementation in a comprehensive document known as a **"Data Governance Programme"**, a document that aims to translate the data strategy into the organisation's day-to-day operations. This document includes the objectives, scope, organisational structure, policies, procedures and implementation plan to subsequently implement the initiatives established in the plan.

By having a strategy, embodied in a Data Governance Programme, organisations will gradually build a mature data ecosystem, which will bring the following benefits:

Illustration 14: Benefits of having a Data Strategy

Therefore, one of the key aspects for an organisation to evolve its data ecosystem is to have a **data strategy**. Just as organisations have a defined strategy to meet their objectives, the same importance must be given to one of their main assets: data.

Having a data strategy consists of:

1. Defining Business Objectives:
    - Identify the strategic objectives of the organisation.
    - Align the data strategy with the business objectives to ensure that data delivers value.
2. Assessing the Current State of the Data:
    - Conduct an analysis of the current state of data management.
    - Identify gaps and opportunities for improvement.
3. Developing a Data Vision and Mission:
    - Create a clear vision of how the data will support the organisation's objectives.
    - Establish a mission to guide the use and management of data.
4. Defining the Data Architecture:
    - Design a data architecture that supports the business strategy and objectives.
    - Consider aspects such as the storage, integration and quality of data.
5. Drawing up an Action Plan:
    - Detail the initiatives and projects needed to implement the data strategy.
    - Establish a timeline and allocate resources.

As mentioned above, the "Data Governance Programme" will be the tool that will ground the strategy defined in the organisation, a document that includes the objectives, scope, organisational structure, policies and procedures of Data Governance that ensure the correct management of data, ensuring its quality, security and regulatory compliance. It also promotes the integration of data throughout the organisation, avoiding the creation of information silos. This facilitates a more global and coherent vision of data, improving strategic decision-making and encouraging the shared use of data between different departments and areas, aligned with the organisation's objectives.

Data governance and management

Another key to having a robust data ecosystem is to have proper **data governance and management** in place, and for this it is necessary to understand the difference between data governance and data management.

The purpose of data governance is to ensure that data is properly managed, according to the policies and best practices defined in the Data Governance Programme.

Data governance guides the data management execution functions, understood as the day-to-day operations related to the data lifecycle, such as the collection, storage, organisation and analysis of data within an organisation. It focuses on ensuring that data is accessible and useful for operations.

Example

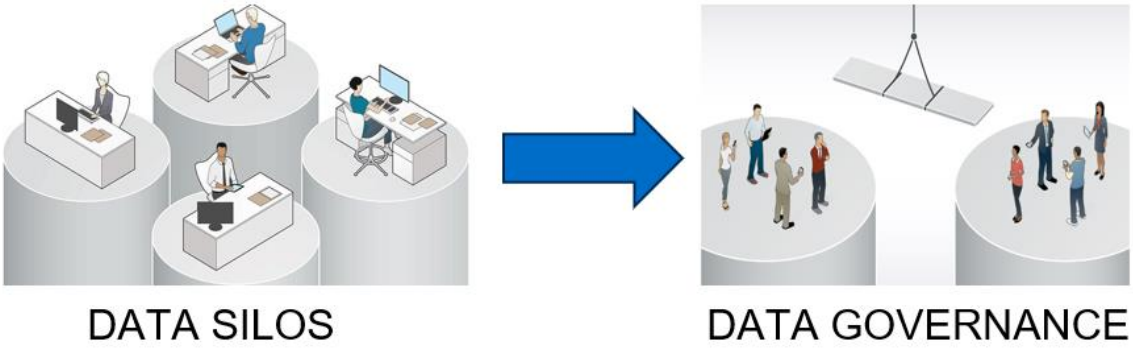
Just as an auditor controls financial processes, but does not execute financial management, data governance ensures that data is properly managed without executing data management activities directly

On the other hand, data governance is more strategic. It establishes the policies, rules and roles to ensure that data is used securely, efficiently and in compliance with regulations, avoiding silos and aligning data with organisational objectives.

Data governance is the driving force that guides all data management functions. Its purpose is to ensure that data is managed appropriately, in accordance with defined policies and best practices.

Data governance builds bridges within the organisation so that information silos do not occur and so that the organisation has an organised and 360° view of the data.

Illustration 15: Data Silos to Data Governance through the implementation of the Data Strategy



Data governance aims to materialise the data strategy, thanks to:

1. The establishment of Policies and Standards:
    - Create data governance policies that define how data will be managed.
    - Ensure compliance with legal standards and regulations.
2. The creation of a Governance Structure:
    - Designate clear roles and responsibilities (e.g., CDO, Data Stewards, Data Owners).
    - Form an organisational structure and its committees to drive, revitalise, oversee and manage data initiatives.
3. The development of Governance Processes:
    - Establish processes for data lifecycle management (creation, storage, use, archiving and erasure).
    - Implement control and audit mechanisms.
4. Promotion of the Data Culture:
    - Foster an organisational culture that values the use of data.
    - Provide training and awareness on the importance of data governance.

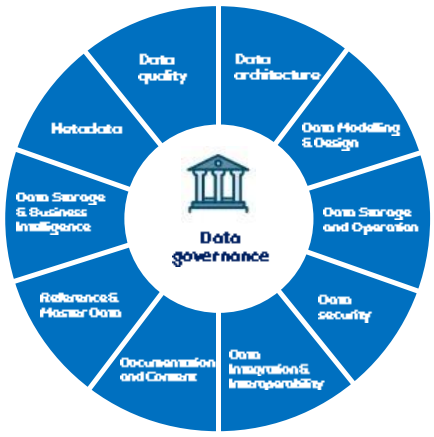
As indicated above, data governance is the auditor that ensures that all areas of data management are being executed correctly and in a uniform manner and based on the standards (policies and procedures) established by the organisation.

On the other hand, data management is the process of collecting, storing, organising, maintaining and protecting the information generated and used by an organisation. This

process encompasses a range of practices and technologies that ensure the quality, accessibility, security and usability of data, thus enabling effective analysis and informed decision-making.

DAMA establishes the "DAMA Wheel", which represents all the areas that must be taken into account for proper data management, highlighting that all areas of data management revolve around a common axis, the policies, rules and roles of Data Governance.

Illustration 16: Data Management Areas - DAMA



The DAMA Wheel represents the different areas of Data Management that must be taken into account for a proper data management, each of which has specific objectives and functions:

- Data Architecture: General structuring of data and associated resources as an integral part of the organisation's architecture.
- Data Modelling & Design: Analysis, design, construction, testing and management of data.
- Data Storage & Operations: Physical assets for data storage and management.
- Data Security: Ensuring privacy, confidentiality and appropriate access to PII, PHI and private data.
- Data Integration & Interoperability: Acquisition, transformation, movement, delivery, federation and virtualisation of data.
- Documents & Content: Storage, protection, indexing and access to data contained in unstructured sources (files) for integration with structured data (databases).
- Reference & Master Data: Management and standardisation of data shared between different domains to avoid redundancies and thus ensure greater quality and security.
- Data Warehousing & Business Intelligence: Management of the analytics process to allow the management and access to informational data to aid in decision-making through reporting and analysis.
- Metadata: Collection, categorisation, maintenance, integration, control and delivery of metadata (data about data).
- Data Quality: Defining, monitoring and maintaining data integrity.

To explain each of these areas in detail, DAMA, in its DMBok guide, establishes Context Diagrams, explaining each area, its goals/objectives, the different activities involved, their inputs and outputs, as well as the roles involved in these activities (providers, participants, consumers), together with the techniques and tools to implement them and the metrics to be able to measure the fulfilment of the objectives set.

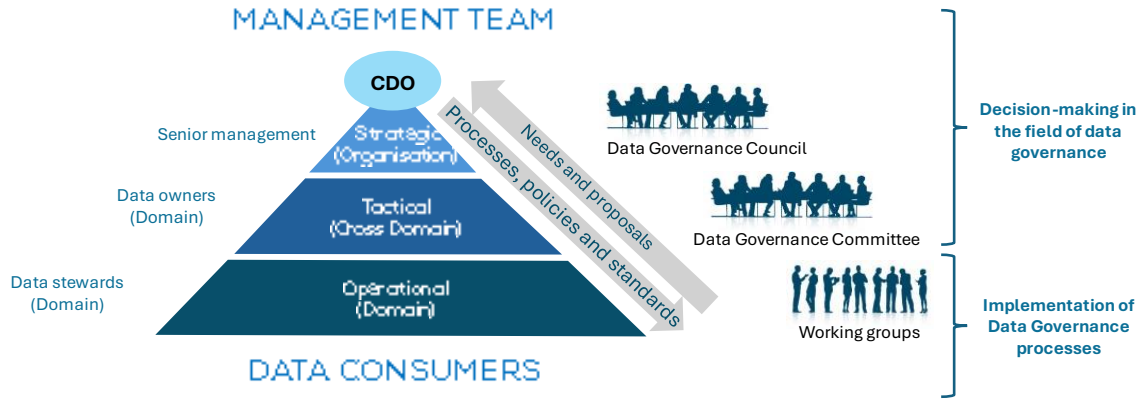
5. The data office

5. The data office

In order for there to be adequate governance and management of data in a global and homogeneous manner, it is necessary to have an organisation around the data where the issues related to this asset are dealt with. Just like any organisation has a group of people or department specialised in finance, production, sales, etc., in order to obtain the maximum performance from data, it is necessary to establish an organisation that leads and coordinates the initiatives linked to data. **The Data Office not only organises and allocates resources, but also coordinates the implementation of the data strategy within an organisation, ensuring that the data governance procedures and standards are met.**

Many organisations are establishing a Data Office that brings together different committees and groups of people to coordinate data governance and management. For this it is essential to have the whole organisation represented, from the management layer that provides the decision-making capacity, through the more functional part where there are specialists and experts in the different realities existing in the company and ending at the operational layer, which is where the data is actually managed.

Illustration 17: Data Governance Office



Within this organisation, a figure that acquires special relevance is the CDO (Chief Data Officer) who serves as a link between all areas and layers. This person knows both the reality of the organisation and the world of data, and acts as the "glue" so that communication flows throughout the organisation in a uniform manner.

Each layer has distinct roles, responsibilities and members, as explained below:

- The **strategic layer** is represented by the **strategic committee**, which approves and sponsors data initiatives.
    - Responsibilities:
        - Sponsoring the Data Governance programme
        - Providing an overview and long-term perspective
        - Leading cultural change in the organisation
        - Prioritising and funding data governance projects and initiatives
        - Approving organisation-wide data policies and standards
        - Resolving unresolved data conflicts

    - Members
        - CDO
        - Senior Management
- The **tactical layer** is represented by the **tactical committee**, which standardises and defines procedures, oversees, organises and proposes data initiatives.
    - Responsibilities:
        - Representing the entire organisation
        - Providing a short-term vision
        - Overseeing the progress of current data initiatives
        - Establishing data policies and standards for the entire organisation
        - Creating working groups for specific data initiatives
        - Facilitating cross-data initiatives
    - Members
        - CDO
        - Managers of the different data domains and specialities
- The **operational layer** does not have a specific committee representing it, but it is common to create **working groups** at this level to address the implementation of specific initiatives.
    - Responsibilities:
        - Executing data initiatives according to established policies and procedures
        - Managing data
        - Identifying and communicating needs and proposals
    - Members
        - Data administrators of the different domains
        - IT roles (data architects, data engineers)

6. Key profiles of the data ecosystem

6. Key profiles of the data ecosystem

As indicated in the previous section "Data Office", it is essential to establish an organisation that leads the governance and management of data, so that, throughout the different levels and committees explained, it is guaranteed that the maximum value is being obtained from the data, thanks to the fact that the established strategy is being followed and the established policies and procedures are being respected.

As mentioned above, there is a figure that is particularly relevant within the organisation, which is the **CDO (Chief Data Officer)**. This **figure is responsible for leading data initiatives, acting as a link between the different levels (strategic, tactical, operational) and, therefore, ensuring that there is uniformity and alignment of the entire organisation towards the established strategy.**

At a "functional knowledge" level, it is a good practice to establish a Data Domain Map, which defines how data is organised and categorised within an organisation, so that the existing knowledge in the organisation can be separated into different data domains (e.g., products, customers, sales, finance, etc.). In this way, those responsible for each domain can be established so that, in the event of a problem with any data in a specific domain, or in order to obtain any "data product" (any result or output generated from the analysis and processing of data within a specific domain), the person in charge can be contacted.

At a domain level, the key profiles are as follows:

- **Data Owner**: They are the ultimate decision-maker for the data in their domain. They are responsible for the management and protection of the data in their domain, collaborating in the definition of the policies for accessing and using the data, as well as providing the necessary information for their domain and guaranteeing the alignment of the data with the organisation's strategy and objectives.

- **Data Steward:** They are technically responsible for the data in their domain. They oversee data management and implementation processes and maintain tools to monitor, validate and improve data quality. They ensure that data is accurate, consistent and accessible, following the policies defined by the Data Owner. They also work closely with technical teams to resolve data issues and ensure alignment with established objectives and standards.

- **Data Translator:** This is the professional who has sufficient business and technical knowledge to express the needs of the organisation in a language that is valid so that the management, analytical and exploitation profiles can create algorithmic models that meet the requirements of the specific business use case.

As for data governance, the following profile stands out:

- **Data Governance:** This person designs and advises on the implementation of strategies for the effective management of data in an organisation and its alignment with organisational objectives. They collaborate in the definition and establishment of policies, processes and structures to ensure data quality, security and regulatory compliance. They also coordinate the implementation and maintenance of data governance policies and tools. They are also responsible for monitoring and resolving technical issues related to data governance.

In terms of data management, analytics and exploitation, the most common profiles are as follows:

- **Data Acquisitor**: They find connectivity solutions between processes and collect data through sensors or data ingestion flows. They know communication protocols, as well as the main components of a network and have knowledge of the software that connects the IT world with the physical world. Their basic programming skills allow them to make the connection between these two worlds. They can also establish cybersecurity standards and are able to audit and make proposals for their assurance. Expert in sensor systems and IoT.

- **Data Engineer:** They are responsible for integrating data from diverse sources, optimising data pipelines and ensuring the scalability and performance of data solutions. They work closely with data scientists and analysts to facilitate access to reliable data that is conveniently prepared for use.

- **Data Architect:** They design and structure an organisation's data architecture, ensuring that storage and processing systems are efficient and scalable. They define standards, data models and strategies to align the data infrastructure to the organisation's objectives.

- **Data Analyst:** They examine and analyse datasets to generate information to support business decision-making. Their responsibilities include cleaning and processing data, performing statistical analysis and creating visualisations that communicate key findings. They work closely with different departments to understand specific needs and provide data-driven solutions.

- **Data Scientist:** They develop advanced analytical models, combining skills in statistics and programming. They apply machine learning techniques using large volumes of data and validate results to resolve complex problems.

7. Enabling technologies

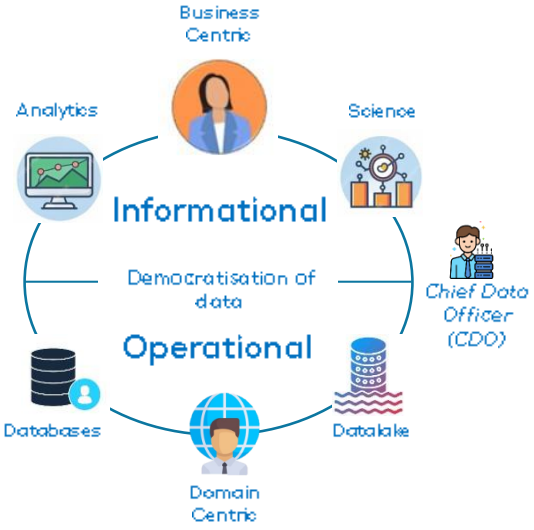
7. Enabling technologies

For the proper governance and management of data there are numerous tools specialised in these tasks, but in order to choose the most appropriate combination of tools, it is first necessary to understand that the data ecosystem consists of different environments with distinct needs, as explained in "section 3. Operational versus Informational Data" of this document.

As indicated above, **the operational environment refers to transactional data that is used in the normal operations of an organisation and for which specific usage tools are generally established**, e.g., a CRM for customer management or an ERP for production processes. This environment is governed by technical staff who are familiar with the technological systems that store the data, so business profiles usually have limited access to the data and are always dependent on what the tools or technicians provide.

On the other hand, **the information environment refers to the data that business managers need for decision-making**. Business needs to be able to have access to quality data that is understandable, free of interpretation errors and that can be used for the organisation's strategic decision-making, so it must be abstracted from any technological or conceptual constraints that may exist. Informational data is data of full quality, clearly interpretable and accepted by the entire organisation in order to make decisions based on it.

Illustration 18: New data ecosystem



Each of these environments requires different capabilities and technologies, and there are currently very different types of solutions:

- By type of implementation: On Premise (local) or Cloud (in the cloud)

    - On-premise technological solutions involve the software and infrastructure being installed and managed on the organisation's own servers, offering greater control, but with higher costs and maintenance responsibilities. Cloud solutions, on the other hand, are hosted on external servers managed by a provider, offering flexibility and scalability, but with less direct control over the infrastructure.
- By type of licence: Open Source or Proprietary (Paid software)
    - Open source solutions consist of software whose source code is accessible, allowing users to freely modify, distribute and improve the program. Proprietary solutions are developed and controlled by a company, restricting access to the source code and limiting its use, modification and distribution according to the terms of the contracted licence.
- Etc.

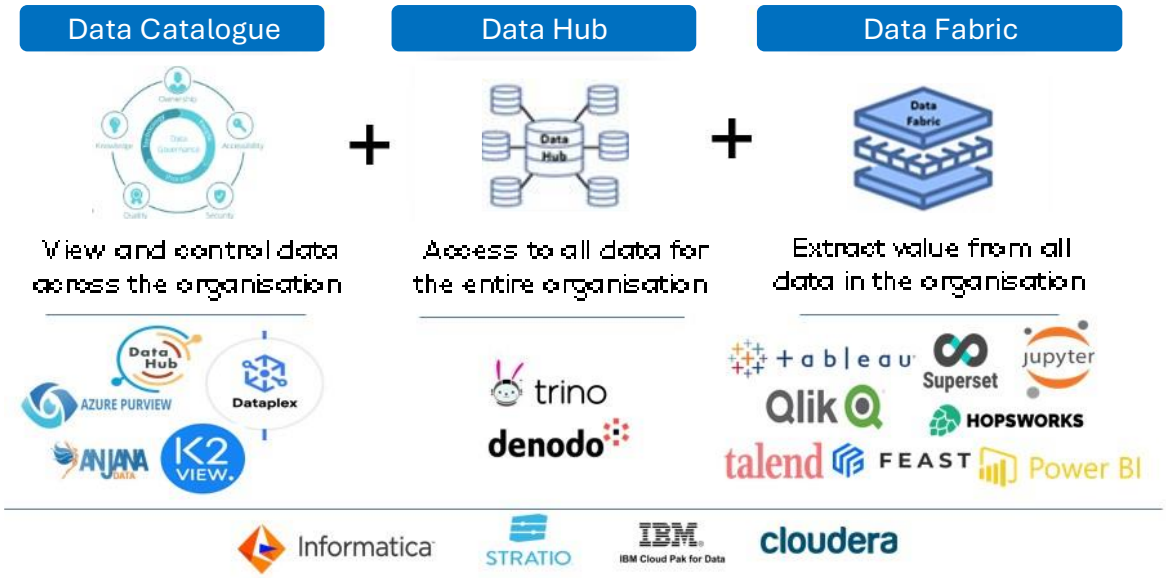
Therefore, when defining the organisation's technological ecosystem, the requirements to be covered for each environment (Operational or Informational) must be taken into consideration, as well as the strategy and objectives of the organisation itself, so that the technological infrastructure is aligned with them.
In the operational environment, it is most common to have relational databases (tables with data distributed in rows and columns) which provide great consistency to the data and speed in response time, but which also suffer when they receive a high volume of requests, making them unsuitable for analytical or data science projects.

One of the problems with existing data in operational environments is that it has been managed for a specific purpose and therefore, technicians generally use specific nomenclatures and definitions to facilitate its development, meaning that it is not a data that can be easily exposed to business profiles, as it would be difficult to understand its meaning and would lead to misinterpretations.

This is why it is necessary to have a semantic layer, which is responsible for translating the operational data into a language that can be interpreted by the organisation and used in informational environments, and here concepts such as ontology acquire special relevance, allowing the identification of the terms used in the organisation, the relationships between them and their definitions.

In order for the organisation to have quality information data, duly protected and accessible to whoever needs it, there are different technological solutions that are extremely useful:

Illustration 19: Technologies for data integration and governance



Data Catalogue solutions or tools

This allows for a centralised view of the existing information in the organisation, with its corresponding cataloguing and including both technical and business information, so that all the knowledge available in the organisation can be visualised in a centralised manner.

The data catalogues allow:

- The visualisation of metadata (additional information about the data, such as the date or the location where a photo was taken), so we can differentiate access to the information in the data (metadata) from access to the data itself.
- Having a technical dictionary (technical information specific to the data) and a business glossary (description of the meaning of the data as stated, for example, in the ontology of the organisation).
- Establishing knowledge domains and identifying responsibilities within the domains (data owner – data steward).
- Visualising the lineage or traceability of the data, in order to know the origins and/or the transformations suffered by certain data.

A Data Catalogue is primarily used to visualise and discover data, providing information about where it is located, how it is structured and who is responsible for it, thanks to metadata management and search tools.

Solutions or tools to Centralise Data – Data Hub

It allows centralised access to the organisation's data, regardless of the technology or system in which it is stored and through a single language (SQL).

Data Centralisers allow:

- Management of access permissions to the different systems where the information is stored.
- Having a unified view of the information, even if it is in different systems (data federation).
- Management of the confidentiality of information by anonymising data according to its cataloguing and the roles of the user accessing it.

A Data Hub focuses on centrally accessing and extracting data. It acts as an integration point where data from multiple sources is connected and distributed for real-time or on-demand use.

Data exploitation solutions or tools – Data Fabric

They offer solutions for data exploitation, for descriptive, analytical or data science purposes.

Examples of the functionalities offered are:

- Business Intelligence (BI) solutions for generating reports or dashboards, for example
- Data ingestion and warehousing for analytical and data science purposes (DataOps, MLOps)
- Automating the generation of Machine Learning models
- Data sharing spaces (Data Marketplaces)