LUISS Guido Carli
LIBERA UNIVERSITÀ INTERNAZIONALE DEGLI STUDI SOCIALI

Department of Management

Master's degree in International Management

Chair: Digital Transformation

Data governance:
securing the future of financial services

SUPERVISOR

Prof. Paolo Spagnoletti

CANDIDATE

Luca Robimarga

682971

CO-SUPERVISOR

Prof. Stefano Za

ACADEMIC YEAR 2017/2018

INDEX

INTRODUCTION

Nowadays the evolution of the business world has reached unimaginable rhythms up until a few decades ago. The war between companies and scholars, in predicting trends and understanding the drivers of tomorrow's success, is becoming more and more "bloody", but there is one thing that unites all current and future trends and drivers: data.

Everything that surrounds us today is governed by data. The data as such, however, are worth very little; what gives value to them is the ability of those who possess them to know how to read, aggregate and manage them. This is the source of the great value of data: from the irrefutable potentials and their ability to determine who wins and who loses in every business sector.

This data potential is slowly becoming apparent to all business actors, and investments in their management are growing day by day. Investments in this field are also increasing as a result of the fact that data, when they assume such value, on the one hand, if well managed, bring great benefit to the company and on the other hand, if mismanaged, they expose to a great risk both the company and third parties. Precisely for this reason, to accompany companies towards a process of data management, regulations and reference principles have emerged that dictate the perimeter and the constraints of a correct management of data. This is how data governance projects begin to emerge within companies.

The Data Governance has the final objective of increasing the awareness of all the actors operating within a company about the data circulating within it and of defining the responsibilities of data production.

This process takes place within each company structure through the use of multiple tools, each with a precise functionality, which we will analyze in detail, accompanied by a particularly exemplary corporate case.

The initial phase of implementation of the Data Governance model takes place through Assessment, a phase in which the Data Governance Office, through the census tools and a questionnaire, collects the necessary preliminary information. The

assessment phase concludes with the analysis of the data collected through the Prioritization Matrix from which the most critical structures and reports (vehicle of data) are evinced and therefore require timely intervention. Through this analysis the construction of the project masterplan takes place, which will be used for the planning of each single wave (ie implementation of the project to each individual company structure).

Once the intervention ranking has been determined, we proceed to the punctual study of every single data produced by each single structure through the Data Lineage. Each datum is mapped, registered and cataloged respectively with three different instruments: the Logical path of data used to graphically represent the logical path carried out by the data, from its creation in the systems up to its final form; Data Registry thanks to which it is possible to record all the controls, transformations, systems and organizations through which the data passes before reaching its final form; Business Glossary, an instrument thanks to which every piece of data is cataloged in order to make it universally recognizable by every actor of the entire company.

After completing the lineage phase, all the information about the controls, the transformations and the mapped systems are combined in order to determine the quality of each single data. Its quality is determined through the Key Quality Indicators that give a vision of how the data has been treated throughout its life cycle in order to determine its reliability. Once the latter has been determined, it will be possible to identify the difference that exists between the actual quality of the data and the expected quality of the same in the planning phase.

The last step of the project consists of two steps: identification of the critical issues that did not allow the data to obtain a sufficiently high level of quality to meet expectations, and implementation of a remediation plan that allows to minimize the gap between as-is model and to-be model.

As we will see later, a Data Governance project has a beginning but not a well-defined end, but to fully understand this statement, let’s retrace the fundamentals of Digital Transformation and then go into the detailed study of data.

CHAPTER 1

1. The definition of Data Governance from theory to practice

Among the words that pertain to the business world one of the most used in recent years is definitely Digital Transformation. Too often we hear companies that speak of radical transformation and change processes, without however having in mind clear objectives and understand the real business needs of the company. Around the term Digital Transformation there is a lot of confusion: in many cases, companies believe that a transformation path corresponds to the inclusion of new technologies in the company structure and do not worry about how this infrastructural change can impact and be absorbed by workers.

1.1 Digital transformation: the world driven by data

Before embarking on a path of Digital Transformation it is good to stop and think first that not everything is good for everyone and that not all changes may be necessary. In any case, a fundamental element must be taken into account by all companies: the data. In today's business environment, data are the new asset on which the value of a company is based, and not just those born on the internet, belonging to the platform economy or startups.

The so-called incumbent companies (organizations with an almost monopolistic past that today are competing with start-ups and new business models) have a very high potential to become companies' protagonists of the data driven economy because they possess the fundamental asset to be part of them: data. The crucial point becomes: how to use this data and how to make it the treasure that can potentially be? Data is important because it brings with it two fundamental characteristics of digital transformation: speed and traceability. In terms of speed, digital is spreading at exponential rates and pervades both our private life and work. But speed and acceleration are typical elements of any major revolution history has brought with it: just think of the industrial revolution and how it has changed the way it works by simplifying and speeding up processes.

Traceability, on the other hand, is an exclusive feature of the digital world and is certainly the aspect of greater value linked to data. Using traceability means being able to analyze data for valuable information and deductions and is not unique to younger companies, as the potential of a company that has collected endless amounts of data over the years is just as infinite. This is why it is becoming increasingly important for companies to exploit suitable analytical tools.

The fact that companies become more and more data driven is now an urgent need: data and their correct analysis and interpretation can promote important changes in terms of products and services, can allow companies to take advantage of new business opportunities and, above all, allow you to have a completely new approach to customers, providing them with personalized services or products.

Often, however, when talking about Big Data, the name is associated exclusively with the concept of volume. In reality there are many other characteristics of Big Data that are summarized in the so-called 7 V: volume precisely; speed, because data must be accessible in real-time; variety and variability understood respectively as the different types of data (often unstructured) and the different meaning that a given can bring with it depending on the use made of it; truthfulness concerning the level of accuracy of the data; visualization, understood as the ability to provide adequate technological tools for the analysis and interpretation of data; finally, but not least, the most important V, that is the one linked to the concept of value, since it is fundamental that once the data are collected and analyzed, the necessary value can be obtained for the company and for the different Business Units.

The data is only the first of many factors to consider when embarking on a coherent and responsible Digital Transformation process. But they are certainly an excellent starting point to understand that, often, what is called Digital Transformation should really be called Business Transformation, because a conscious use of data brings with it a series of chain reactions, including a new customer approach, a change in corporate assets and new organizational models that substantially change the business and business model of companies, thus preparing a radical transformation.[1]

[1] Consuelo Sironi, 06 March 2018: https://www.ilsole24ore.com/art/management/2018-01-19/il-valore-dati-e-asset-chiave-digital-trasformation-154035.shtml?uuid=AEzwdclD

What is a data driven organisation?

A Data Driven organization is based on data and not on opinions to make its own decisions. The abundance of data, together with the development of the technology that characterize our era, have made it easier to register and store in an ever-increasing quantity and they can therefore be profitably used for the Decision-Making process.

Good data, however, do not guarantee good decisions, as it is essential to know how to interpret and direct all the company organization towards the "culture of data", so that every subject that is part of the process is ready to receive them.

What are the issues to be considered to develop a data driven organization?

The main topics that a company must address to develop a data driven organization are the following:

1) Incorrect diffidence towards a decision making based on data: decisions based on data are the best because they are based on facts rather than on personal opinions or experiences;
2) Data visualization theme: often the data you need are not immediately reachable and are not displayed in the right way, therefore they cannot be correctly interpreted and cannot be used to support decision-making processes. It is essential to have synthetic and detailed dashboards built specifically for the business needs that allow consultation of the data. Data visualization is essential for the decision maker, in order to focus on the really important elements, without wasting time in unnecessary details, in order to concentrate on solving problems and seizing new opportunities;
3) KPIs theme: It is important to build indicators, at all company levels, to evaluate the performance of data driven decision-making processes as otherwise it is not possible to evaluate the improvements or to set objectives;
4) Shortage in the company of competent figures ("people-oriented data experts") who are able to extract value from the data. Once the data are stored, it is essential to understand how to treat them: what to keep and what to delete, so as to transform the raw data into reports ready for consultation at the front - end of the subjects responsible for making decisions;

5) Silos company organization: the company is often structured in watertight departments that do not communicate with each other, each with its own data. The lack of integration between the data does not allow the validation of the same, and therefore prevents to have the security of the use of the correct data at the basis of a decision-making process. It is therefore necessary, at the information system level, to integrate company data at every level, that is between sources and databases of different compartments, and between physical and non-physical data.[2]

1.2 Data governance: transforming numbers in assets

As we saw in the previous paragraph, the informative patrimony of any organization, that is the whole of the information related to all the operations that are part or influence its activity, is, together with people and financial resources, the most valuable asset available.

Transforming data, the mass of data that flows with virtually continuous flow into Information Systems in order to have the 'unique and true' knowledge of what has happened and takes place in a company is a process that absorbs important human, economic and technological resources. It is all the more complex and onerous as the starting data come from different sources with different structure and all the data, in general, are processed with unreliable means. But it is precisely based on data from different sources that gives value to information, merging into a single vision the points of view that the different actors have of the same event.

It is from this observation that the concept of Data Governance is born, which can be defined as the set of activities aimed at managing people, processes, methodologies and information technologies in order to achieve a constant and correct treatment of all data that have importance for an organization.

[2] Sara Pea - Specialista Business Intelligence – Gruppo Sme.UP, 11 August 2018: https://www.smeup.com/blog/blog-business-intelligence/organizzazione-data-driven/

The definition of Data Governance is not a technology, but a set of strategies, processes and rules that allow data to be processed and exploited in the best possible way.[3]

In practice, most Data Governance projects aim to standardize the definition of data by the various company functions, to establish common access and use rules and to identify the subjects involved, defining their responsibilities. Later in the thesis we will see how all these results can be achieved and through which tools.

Therefore, even if includes the use of Data Quality tools, it is not a technology, but a set of strategies, processes and rules to be defined upstream of the use of data, purpose of exercising an effective control on the processes and methods used by the administrators to prevent errors and to suggest the necessary interventions to solve the problems created by poor quality data.

The first advantage given by Data Governance is, obviously, to allow the so-called 'knowledge workers' (which are not only decision makers but begin to include employees with operational tasks) of access and share information provided by different applications and databases and be able to count on quality data in order to do their job better. But effective governance and data control also have important effects on safety, with reduction of risks deriving from operations and the possible failure to comply with laws and regulations (this aspect will be soundly examined later).

Before to going on, it is important to make it clear on the subject the definitions and the differences between Data Governance and Data Quality Management.

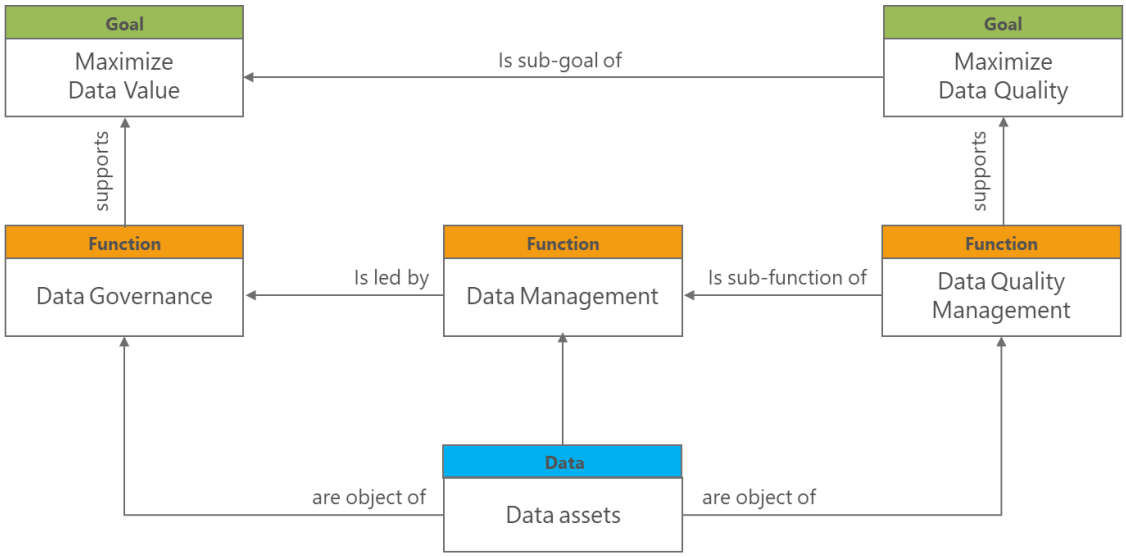
A unique means, the assumption and the vision of the "Data Governance "cannot be found between the research community and the community of information systems professionals, however, both communities agree that Data Governance refers to the assignment of decision-making rights and related duties in data management in companies. According to Weber et al (2009), for example, Data Governance specifies a framework for decision-making of rights and responsibilities regarding the use of data in a company. Khatri and Brown (2010) see Data Governance as referring to the assignment of decision rights in relation to a company's proprietary data. Data Governance aims to

[3] Giampiero Carli Ballola, 4 October 2010: https://www.zerounoweb.it/techtarget/searchsecurity/la-definizione-della-data-governance-il-primo-passo-per-la-visione-unica-della-realta/

maximize the value of data assets in enterprises. They are discussing whether the value of data can and should be determined for financial accounting purposes (Atkinson and McGaughey 2006). Data's "fitness for use" is what Wang (1998) considers as data quality. Poor data quality reduces the value of data assets in an enterprise if their utility is low (Even and Shankaranarayanan 2007).

So, businesses are anxious to maximize data quality. Maximizing data quality is the goal of data quality management. DAMA International (2009, 20) defines data quality management as a function to "measure, evaluate, improve and guarantee the fitness for use". Data quality management is therefore a sub-function of data management, which includes planning, control and provisioning of data resources (DAMA 2009). The relationship between Data Management and Data Governance is based on a differentiation proposed by the International Organization for Standardization (ISO) on governance and management (ISO / IEC 2008). Following this differentiation, Data Governance is the main function of data management as it specifies what decisions must be made in data management and who makes these decisions. Data management ensures that these decisions are taken and applied.

The graph 1.1 shows the relation between Data Governance, Data Management and Data Quality management and the process thanks to which the value of data can be maximized.



Graph 1.1: Relation between Data Governance, Data Management and Data Quality management. Source: Otto, Boris, and Kristin Weber. 2011.

According to a 2006 survey (Russom 2006) it has been shown that data governance is rarely adopted. Only 8%, between companies and banks, implemented a data governance initiative, and 17% were in the planning or implementation phase.

Within the scope of a data governance model in a real case, we must also start talking about IT governance and therefore of all the architecture supporting the Data Governance process.

With the term IT Governance, or Government of IT, we mean all those processes aimed at managing the IT systems of a company. In particular, this discipline is interested in analyzing and managing everything related to IT risks and the processes that regulate its activity. IT Governance is only one of the gills of the broader Corporate Governance, a discipline that, as the term suggests, corresponds to that set of rules governing the administration of a company. It follows that IT Governance establishes the rules to ensure compliance with the company's information technology standards and procedures.[4]

Research on IT governance is more advanced than research on data governance, with the first publications dating back to 25 years ago (Brown, C.V. 1997). IT governance allows to define roles and responsibilities among the various business structures.

The research on IT governance proposes three elements that make up an IT governance model: roles, the main decision areas and the assignment of responsibilities. Suppose such flexible models - instead of universal approaches to data governance postulated by previous research - would help companies' structure and document their specific decision-making framework for the Data Quality Management. Therefore, we adopt the idea of the IT governance model to build a model for data governance.

[4] Sabrina Poli, 8 May 2017: https://www.matika.it/it-governance-gestione-rischi-processi-it-framework-best-practice-standard-internazionali/

Usually, to define Roles, decision areas and responsibilities, companies use the RACI matrix. It is an important tool to support project planning. It integrates the information of the WBS (Work Breakdown Structure) and the OBS (Organization Breakdown Structure), essentially defining the "who does what". In this sense it contributes to:

- highlight immediately what should be done, who should do it and with what organizational role;
- formalize the role not only of those who will actually have to do the job but also of those who will have to support them;
- favor a better exploitation of the costs of each activity, incorporating not only operational / executive but also support ones;
- create awareness of the impact of everyone's work on the work of other team members;
- create responsibility among the members of the project team;
- favor the commitment also by the managers of the resources involved.

The name "RACI" is an acronym of the 4 possible roles that can be associated with an activity:

1) **Responsible** - is the role of the one who is called to perform the task operationally (for each task it is possible to have more Responsible)
2) **Accountable** - it is usually the role to which the Responsible reports in the project organization chart or which will have to play a supervisory role of the Responsible work (must be univocally identified and will have to account for the work of the resources he coordinates performance of the activity)
3) **Consulted** - is the role of who will support the Responsible in carrying out the task providing useful information to complete the work or to improve the quality of the work itself
4) **Informed** - is the role of who should be informed about the work of the Responsible and who will have to make decisions based on the information received

The RACI is therefore constructed by associating to each activity the individuals or the organizational roles that will participate in it, indicating the specific role for each activity.

It is therefore possible that the same person has a role on an activity and a different role on another activity.[5]

The graph 1.2 shows an example of allocation of roles and responsibilities in a Data Governance project (them change bank by bank according to the structure and needs)

RACI MATRIX		ROLES				
		Executive Sponsor	Data Governance Council	Chief Steward	Business Data Steward	Technical Data Steward
DECISION AREAS	Plan data quality initiatives	A	R	C	I	I
	Establish a data quality review process	I	A	R	C	C
	Define data producing processes		A	R	C	C
	Define roles and responsibilities	A	R	C	I	I
	Establish policies, Proc. and standards for DQ	A	R	R	C	C
	Create a business data dictionary		A	C	C	R
	Define information systems support		I	A	C	R

Legend: R – Responsible; A – Accountable; C – Consulted; I – Informed

Graph 1.2: RACI matrix. Source: Wende, Kristin. 2007

However, it is important to stress that data governance is not a complete subset of IT governance. As outlined above, achieving the quality of corporate data requires close collaboration between IT and business professionals who understand the data and their business purpose. Therefore, we argue that data governance and IT governance are co-equal and both must follow the principles of corporate governance.

The application of a Data Governance model obviously changes according to the needs of each company / bank, their organizational structure, the initial level of data governance and results of the risk analysis process. Above all, this last aspect can be decisive in the drafting of a project of Data Governance and in the determination of budget, effort and development/application timing of the project. This thesis, in order to bring the maximum

[5] http://www.humanwareonline.com/project-management/center/ram-matrice-assegnazione-responsabilita/

possible benefit to the reader, will take as an example a case in which there is a model of data governance at an early stage, where the circulation of information within the bank is entrusted almost exclusively to the manual work of operators through basic computer tools (better defined as "EUC" or End User Computing), and the individual responsibilities for the data processed are not well defined.[6]

[6] Giampiero Carli Ballola, 4 October 2010: https://www.zerounoweb.it/techtarget/searchsecurity/la-definizione-della-data-governance-il-primo-passo-per-la-visione-unica-della-realta/

CHAPTER 2

2. Focus on normative and principles

As we have said, the implementation of a Data Governance model, that ensures Data Quality, changes from bank to bank and from company to company, but all the Data Governance projects have as common denominator the guidelines drawn up by the reference supervisory body. In our case, the relevant normative are the "BCBS 239, January 2013" (Basel Committee on Banking Supervision) of the Bank for International Settlements and the Supervisory Provisions for banks with the "Circolare n. 285 of 17 December 2013 "with the 11th Update of 21 July 2015 of the BANCA D'ITALIA.

The BCBS 239 gives the general principles to strengthen banks' risk data aggregation capabilities and internal risk reporting practices. In turn, effective implementation of the Principles is expected to enhance risk management and decision-making processes at banks.

The adoption of these Principles will enable fundamental improvements to the management of banks. The Principles are expected to support a bank's efforts to:

- improve the infrastructure for reporting the most important information, in particular those used by the board of directors and senior management for identify, monitor and manage risks;
- improve decision-making in the entire banking organization;
- improve the management of information among the various legal entities, favoring at the same time a comprehensive assessment of level risk exposures global consolidated;
- reduce the likelihood and severity of losses linked to deficiencies in the management of risks;
- reduce the time taken to prepare information and therefore speed up the process decision-making;

- improve the quality of an institution's strategic planning and its capacity to manage the risks inherent in new products or services.

Obviously, risk data aggregation capabilities and risk reporting practices cannot exist in isolation and they are clearly inter-linked. The quality of reporting for the risk management depends on the aggregation capacity of the data, and the relative flows of information from the presence of solid infrastructures and corporate governance mechanisms. Banks should comply with all the Principles on aggregation and reporting of risk data simultaneously. However, in exceptional circumstances compromises among different Principles could be accepted, for example in the case of urgent or timely requests for information on new or little-known areas of risk. These compromises should in no case have significant consequences on decisions in risk management. The decision-making bodies of banks, in particular the board of administration and senior management should be aware of such compromises the limits or shortcomings they entail.

Supervisors expect banks to adopt policies and processes for the adoption of compromise solutions of this kind. Banks should be in able to explain the consequences of these choices on their decision-making process qualitative relationships and, as far as possible, quantitative measures.

Banks should develop the ability to produce prospective reports capable to provide early indications on violations of risk limits potentially exceeding the bank's risk propensity / tolerance. This capacity should also allow the bank to carry out flexible and effective stress tests, able to provide one prospective risk assessment. Supervisors expect that reporting risk management allows banks to anticipate problems and provide one prospective risk assessment.

All the data creation and its flow has to be automated but occasionally the expert judgment may be applied to incomplete data to facilitate the aggregation process, as well as the interpretation of results within the risk reporting process. The appeal to the judgment of an expert rather than complete and accurate data it should only take place exceptionally, and it should not have a significant impact on the Bank's compliance with the Principles. In case of appeal to an expert, supervisors expect this to happen in a transparent manner

clearly documented, to allow an independent verification of the procedure followed and of the criteria used in the decision-making process.

2.1 The BCBS 239 of the Bank for International Settlements

The BCBS 239 give the Principles for effective risk data aggregation and risk reporting. Those principles can be allocated in four closely related topics:

1) Overarching governance and infrastructure
2) Risk data aggregation capabilities
3) Risk reporting practices
4) Supervisory review, tools and cooperation

2.1.1 Overarching governance and infrastructure

Banks should have a solid corporate governance system, one robust risk data architecture and a reliable IT infrastructure. These elements are essential to ensure compliance with the other principles. In particular, the bank's board of directors should monitor the responsibility of senior management to apply all relevant principles aggregation and reporting procedures of risk data, as well as on the strategy adopted to meet these standards in the time agreed with the supervisory authorities.

Principle 1 – Governance

"A bank's risk data aggregation capabilities and risk reporting practices should be subject to strong governance arrangements consistent with other principles and guidance established by the Basel Committee."

The board of directors and senior management should promote the identification, assessment and management of data quality risks in the context the overall framework for risk management. The latter should contemplate service level agreements for both

outsourced and internal processes related to risk data, corporate policies on confidentiality, integrity and availability of data, as well as policies risk management.

The aggregation capabilities of risk data and related reporting practices of a bank should be:

- Fully documented and meet high standards of validation. The validation process should be independent and verify compliance from part of the bank of the Principles. The main purpose of the independent validation is to ensure that, aggregation processes and risk data reporting, work correctly and is appropriate to the profile of risk of the bank. Independent validation activities should be aligned and integrated with other independent verification activities under the risk management of the bank;
- Be considered an integral part of any new initiative, including acquisitions and / or disposals, development of new products, as well as more general initiatives of change in processes or IT systems. In evaluating the opportunity of a significant acquisition, the due diligence process of the bank should consider the aggregation capacity of risk data and the related practices of reporting of the entity being acquired, as well as the impact of the transaction would have on its own abilities and procedures on the matter. The consequences for the aggregation of risk data should be explicitly considered by the Board of Directors and re-enter the decision to proceed with the operation. The bank should define the period within which to integrate and align the data aggregation capabilities of risk and related reporting practices of the acquired entity;
- Not be affected by the structure of the banking group. The structure of the group does not should hinder the aggregation of risk data or at the level consolidated or at any relevant level within the organization.

Senior management should have full awareness and understanding limits that prevent complete aggregation of risk data, such as limits in terms of coverage (undetected risks or affiliates not considered), technical limits (indicators model performance or recourse to manual procedures) or legal limits (obstacles of a nature legal data sharing between jurisdictions). Furthermore, management should ensure that the bank's IT strategy includes ways to improve risk data aggregation capabilities and risk reporting practices

and to remedy any shortcomings against the Principles and also identify data critical to risk data aggregation and IT infrastructure initiatives through its strategic IT planning process.

Principle 2 - Data architecture and IT infrastructure

"A bank should design, build and maintain data architecture and IT infrastructure which fully supports its risk data aggregation capabilities and risk reporting practices not only in normal times but also during times of stress or crisis, while still meeting the other Principles."

The aggregation capabilities of risk data and related reporting practices should be explicitly considered in the continuity planning processes operating the bank and be subjected to a business impact analysis.

Banks should define taxonomies and an integrated data architecture for the whole group, including information on the characteristics of the data themselves (**metadata**), as well as the use of unique identifiers and / or unified conventional designations for mark data such as legal entities, counterparts, customers and accounts. Specific roles and competences regarding "ownership" should be established (**ownership**) and the quality of data and risk information for both operational functions both for information technology. The owners (operational and IT functions), in collaboration with risk managers, should ensure the presence of adequate controls throughout the whole life cycle of data and for all aspects of the technological infrastructure. The owners of the operational functions must in particular ensure that the data are correctly entered from the relevant front office unit, are updated and aligned with the definitions provided, and that the aggregation of risk data and related reporting practices are consistent with the bank's policies.

2.1.2 Risk data aggregation capabilities

Banks should develop and maintain solid aggregation capacity risk data to ensure that the relevant reports provide a reliable representation risks (meeting expectations for data

aggregation is necessary condition for meeting reporting expectations). The Respect for a Principle should not take place at the expense of compliance with another Principle. The aggregation of risk data should satisfy all the Principles concurrently.

Principle 3 - Accuracy and Integrity

"A bank should be able to generate accurate and reliable risk data to meet normal and stress/crisis reporting accuracy requirements. Data should be aggregated on a largely automated basis so as to minimise the probability of errors."

Despite the importance of accurate data, one study estimates that between 15 and 20 percent of a typical organization's data are wrong or unusable (Cohen, Shoshanah, and Joseph Roussel 2005). Banks should aggregate risk data accurately and reliably. Risk data checks should be as strict as those applicable to accounting data. Where a bank uses manual procedures and IT applications desktops (e.g. spreadsheets and databases) and have specific units of risk using these applications to develop software, the bank should prepare effective risk mitigation tools (e.g. procedures and policies on end-user computing) and other effective controls to be applied uniformly to all bank processes. In order to guarantee accuracy, the risk data must be connected to the sources of the bank, including where appropriate accounting data. Furthermore, persons in charge of risk control functions should have access sufficient to the risk data, which allows them to aggregate, validate and adequately link these data for reporting purposes.

As a prerequisite for the fulfillment of this principle, banks should get a "glossary" (which, later in the thesis, will be named "Business glossary") of the concepts used, so that the data receive the same definition throughout the organization.

Obviously, does not exist a perfect theoretical balance between automated and manual systems. In cases requiring the application of professional judgment, a manual intervention may be appropriate, while for many other processes a greater one is desirable automation level in order to reduce the risk of errors.

Supervisors expect banks to document and illustrate all data aggregation processes, both automated and manual (based on judgment professional or other). The documentation should, among other things, explain the opportunity alternative manual solutions that may be adopted in case of system problems information technology, their relevance for the accuracy of the aggregation of risk data and measures proposals to reduce its impact. Supervisors expect banks to measure and monitor accuracy of data and prepare appropriate channels for relevant issues to be brought to the attention of the appropriate hierarchical levels, as well as operational plans to remedy one poor data quality.

Principle 4 – Completeness

"A bank should be able to capture and aggregate all material risk data across the banking group. Data should be available by business line, legal entity, asset type, industry, region and other groupings, as relevant for the risk in question, that permit identifying and reporting risk exposures, concentrations and emerging risks."

In order to be compliant with these principles, it is not necessary for a banking organization to express all forms of risk with a single measure or on a common basis, however the data aggregation capabilities of risk should be the same regardless of the choice of aggregation systems used. Nonetheless, every system should clearly indicate the specific approach used to aggregate the positions related to each risk measure, in order to enable the board of directors and senior management to evaluate the results.

Supervisors expect banks to produce complete risk data and that measure and monitor the completeness of the data. The possible presence of data gaps risk should not affect the bank's ability to manage risks effectively. The supervisory authorities expect that the data of the banks are substantially complete, and that any exceptions are identified and motivated.

Principle 5 - Timeliness

"A bank should be able to generate aggregate and up-to-date risk data in a timely manner while also meeting the principles relating to accuracy and integrity,

completeness and adaptability. The precise timing will depend upon the nature and potential volatility of the risk being measured as well as its criticality to the overall risk profile of the bank. The precise timing will also depend on the bank-specific frequency requirements for risk management reporting, under both normal and stress/crisis situations, set based on the characteristics and overall risk profile of the bank."

The aggregation capabilities of a bank's risk data should be such to ensure that it can produce aggregate risk information in a short time, as well to comply with all the requirements for risk management reporting. The Basel Committee recognizes that the different types of data must be available at different times, depending on the type of risk, and that some data may be made needed more quickly in a situation of tension / crisis. Banks must structure theirs systems in such a way as to allow them to be produced in a situation of tension / crisis timely aggregate data on all fundamental risks.

Fundamental risk data include:

- aggregate credit exposure to an important corporate customer. By way of comparison, retail exposure groups may not vary in such a way just as relevant in a short time, but could still present considerable concentrations;
- counterparty credit risk exposures, for example for transactions in derivatives;
- exposures, positions, operating limits and market concentrations by sector and by regions in relation to trading activity;
- liquidity risk indicators, such as cash flows / regulations and funding;
- urgent indicators of operational risk (e.g. availability of systems, unauthorized access).

Supervisors will verify that the requirements of individual banks in terms of periodicity, both for normal situations and for cases of tension / crisis, allow for generate in a short time aggregated and updated risk data.

Principle 6 - Adaptability

"A bank should be able to generate aggregate risk data to meet a broad range of on-demand, ad hoc risk management reporting requests, including requests during

stress/crisis situations, requests due to changing internal needs and requests to meet supervisory queries."

The aggregation capabilities of banks' risk data should be flexible and adaptable to meet ad hoc requests for information and for evaluate emerging risks. This adaptability will allow banks to improve risk management, in particular regarding forecast data, as well as facilitating conducting stress tests and scenario analysis.

Adaptability presupposes:

- flexible aggregation processes that allow aggregation of risk data for make assessments and take decisions quickly;
- the possibility of adapting data to the needs of users (for example, creating dashboard, or highlighting the salient points and anomalies), get the degree of researched detail and quickly produce summary reports;
- the ability to integrate new developments related to business organization and / or factors external factors that influence the bank's risk profile;
- the ability to integrate changes to the regulatory framework.

Supervisors expect banks to be able to generate subsets of data based on the required scenarios or following particular events economic. For example, banks should be able to quickly aggregate risk data on credit exposures by countries on a specific date for a list of countries, as well as data on credit exposures by economic sectors at a specific date for a list of sectors, and this for all operational lines and geographical areas.

2.1.3 Risk reporting practices

Accurate, complete and timely data are the foundation of a management effective risk. However, they do not guarantee on their own that the board of directors and senior management receive adequate information to make effective decisions regarding risk. For the purpose of proper risk management, it is necessary that the appropriate information be presented to the right people at the right time. Reporting based on data from risk should be accurate, clear and complete. Its content should be corrected and presented to the appropriate decision-making bodies in time to allow one adequate response. To be effective, risk reports should meet the principles shown below.

Principle 7 - Accuracy

"Risk management reports should accurately and precisely convey aggregated risk data and reflect risk in an exact manner. Reports should be reconciled and validated."

Risk reporting should be accurate, to ensure that the board of directors and senior management can rely on the information aggregated to take relevant risk decisions. In order to ensure the accuracy of reporting, banks should equip themselves at least with defined requirements and processes for linking reporting with risk data; verification of automatic and manual changes and checks of reasonableness, including an inventory of validation rules applied to information quantitative (the inventory should include the explanation of the conventions used to describe any logical or mathematical relationships to be submitted to verification through validations and checks); integrated procedures for identifying, reporting and explaining errors or shortcomings in the integrity of data through the reporting of anomalous values (exception report).

The approximations are an integral part of reporting and management of risks. The results obtained from models, scenario analysis and stress tests are examples of approximations that provide fundamental information for risk management. Although expectations regarding approximations may differ from those relating to other types of reporting, banks should follow the principles of reporting and establish expectations in terms of reliability approximations (accuracy, timeliness, etc.) to ensure that the direction can rely on the information received to make relevant decisions on the subject of risk. These expectations should also relate to the data used to obtain the approximations.

Supervisors expect that the bank's top management establish accuracy and precision requirements for both routine and in-kind reporting tension / crisis situation, including information on positions and exposures criticism. These requirements should be commensurate with the importance of the decisions you make will base on the information in question. The supervisory authorities expect that, for the purposes of accuracy requirements, the banks adopt an approach similar to that of accounting relevance. For example, an omission or an inaccuracy that could influence them could be considered relevant risk decisions made by data users. Banks should be in able to motivate accuracy requirements. Supervisors expect that the banks consider the possibility of

applying precision requirements according to the processes and the results of validations, verifications and linking operations.

Principle 8 - Comprehensiveness

"Risk management reports should cover all material risk areas within the organisation. The depth and scope of these reports should be consistent with the size and complexity of the bank's operations and risk profile, as well as the requirements of the recipients."

Risk reporting should report information on exposures and events positions relative to all significant risk areas (for example, credit risk, risks of market, liquidity risk, operational risk) and to all their significant components (for example, single borrower, country and business sector with regard to the risk of credit). It should also contain information on the magnitudes related to the risk (for example, regulatory capital and economic capital). Reporting should detect emerging risk concentrations, provide information in the context of limits and risk appetite / tolerance and, where appropriate, make recommendations for possible interventions. It should describe the status of implementation of the measures approved by the board of directors or senior management for reduce risk or address specific risk situations. Reporting should allow monitoring of emerging trends through forecasts and stress tests perspective. Supervisors expect banks to determine, in matters of reporting, the requirements that best suit their operating models and risk profiles. The supervisory authorities will have to consider satisfactory the choices made by the banks in terms of risk coverage, analysis and interpretation, scalability and comparability between entities of the banking group. For example, a report on aggregate risk data should contain at least the following information: capital adequacy, regulatory capital, projections on capital and liquidity ratios, credit risk, market risks, risk operational, liquidity risk, stress test results, concentrations within one same risk category or between different categories, positions and financing plans. Supervisors expect that risk reporting to the Board of Directors and senior management provides a forward assessment of risks and is not based only on current and past data. It should contain forecasts or scenarios

related to the main market variables and the effects on the bank, so inform the board of directors and the top management of the likely future performance of the bank's capitalization and risk profile.

Principle 9 - Clarity and usefulness

"Risk management reports should communicate information in a clear and concise manner. Reports should be easy to understand yet comprehensive enough to facilitate informed decision-making. Reports should include meaningful information tailored to the needs of the recipients."

Risk reporting should promote proper risk management and facilitate the decision-making process of recipients, in particular the board of directors and the high-level direction. It should ensure that information is relevant and adapted to need for the recipients.

Reporting should present the right balance between risk data, analysis and interpretations, and qualitative explanations. The balance between quantitative and qualitative information will vary according to the levels of the organization, as well as the degree of aggregation applied. At the highest levels of the organization, aggregation will presumably be greater, and yes it will therefore require a greater degree of qualitative interpretation.

Reporting policies and procedures should incorporate different needs information from the board of directors, top management and other levels organization (for example, risk committees). The Board of Directors, being one of the main recipients of the reporting on risks, is responsible for defining its needs in this area and for observing the its obligations to shareholders and other interested parties. The board of directors should make sure to request and receive appropriate information, which will allow it to fulfill its mandate vis-à-vis the bank and the risks to which it is exposed. This will allow the board itself to ensure that its work is done in compliance with the limitations of tolerance / risk propensity set. The board of directors should notify senior management if the risk reporting does not meet the requirements defined by it and does not provide the level or type adequate information to determine the tolerance / risk propensity of the bank and check that it is respected. The board of directors should indicate if the reports received

are sufficiently accurate and have the right balance of information qualitative and quantitative. Top management is also among the main recipients of risk reporting and is responsible for defining their needs in this regard. It should make sure you receive appropriate information, enabling it to fulfill its mandate with respect to it of the bank and the risks to which it is exposed. Banks should draw up an inventory and classification of risk data that refer to the concepts used to process reporting.

Supervisors expect risk reporting to be clear and useful. Individual reports should present the right balance between detailed data, qualitative analyzes, interpretations and final recommendations. The interpretation and explanation of the data, including observed trends, should be clear. Furthermore, they expect banks to check periodically with the recipients that the aggregated and reported information is relevant and appropriate, in terms both quantitative and qualitative, for the purposes of corporate governance and decision-making.

Principle 10 - Frequency

"The board and senior management (or other recipients as appropriate) should set the frequency of risk management report production and distribution. Frequency requirements should reflect the needs of the recipients, the nature of the risk reported, and the speed, at which the risk can change, as well as the importance of reports in contributing to sound risk management and effective and efficient decision-making across the bank. The frequency of reports should be increased during times of stress/crisis."

The frequency of the reports will vary according to the type of risk, of the objectives and recipients. Banks should periodically evaluate the goal of each type of report and establish the times within which the reports should be produced in situations normal and in the event of tension / crisis. Banks should check their own regularly ability to produce accurate reports on time, in particular in the situation of tension / crisis. Supervisors expect that during the periods of tension / crisis the whole relevant and relevant reporting on credit risk positions / exposures, of market and liquidity is available in a very short

time to react effectively to the evolution of risks. In this regard, some information on positions / specific exposures may become necessary immediately.

Principle 11 - Distribution

“Risk management reports should be distributed to the relevant parties while ensuring confidentiality is maintained.”

Procedures should be put in place to allow risk data are collected and analyzed quickly and the related reports are promptly received all recipients of the case. This need must be weighed against the need to guarantee appropriate confidentiality. Supervisors expect a bank to periodically confirm that the case recipients receive timely reports.

2.1.4 Supervisory review, tools and cooperation

Supervisors have an important role to play in encouraging and monitor the implementation and compliance with the Principles by banks. They must also verify compliance with the Principles with the various banks to ascertain that the principles they are producing the desired outcomes and assessing the opportunity for further improvements.

The following principles regards the supervisory bodies duties which are not the objective of this thesis and they will be only enunciated for completeness on the argument, without going deep on the topic.

Principle 12 - Review

"Supervisors should periodically review and evaluate a bank's compliance with the eleven Principles above."

Principle 13 - Remedial actions and supervisory measures

"Supervisors should have and use the appropriate tools and resources to require effective and timely remedial action by a bank to address deficiencies in its risk data aggregation capabilities and risk reporting practices. Supervisors should have the ability to use a range of tools, including Pillar 2."

Principle 14 - Home/host cooperation

"Supervisors should cooperate with relevant supervisors in other jurisdictions regarding the supervision and review of the Principles, and the implementation of any remedial action if necessary."

2.2 Circolare n. 285 of BANCA D'ITALIA

The data recording and reporting system is designed to promptly track all business operations and management events in order to provide complete and up-to-date information on company activities and the evolution of risks. Ensures continuous integrity, completeness and correctness of the archived data and the information represented; Moreover, guarantees accountability and easy verifiability (for example, from control functions) of recorded operations.

In particular, the data management system meets the following requirements:

- the registration of company facts is complete, correct and timely, in order to allow the reconstruction of the activity carried out;
- a corporate governance standard (which will be described in the next chapter and on which will be developed the data governance case) is defined that identifies roles and responsibilities of the functions involved in the use and processing, for operational and management purposes of the company information; in view of their relevance in the information system, measures to ensure and measure quality are defined, e.g. through the quality of the key indicators periodically referred to business users, control functions and to the organ with management function;
- identification, measurement or evaluation, monitoring, prevention or mitigation of risks associated with data quality is part of the management process risks; in case of acquisition or incorporation of external subjects, the due diligence includes the assessment of the impact of the operation on the management procedures and aggregation of data; the use of sectoral procedures (accounting, reports, anti-money laundering, etc.) does not compromise the overall quality and consistency of company data; at the consolidated level, the group system ensures the integration of information coming from all the members of the group;
- in case of recourse to a company data warehouse for analysis and reporting purposes, the procedures of extraction, transformation, control and loading of data in centralized archives - so as the functions of data exploitation - are documented in detail, in order to allow verification of data quality;
- data management and aggregation procedures are documented, with specific provisions the circumstances in which manual entry or correction of company data

is allowed, registration date, time, author and reason for intervention, the affected operating environment and data previous modification;

- data acquisition processes from external information providers are documented and manned;
- the data are stored with a granularity suitable to allow the different analyzes and aggregations required by the exploitation procedures;
- product reports show the main assumptions and estimation criteria adopted (for example, as part of the monitoring of business risks);
- the reporting system makes it possible to produce timely and high-quality information for the supervisory authority and the market.

CHAPTER 3

3. Case description

The case I am going to discuss and explore is one of the major Italian companies, rooted in the territory for decades and a point of reference for most Italians because it operates in different industrial sectors: finance and banking, mobile operating system, delivery services and insurance sector.

The company provide the so-called "universal service", i.e. it must provide some essential services at a controlled price and since 2015, the company has been listed on the stock exchange after 153 years since its foundation. In fact, starting from October 27, 2015, the public limited company, with a 60% stake in the Ministry of the Economy and Finance, is listed on the Milan Stock Exchange with a float of almost 40% and an initial capitalization of 8.816 billion euros. 70% of the shares sold belong to institutional investors while 30% were reserved for the general public residing in Italy and for employees residing in Italy.

With a coverage of almost 100% of the Italian territory, today the company is organized in 5 Divisions and 13 Directions. It has 132 branches, 12,000 offices, 16 mechanization centers, more than 2,000 delivery offices, more than 7,300 ATMs, 18 daily air connections, more than 33,000 vehicles for a total of more than 140,000 employees[7].

The business unit to which the data management model is applied is the banking business. Obviously, it is the most exposed branch in absolute and is the one that manages the riskiest products both for customers and for the company itself. This type of enterprise is a practical example to study because it starts from a very primordial state of Data Governance and we can follow all the initial steps, usually similar to all companies, of how this type of project born and developed. The business unit in object, is divided into 11 structures, each with a different task. Each of these structures will have to adopt a model of Data Governance (later we will see how the application of this model takes place, which structures need timely intervention and with which depth)

[7]Source: Companies websites

3.1 Pillars of Data Governance

Each data management model refers to the regulations and principles set out in the second chapter. Despite this, every bank needs a custom model that fits perfectly with the organizational and architectural structure of the enterprise. For this reason, every bank, before the beginning of the works, publishes a Data Governance Standard in which the modalities of application of the model to the enterprise are declined.

3.1.1 Objectives and scope of application

The object of a Governance Data Standard is therefore the definition of a data management system that, in compliance with the regulatory provisions, continuously pursues the completeness, correctness and timeliness of the data stored and the information represented.

The data management system is the model that establishes the rules, the actors, the responsibilities, the control models with which the information managed by the information system must be treated, throughout their life cycle, guaranteeing its accountability.

The Data Governance introduces the following principles, also taken from the legislation, that the company intends to make its own:

- the data must be reliable and assessable in terms of completeness, accuracy and timeliness;
- people must be responsible for data governance, for the prevention and management of problems that may occur on them, including loss of quality;
- data governance is an evolving process that is continuously improving within the company;
- technology is an indispensable support for ensuring the correct implementation of data governance processes.

The persons involved in the application of the Data Governance model must operate in compliance with the regulatory and organizational system and are required to operate in

compliance with the laws and regulations in force and in compliance with the principles described at the outset in chapter 2 and they are briefly summarized below:

1) **TRACEABILITY** - The persons involved in the data governance process must guarantee, each for the part of their competence, the traceability of the activities and documents related to the process, ensuring the identification and reconstruction of the sources, information elements and controls carried out at support of activities.
2) **SEGREGATION OF TASKS AND ACTIVITIES** - The process of data governance involves the segregation of tasks and responsibilities, between different organizational units or within them, in order to avoid that incompatible activities are concentrated under common responsibilities.
3) **COMPLIANCE WITH LAWS AND CONSISTENCY WITH THE GENERAL REGULATORY FRAMEWORK** - The data governance process is defined in compliance with applicable regulations, in line with the internal reference framework and national and international best practices.
4) **CONFIDENTIALITY** - Without prejudice to the transparency of the activities carried out and the information obligations imposed by the provisions in force, the persons working in the data governance process ensure the confidentiality required by the circumstances for each news item / information learned on the basis of their work function.
5) **CONFLICT OF INTEREST** - The people involved in the data governance process act towards their counterparts according to relationships marked by the highest levels of behavioral ethics, in compliance with the Code of Ethics, avoiding decisions and carrying out activities, in conflict, even if only potential with the interests of the Company or in any case contrary to its official duties.
6) **APPROACH BASED ON RISKS AND PROCESSES** - The process of data governance, inspired by a process logic, is based on a preventative approach to risks, contributing to the assumption of informed decisions, and, where possible, the translation of the main risks into opportunities.
7) **RESPONSIBILITY MANAGEMENT (ACCOUNTABLITY)** - Management, within the scope of the functions covered and in achieving the related objectives,

ensures the application of the data governance process for the activities of competence, actively participating in its operation.

8) **COMMUNICATION AND INFORMATION FLOWS** - The information necessary to fulfill its responsibilities, including those regarding data governance, is made available to every corporate body and structure.

3.1.2 Data management system

The data represent one of the main assets of the organizations and are considered the "fuel" of the engine for growth and innovation.

Hence the great attention to their management and their control and the drive towards the introduction in each company of a standard of Data Governance which mainly responds to two objectives: a regulatory one, to ensure compliance with the supervisory regulations and to the guidelines of reference dictated by national[8] and international[9] bodies and the other strategic, aimed at the use of data as a vehicle for the creation of value.

The Data Governance framework (exemplary model in the first chapter) that each company must propose should recall processes, rules and technologies that already permeate the entire company information system, with the aim of drawing a link between them and the ultimate purpose to provide an "identity card" of the data (Data Owner, traceability and quality level)

In order to build a reference model that summarizes the main domains that characterize the Data Governance, the company must draw on national and international frameworks

[8] **Bank of Italy Circular n. 285** of 17 December 2013 (Title IV, Chapter 4, Section V) - 11th update of 21 July 2015. (the data management system); Basel Committee on Banking Supervision 239 "Principles for effective risk data aggregation and risk reporting" (January 2013); Basel Committee on Banking Supervision 268 "Progress in adopting the principles for effective risk data aggregation and risk reporting" (December 2013).

[9] **GDPR** (General Data Protection Regulation) - The general regulation on data protection with which the European Commission intends to strengthen and make more uniform the protection of personal data of EU citizens (EU Regulation 2016/679, From 25 May 2018, the GDPR will replace the directive on data protection (officially Directive 95/46 / EC) [established in 1995 and will repeal the provisions of the Code for the protection of personal data (Legislative Decree No. 196/2003)

and experiences to measure the level of maturity on the Data Governance and define the related evolutionary roadmap according to the approach CMMI[10].

3.1.2.1 Methodological approach

The adoption of a Data Governance standard represents a step towards an increasingly detailed and widespread governance of the company's information assets. As regards the case dealt with in the next chapter, it envisages a modular "risk-based" approach with regard to the definition of the scope and the intensity of the preventive measures of data-related risks, which, through the role of Data Governance Office, will be subject to review, monitoring and continuous evolution. The Data Governance Office will identify, measure, monitor and prevent risks related to data quality.

The risks associated with data management can be mapped into the following main types:

1) Operational risk, which can be divided into:
    - Risk events that generate penalties - for example quality defects found on data intended for mandatory reporting purposes that provide for a sanction regime;
    - Risk events that generate losses other than sanctions - quality defects on those data whose erroneous management may involve any other type of operational loss (for example: wrong strategic decisions taken by the summit, judicial or extra judicial reimbursements, devaluations, etc.) .
2) Reputational risk - incorrect data management that can cause effects on the negative perception of the company image by customers, shareholders, supervisors and other stakeholders.

In general, the potential risk related to a data item is expressed by the product of the probability that it does not comply with the requirements defined for its use (for example, it has a higher percentage of anomalies than that considered acceptable) multiplied by the

[10] Capability Maturity Model Integration (CMMI) is a process improvement approach that aims to help an organization improve its performance.

impact of the damage that it follows that it is a function of its recipient of use (for example the payment of an administrative penalty for irregularities in reporting to regulators).

Data Governance is the tool to mitigate these risks, through a coordinated set of processes, people and technologies.

3.1.2.2 Perimeter of Data Governance

"Data" is defined as the information managed and processed by an IT tool that the company uses and manages for its strategic, reporting and operational purposes.

For each data it is necessary to identify two main aspects:

- its type (depending on its life cycle);
- the uses for which it was generated (recipient of use);

Data must be governed regardless of their source or destination[11]; the company must ensure that, data provided or disseminated in its own name, comply with the rules in force to guarantee regulatory compliance and to protect any consequences (economic, regulatory or reputational) of inadequate management.

However, in order to identify an intervention priority, it is essential to establish a relevant perimeter for data governance purposes, in the sense that the perimeter data will be assigned an entry order in the program and, if necessary, the appropriate corrective measures will be defined and adopted. the quality level should not be adequate.

Data that fall into at least one of the following two categories are defined as relevant:

- high risk use destination;
- type of data at risk.

The high-risk use destinations are as follows:

[11] Governance is therefore necessary: on the own data, that is data that are managed throughout the entire life cycle by the company and on all those acquired by third parties through applications or managed by third parties on behalf of the company, ensuring that even for these, suitable management and quality control processes are envisaged.

- reporting or information purposes for top management;
- periodic financial reporting purposes for shareholders and stakeholders in general;
- valuation purposes on company processes and systems for the Control Functions;
- reporting or information purposes to the Supervisory Bodies;

The main types of data at risk are the following:

- from external sources: the data provided by external information providers are potentially at risk as they are not directly subject to the company acquisition procedures and the related controls;
- subject to manual corrections: the data frequently corrected or manual processes on applications developed by end users ("end user computing" EUC) are to be included among the critical ones as automatic controls are not applied to them;
- intended for a variety of uses: the data stored in a data warehouse must be managed with particular care as they are subject to substantial transformation and standardization processes and destined for a variety of uses;
- from company acquisitions or mergers: in the case of data migration from a different IT system, data quality could be compromised by conversion processes. In this case, appropriate management and control measures must be provided for the information assets for the initial period of operation with the new system;
- managed by applications that are not consolidated or valued at risk: data managed by new applications, or those modified significantly, are more critical in terms of data quality than those that have been operating for a long time and management measures must therefore be envisaged and reinforced control. In addition, controls are to be strengthened for those applications which, following an IT risk analysis, present levels of risk above the acceptance threshold.

In determining the perimeter, an approach is therefore adopted, which enhancing what is already present in the company at the level of description of processes, applications, information and risks, will introduce the Services dimension and data which will be accompanied by an identity card compiled on the basis of their life cycle and integrated with the Key Quality Indicators that will be defined.

In particular, the implementation of the Data Governance process starts from a functional perimeter, defined by company to company, called "priority", in which services and data

will be collected from the relevant Data Owners with the objective criteria described above.

For all relevant data, specific measures to mitigate the risk of alteration of their quality are foreseen: the whole information generation and transformation chain is governed, starting from the data or from the source data.

For the remaining data considered as non-critical, the level of original risk is accepted without specific mitigation (also called “inherent risk”).

3.1.2.3 Data Life Cycle

Effective data governance involves managing data throughout the relevant life cycle in order to support the achievement of business objectives and mitigate the risks that may arise from a lack of adequate controls in business processes.

The term “Data Life cycle “identifies the set of all the phases that describe the existence of a data within the company’s information assets, from its creation to its elimination. The main phases that make up the life cycle of a data are summarized below:

1) **Creation / Classification**

Data can be collected through various channels: by mail, by partners, or through IT applications used by internal or external staff (for example by customers through mobile APP applications).

With respect to the computer applications with which they are processed, it must be established whether this is the first time that data is created (determination of the “master” source) or is transmitted by another internal application to its own domain. In particular, the recording of company facts must be complete, correct and timely, in order to allow the reconstruction of the activity carried out and with a level of retention aligned with the conservation policies. The information must be classified appropriately, allowing the application of the most appropriate security measures and privacy guarantees.

2) **Access / Use**

Information is frequently the subject of processing or use by multiple applications and persons, including subjects operating outside the company function in question. The company must ensure that only authorized users can access the processed data, with the appropriate access profiles defined according to the current guidelines and procedures, and must impose strict conditions for the transfer of data on Office tools or devices outside the function.

3) **Update / Transformation**

Data is typically updated several times during their life cycle. The criteria for transformation, aggregation, transfer, distribution and use of data must be adequately documented and managed safely. At this stage there are multiple problems that can undermine data integrity: repeated updates (manual or automatic, through batch procedures), human errors, and malicious activity can compromise the integrity and accuracy of information.

4) **Cancellation**

A policy of deleting data from archives must also be defined, which may also consist in keeping data in the archive indefinitely, in compliance with regulatory or legal obligations, and / or deleted if expressly required.

3.1.2.4 Data Quality

To ensure efficient data management throughout their life cycle, indicators are introduced / census to ensure and measure quality, the Key Quality Indicators (KQI).

The Key Quality indicators aim to measure the qualitatively most relevant aspects with respect to data management. They also make it possible to extend the control domain to all phases of the life cycle by relating the results of the checks carried out in several phases / processes.

These indicators measure the quality level according to the three main directions:

1) completeness;
2) accuracy and integrity;
3) timeliness;

Each indicator is grouped into specific control families that guarantee quality levels according to the guidelines described. For each family, target values and tolerance thresholds are defined, and a re-entry plan will be established in cases where a satisfactory evaluation is not achieved.

The measurement of the defined Key Quality indicators will be periodically reported, with periodicity and level of detail, different to the individual Data Owner, the control functions, the Company Manager and the management body.

The Key Quality Indicators will be synthesized on a dashboard (Data Quality Dashboard) whose consultation by the various stakeholders will promote an awareness on the health status of the data of its information assets and in general on the progress of the governance program.

3.1.2.5 Roles and responsibilities: Data Governance office, Data Owner and User

Effective Data Governance requires an appropriate governance and management model, with well-defined roles and responsibilities, sufficient resources to perform the required tasks and clear guidelines on both the general and specific objectives.

The subjects / structures involved in the Data Governance process hold specific responsibilities in this area, in line with the company's overall organizational and governance structure.

The roles and responsibilities defined within the overall Data Governance process are as basically three: Data Governance office, Data Owner and Data User.

Data Governance Office

The Data Governance Office is the center of competence of Data Governance for the company, with the aim of supporting the various processes of Data Governance, by virtue of specific expertise on the subject. It is entrusted with tasks of direction and control and operational tasks in collaboration with the other roles involved in the Data Governance.

Among the main activities and responsibilities of Data Governance we find:

- proposal to define and update the Data Governance Standard;
- update of the scope of the Data Governance and the related Data Owner and data user;
- detection and updating of the Data Governance metrics;
- measurement of the level of maturity achieved compared to the standard;
- proposal for the definition and updating of the procedures and tools of Data Governance;
- preparation of the Data Governance Report;
- support to the Data Owner.

The Data Governance Office coordinates with the structures within the company for the updates of specifications and the interface to IT.

Data owner

The role of Data Owner is assigned on the basis of functional responsibilities that perimeter a set of services / applications and therefore data.

The Data Owners are identified in the managers of the company functions involved in the Data Governance process, on the basis of the identified perimeter, as described above. The Data Owner is a figure who is given the Accountability of a given datum as a connoisseur of his life cycle. The assessment activities allow each Data Owner to identify the main data managed, with a known life cycle and for the use destinations for which they are responsible.

The responsibility of the Data Owner is therefore related to his ability to assess the risks related to the specific use of data and the ability to activate / request specific safeguards

by competent figures (remediation action proposal).The Data Owner has the function of guaranteeing the quality of the data (through the tools provided by the Data Governance system) of which it is the end user and this responsibility remains unchanged regardless of whether the same data, in whole or in part, is produced and / or updated by other actors, internal and / or external to the company function. The Data Owner activates the competent functions for the technical and process implementations aimed at guaranteeing the quality of the data for which it is responsible, where, in the face of a specific recipient of use, criticalities should be highlighted.

The Data Owner is responsible for the following activities:

- identification of relevant data in the specific area of competence;
- definition of the necessary controls for data control (with support of the Data Governance Office);
- KQI monitoring (with support of the Data Governance Office);
- reporting of anomalies found, analysis of the same and follow-up of remediation activities.
- proposition and sharing with the Data Governance Office of the interventions necessary for the continuous improvement of data quality in its field of competence, involving, where necessary, the internal and external functions of the company.
- definition of the needs and prioritization of the evolutionary interventions necessary to align with the data governance standard in order to respect the quality levels (with support of the Data Governance Office).

Data User

The Data User is a figure with in-depth professional skills related to a specific business operation of which he knows the functional logic and the first level interrelations with the other related areas; in particular, it has a specific knowledge of the value of data and their treatment in this operating context. He is therefore the primary interlocutor of the Data Owner and, by virtue of his specific skills on the subject he oversees, he is directly responsible for the data quality control process.

In fact, consistently with the addresses expressed by the same Data Owner, it supports the work and is responsible for the following activities:

- defines, also upon specific request of the Data Owner, the controls necessary to ensure the objectives of Data Governance and data quality, with reference to that specific process supported by the business application for which it is responsible;
- collaborates with the Data Owner in determining the Key Quality Indicators and participates in the execution of controls and measurements of the KQIs themselves;
- intervenes in determining the causes of deviations and anomalies found and supports the competent functions / performs the necessary corrections. Generally, the Data Users are identified in the operating personnel employed by the Data Owner.

CHAPTER 4

4. Implementation of the data governance project

The implementation modalities of a data governance project change from company to company. In most cases the planning, approach and tools used vary depending on the starting level of a company. Just think of a company that, on the IT side, already has many applications for calculating and managing data, and a company that is instead in a "primordial" state and data processing is almost completely done manually with the help of those that we call EUC tools "End User Computing" (Excel, Access, PowerPoint etc ...). In the first case the data have already been mapped and inserted into applications and the Data Governance intervenes in a very advanced state and therefore customized from company to company. In the second case, however, the data and their management are still fully managed by the individual operator and exposed to all the risks involved; in this case, Data Governance must intervene from the foundations and the process of mapping and securing data is almost the same for all companies (except for possible customizations due to the company structure)

The case study we are about to analyze has the objective of creating a general idea of implementation of a Data Governance project and it is precisely to study a case in which a data management model is almost completely absent.

This case allows us to go and study the fundamental phases of defining and securing data within a company.

Before going into the case study, it is good to define the standard path of a given and its purposes.

Usually, within a company, data can be taken from business systems, from external providers or created from scratch.

Individual data is rarely useful for business purposes. Usually the creation of a single useful data derives from a large number of data passing through a transformation process (they are aggregated, disaggregated, transferred from one system to another, etc.) which

allows the “refining” of the same and the production of data useful for business purposes. After its creation, the data, depending on its type, is used for different purposes that can be used internally in the company or external use (analysts or supervisory bodies). The means by which data circulate inside or outside the company is called “Report”. Usually a large company produces several hundred reports that contain dozens of data inside them; they, depending on their relevance, are produced daily, weekly, monthly, etc.

Defined that, it is possible to go on and enunciate the steps in detail that have to be followed in the implementation of a Data Governance model:

1) **Assessment and project masterplan**: Analysis of information produced within the company and data from external sources. This phase consists of several processes ranging from the census of reporting to the prioritization of the areas of intervention.
2) **Information prioritization**: At the end of the assessment phase, starting from the intervention priority area, it is possible to create a ranking of the individual information contained in the produced reports. For each information we proceed to identify all its components and to trace the relative life cycle including the processes of control and transformation.
3) **Data lineage and business glossary**: The previously prioritized information must be analyzed individually in all its parts, highlighting, by levels, every single decomposition, transformation, control (which take place during production and therefore composition of the final data) and support systems. After that, the next step is the definition of each individual decomposition, transformation, control and support systems identified in the previous chapter and define a dictionary of data that allows anyone in the company to identify a meaning univocal for every information.
4) **Definition and calculation KQI**: implementation of indicators that are able to assess the quality of the data produced through its monitoring throughout the life cycle, between transformations and controls.
5) **Gap analysis and remediation plan**: identified the KQIs, it is possible to understand all the shortcomings in the data crunch process and to intervene with modifications on the process in order to reduce the gap between the as-is quality of the data and the to-be quality initially expected

4.1 Assessment and project masterplan

In the very first phase of the project, before starting the actual assessment, it is essential to identify the organizational areas responsible for producing the report, the vehicle through which data transmission takes place. These organizational areas are those on which it will be necessary to intervene.

The assessment phase aims to create a complete picture of the relevant information managed within the competence of the Data Owners of a company (remember that the data owners are usually the managers of the structures, those who have full responsibility of the data produced by their structures), as well as tracing the means by which this information is transmitted / communicated inside or outside the company (eg report, information flow, printouts), which we have identified as “Report”.

The final aim of the assessment phase is to obtain a ranking of the company organizational areas based on two main drivers: **the relevance of the product reports (recipient of use)** and **their management methods**.

The assessment phase consists of a reporting census phase whereby all the reports of interest can be defined, their content and above all their recipient of use, and an assessment phase (through a questionnaire administered to each date User) of the report management mode (production mode and control on the same in order to understand superficially the quality of the data). After collecting all the censuses and questionnaires, the aggregated information is aggregated and the reports and the structures are prioritized in such a way as to create an intervention plan based on the urgency of the intervention found.

Once the organizational areas of reference have been identified, the assessment process begins with the census of reporting.

4.1.1 Census of the Reports

The final objective of the census is to identify the information contained within each individual report.

The identification of the reports is entrusted to the individual Data Owner, the manager of the data produced and / or transformed and shared outside the single organizational area.

The census takes place by collecting the following information summarized is the table 4.1[12].

CENSUS SHEET								
Name of the Report	Description	Data conveyed	Organizational unit (Data Owner)	Recipient of use	Data source	Control over the components	Owner of the control	Periodicity
Report 01	Used to communicate the leverage ratio to the supervisory body	β,α,γ	Risk management	Regulatory	Data Warehouse	Manual	Risk management	Monthly
Report 02	Used to communicate internally the budget estimate of investments	A,λ	Informative system dep.	Board	Internal System X	Automated	Informative system dep.	Annually
Report 03	Used to communicate...	Θ,α,ι,ζ	Sales department	Operational unit	Computed Manually (EUC)	EUC	Sales department	Quarterly
Report 04	Used to communicate...	E,δ,γ	Department X	Directional	External provider	Automated	Department X	Daily
Report 05	Used to communicate...	Θ,β,λ,ι	Department Y	Market	External provider	EUC	Department Y	Weekly
Report 06	Used to communicate...	H,α,κ,ζ,δ	Department Z	Regulatory	Data Warehouse	EUC	Department Z	Weekly
Report 07	Used to communicate...	B,α,λ,ε,γ	Department W	Operational unit	Internal System Y	Manual	Department W	Monthly
Report X	Used to communicate...	Θ,λ,ζ,ε,δ	Department H	Board	Internal System Z	EUC	Department H	Quarterly

Table 4.1: Census sheet. Source: self-elaboration

The census sheet collects:

- **Report title**: name given to a report, recognizable inside and outside the company (Report 01, Report 02, Report 03, etc ...)
- **Report description**: Brief summary of the information conveyed by the report
- **Relevant indicators conveyed by the report**: list of all the relevant information used for the analysis object of the report (eg Report 01 convey the following data: β, α, γ)

[12] Representation of a typical census sheet used to collect all the fundamental information of the report.

- **Organizational level which produces the report**: the organizational unit in the company that produces the report, compute the data in it and due to that is responsible of the report (Data Owner)
- **Recipient of use**: Indicates the type of end user. This is one of the main drivers of our analysis because thanks to it we can determine the importance of the report. We will use it combined with the results of the questionnaires to determine the ranking of intervention.

  The type of end users can be declined as follows:

  - **Operational unit**: Operational Analysts within the company
  - **Directional**: Managers of the company
  - **Board**: CEO and Board of Directors of the company
  - **Market**: Financial analysts external to the company
  - **Regulatory**: Supervisory bodies
- **Data source**: Indicates the source(s) from which the information contained in the report comes from (Internal systems, external providers, internal computing and so on…)
- **Control on components**: Indicate the ways in which checks are carried out on the information present in the report (for example: EUC, manual controls or automated)
- **Owner controls**: Indicates the organizational area to which the responsibility for controls is assigned
- **Frequency**: Indicates the frequency with which the report is produced (daily, weekly, monthly, quarterly, half-yearly or yearly)

Once the census phase is completed, which allows a complete picture of the reports to be obtained, it is possible to proceed with the questionnaire.

4.1.2 Questionnaire on each report

The questionnaire consists of two types of questions which combination will allow us to define the **level of risk** of the report. The latter is the second fundamental driver that will be combined with the recipient of use for the purpose of creating an intervention ranking. In particular, it contains 5 questions on **monitoring the type of data** and 4 questions on **management maturity**. To each question, based on the answer given by the data owner, a score between 1 and 5 is associated.

Through a process of aggregating individual scores, it is possible to quantify the level of risk of the single report that can take a value between 1 and 5, in particular the higher the score, the better the monitoring of the risk level of the report.

The aggregation process, which takes into account a discount rate for certain values that have reported a good score compared to the average, takes place as follows:

- If the average of the management maturity score is less than 3, the risk level will be equal to the average of the score of the type of data;
- If the average of the management maturity score is greater than 3 and less than 5, the risk level will be equal to the average of the score of the type of data increased by ONE unit;
- If the average of the management maturity score is equal to 5, the risk level will be equal to the average of the score of the type of data increased by TWO units.

As we said, the questionnaire is composed by two parts which are the following.

Monitoring the type of data

The monitoring of the type of data investigates in the following areas:

1) Data origin

In this context, we investigate the provision of data used to compile the reports. The data are potentially at risk when, the components used for their creation, differ in the type of source information feed (e.g. more internal sources for the feeding of the same indicator

/ data, presence of external sources ...) in how much they may not be directly subjected to compliant business acquisition procedures and related controls.

2) Data treatment

We investigate the transformation of data. The data frequently subjected to rectification or manual processing on applications developed by end users (end user computing "EUC") are considered critical because they are not applied to automatic controls.

3) Data used for multiple purposes

We investigate the data stored in common / transversal databases that must be handled with particular care as they are subject to substantial transformation and standardization processes and intended for a variety of uses.

4) Data transfer

Investigate the migration of data from one computer system to another. such migration could compromise the quality of the data as a result of the conversion processes that take place during the transfer.

5) Data management application

We investigate the sources of data managed by new applications, or changed significantly. These sources are more critical with regard to the quality of the data compared to those that have been operating more than once, and therefore stronger management and control measures must be envisaged.

Management maturity

The management maturity of data investigates in the following areas:

1) Data governance

Investigation of data protection policies. The data used to compile the reports must be overseen by policies defined and shared with the entire company. Within the policies must be defined in particular the ownership of those who use and oversee the data used.

2) Data life cycle

The life cycle of the individual data used to compile the report is investigated. All data must be clearly mapped, recognized and shared by all users.

3) Data Quality:

We investigate the checks and the quality of the data. The data used to compile the reports must undergo a periodic monitoring activity that covers the entire life cycle of the data. The data quality assessment metrics must be clearly identified and updated periodically.

4) Data usage & data architecture

We investigate the data architecture that must be integrated for all types and for all functions. The architecture must be aligned with the different principles of data quality, scalable and flexible to adapt to changing regulatory needs, so as to provide timely information in real time and without losing reliability.

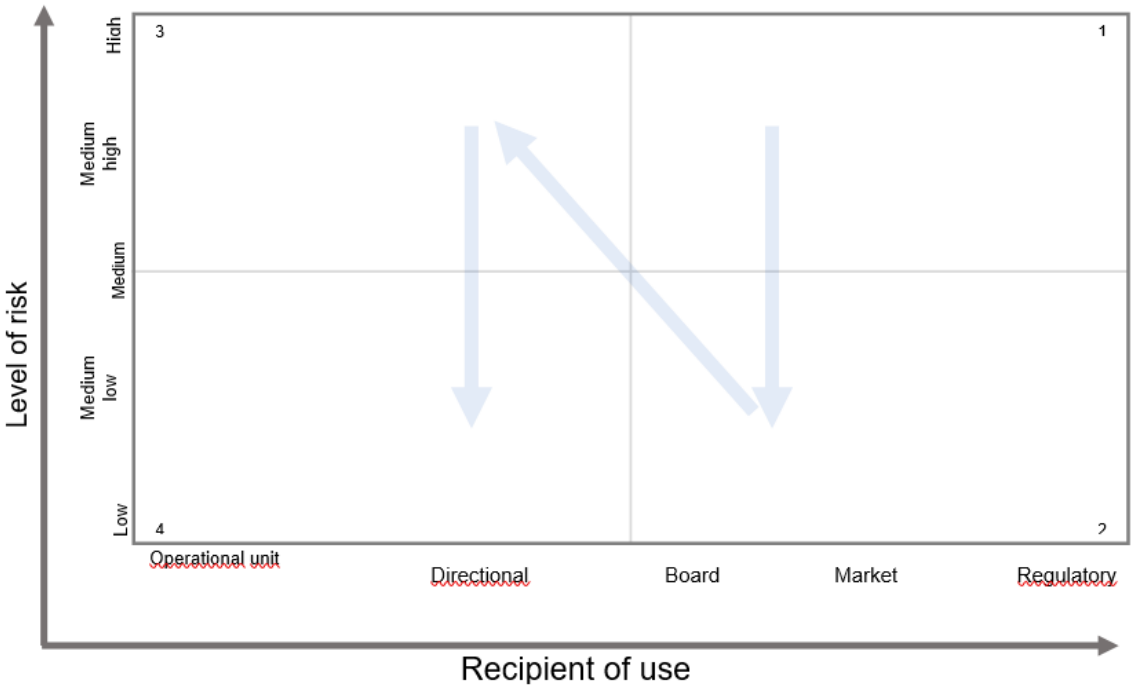
With the completion of all the questionnaires, the data collection phase is completed and it is possible to proceed with the analysis aimed at identifying an intervention ranking.

4.1.3 Prioritization matrix and masterplan

The prioritization matrix is a tool for the synthesis of all the information gathered. It builds a ranking of the organizational areas in such a way as to define which of these needs an immediate intervention.

The following dimensions (which are the main driver described previously) are represented on the matrix (Graph 4.1):

- Recipient of use: Defines the relevance of the final user of the report. The more the end user is relevant, the greater the priority assigned to him.
- Report risk level: that is the aggregate measure of the score of the type of data and management of maturity. The score assigned to a report is negatively correlated with the degree of priority. In particular, with the same usage destination, a report with a score of protection of the lowest level of risk requires a higher priority of intervention.



Graph 4.1: Prioritization matrix. Source: self-elaboration

Each quadrant is associated with a different degree of intervention priority:

- 1 ° Quadrant: All the reports placed inside see as the last recipient the following subjects: the top management of the company (Board) and figures outside the company (Market, Regulatory) for which a high level was found risk (interval [Medium, High]). These reports need the highest intervention priority

- 2 ° Quadrant: All the reports placed inside see the following subjects as the final recipient: the top management of the company (Board) and figures outside the company (Market, Regulatory) for which a low level was found risk (interval [low, medium). These reports require a high priority of intervention

- 3 ° Quadrant: All the reports placed inside see as last recipient a subject within the company (Operational unit, Directional) for which a high level of risk was found (range [medium, high]). These reports require a medium priority of intervention

- 4th Quadrant: All the reports placed inside see the last recipient as a subject within the company (Operational unit, Directional) for which a low level of risk was found (range [low, medium]). These reports require a low priority of intervention

Once all the reports have been positioned on the prioritization matrix, it is possible to understand which are the most critical reports both from the management point of view and from the point of view of importance (end user).

The ultimate aim of this study was to create a ranking of the company structures based on the priority of intervention, so it will be essential to average the risk and recipient of use of each report for each structure. This calculation will allow us to identify precisely which structure produces critical reports with respect to management and which have a high importance and gradually all the other less exposed structures.

The identification of this ranking is fundamental in the planning of a Data Governance project. In fact, this project requires an important effort from every point of view (costs,

personnel and so on) and it is not possible to apply it to all the structures at the same time, but it is necessary to implement structure by structure (each structure will have what we will call its own analysis “wave”). For this reason, thanks to the ranking identified (in addition to the census that allowed us to understand the amount of components that will be improved), it is possible build a plan in terms of time, costs, personnel employed and applications to be used / integrated.

4.2 Information prioritization

At this point in the project in which we collected all the information regarding the various structures, we move to the individual wave, that is the structure-by-structure analysis in order to understand the life cycle of each data, the transformations it undergoes and the controls on it (the ultimate aim of this analysis is to determine the Key Quality Indicators).

The prioritization phase of the information (data) present in the reports of an organizational area has the objective of identifying the most relevant information within an organizational unit. Every information used is not unique to a single report, but the same information can be used multiple times in multiple contexts. For this reason, a prioritization mechanism must be activated that indicates an information ranking within an organizational area through the use of two basic criteria:

- The frequency of use, i.e. the number of reports containing the specific information;
- The recipient of use of reports in which the information is reported.

The ranking of the relevant information prioritization is carried out through a descending order of the individual information recorded on the basis of a weighted average number of reports within which the individual information is present with different weights depending on the recipient of use of the reports.

4.3 Data Lineage

In the lineage phase, each single data, according to the order established in the prioritization phase, must be analyzed individually through a process of logical lineage, which allows to highlight the entire process of building information.

The lineage can be defined as a synoptic map of the data path in the reporting chain (Data lifecycle), including ownership and tracking of all the checks performed in the steps within the company information systems.

The term Life cycle of a data identifies the set of all the phases that describe the existence of a data within the company, from its creation to its elimination / archiving.

The data lineage occurs through the use of three main tools:

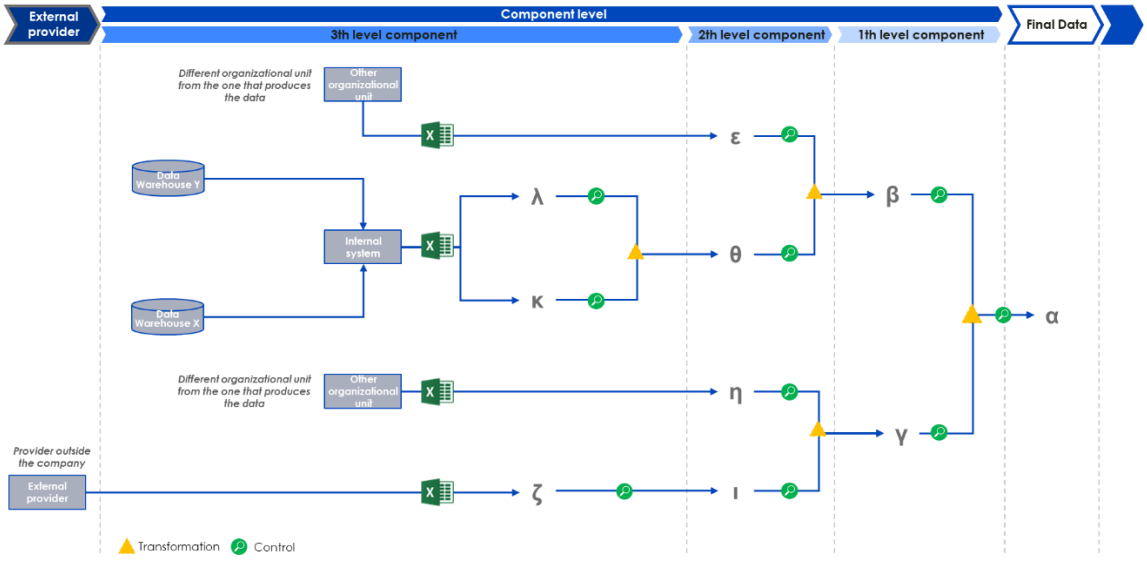
- **The logical path of data:** In order to correctly represent the entire logical line, the data must be decomposed graphically into all its components, highlighting, by levels, each single decomposition, transformation, control and support systems.
- **Data registry:** The objective of this analysis is to define every single control, transformation, system and organization that through which the data pass through during the composition of the final information
- **Business glossary:** The Business Glossary serves for the cataloging of logical data and related metadata[13] (the one that in the previous chapters we have called "identity card" of the data), describes at the level of functional semantics the information of the company, through the "Logical Map of Information ", Integrating it with additional notions relevant for proper management of the same (i.e. information object, data owner, etc.).

[13] The metadata is the identity card of the data. They are all the data we previously collected through the logical path of data and the registries.

4.3.1 The logical path of data

In order to correctly represent the entire logical line, the data must be decomposed into all its components, highlighting, by levels, each single decomposition, transformation, control and support systems.

Before explaining in detail the composition of the logical data path, a typical graph is shown below, thanks to which the graphic representation of this logical path takes place (Graph 4.2).



Graph 4.2: The logical path of data. Source: self-elaboration

Thanks to this type of representation it is possible to graphically trace the data path from the source systems up to its final shape after a process of transformations and controls.

In detail, the graph 4.2 shows:

1) **Composition levels**: indicate the degree of depth of the analysed component, in particular:
    - The final information object of study is positioned on the first level of lineage;
    - The progressive decompositions of the information, starting from the final indicator, determine the successive levels of lineage. The number of decomposition levels is determined by the complexity of the final data. In

fact, some data are the result of the aggregation of few components and it may be sufficient to stop at the second level, while others may be the result of complex transformations due to the aggregation of many components making it necessary to use more than 3 levels of decomposition;

- The last level of lineage, which is obtained at the end of the decomposition process, includes the systems of origin in which the "birth" of the data takes place and the information sources aimed at acquiring the elementary data.

2) **Controls on components**: both along the path and at each level of decomposition it is necessary to indicate any type of control that is carried out on the data. The type of control must be indicated in all its forms afterwards, in the analysis of the registry.
3) **Transformation of components**: whenever a datum, at each level of decomposition, changes its nature undergoing a transformation, it is necessary to indicate the moment in which it occurs. The type of transformation must be indicated in all its forms afterwards, in the analysis of the registry.
4) **Systems**: it is necessary to indicate all the systems on which the information:
    - Born
    - Transit
    - They are subject to processing
    - They are archived / deleted

4.3.2 Data Registry

At the end of the complete mapping of the information, we proceed with the detailed analysis of the actions carried out on all its components, various levels of logical lineage, through the registry.

The registry is divided into 4 basic sections:

- **Registry of controls**
- **Registry of transformations**
- **Registry of systems**
- **Registry of organizational units**

4.3.2.1 Registry of controls

The objective of this analysis is to define every single control that occurs during the composition of the final information.

Before explaining the composition of a registry of controls, a typical table is shown below (Table 4.2), thanks to which collect in detail all the controls highlighted in the logical path of data (Graph 4.2)

REGISTRY OF CONTROLS								
Final data	Component	Description of the control	Type of control	Methods of execution	Frequency of execution	Blocking control	Responsible structure of the control	Responsible structure of the remediation
α	β	Checks of the trend of the data respect to the previous year	Trend	Manual	Annually	No	Department X	Department X
α	η	Checks of completeness of the data received	Completeness	Manual	Quarterly	Yes	Department X	Department X
α	ι	Data User checks if the data is updated	Updating	Automated	Quarterly	Yes	Department X	Department X
α	ε	Data user checks...	Accuracy	Manual	Quarterly	Yes	Department X	Department X
α	γ	Data user checks...	Timeliness	Manual	Quarterly	No	Department X	Department X
α	θ	Data user checks...	Uniqueness	Automated	Monthly	No	Department X	Department X
α	λ	Data user checks...	Integrity	Manual	Monthly	No	Department X	Department X
α	ζ	Data user checks...	Consistency	Automated	Monthly	Yes	Department X	Department X

Table 4.2: Registry of controls. Source: self-elaboration

For each control the following dimensions are defined:

- **Final data**: Represents the main indicator to which the control refers;
- **Component**: Indicates the exact data, along the entire lineage of the final data, to which the control refers
- **Description of the control**: Contains a brief summary of the operation of the control;
- **Type of control**: indicates the type of control[14] chosen between the following:
    - **Completeness**: Controls aimed at verifying the presence of information (e.g. that a field within the application or excel sheet is always valued);
    - **Accuracy**: Controls aimed at verifying the accuracy of the data entered in the field (e.g. that a field in which there must be a date is filled in with the correct format);
    - **Consistency**: Controls aimed at checking the consistency between two or more fields (cross field check) or reconciliation checks;
    - **Timeliness**: Controls aimed at verifying that the information is actually usable when it is specifically requested;
    - **Univocity**: Controls aimed at verifying that the information is unique and not duplicated;
    - **Integrity**: Controls aimed at verifying the integrity of information in the steps between different applications;
    - **Trend**: Controls aimed at verifying any deviations between the observations in different reference periods;
    - **Update**: Checks to verify that the information is updated;
- **Methods of execution**: Indicates whether the control is automated or if it is carried out manually;
- **Frequency of execution**: Indicates the periodicity of execution of the control (daily, weekly, monthly, quarterly, half-yearly or yearly);

[14] The typical types of controls listed, have been studied in order to ensure full compliance of the data with the reference legislation analyzed in Chapter 2 "BCBS 239, January 2013 (Basel Committee on Banking Supervision) of the Bank for International Settlements)"

- **Blocking control**: indicates, with a yes or a no, if such control, if it detects an anomaly, would cause the interruption of production of the final data or not;
- **Responsible structure of the control**: Indicates the organizational area responsible for the control;
- **Responsible structure of the remediation**: Indicates the structure responsible for the remediation action;

4.3.2.2 Registry of transformations

The objective of this analysis is to define every single transformation that occurs during the composition of the final data.

Before explaining the composition of a registry of transformation, a typical table is shown below (Table 4.3), thanks to which collect in detail all the transformation highlighted in the logical path of data (Graph 4.2)

REGISTRY OF TRANSFORMATION					
Final data	Component	Description of the transformation	Type of transformation	Methods of execution	Responsible structure of the transformation
α	β	Sum between ε and θ	Aggregation	Manual	Department X
α	θ	Ratio between λ and κ	Aggregation	Manual	Department X
α	γ	Subtraction between η and ι	Disaggregation	Automated	Department X
α	ε	Mapping between...	Mapping	Manual	Department X
α	ι	Shortcut of ...	Shortcut	Manual	Department X
α	η	Manual adjustment because...	Manual adjustment	Automated	Department X
α	λ	Computing of...	Computing	Manual	Department X
α	ζ	Ration between...	Ratio	Automated	Department X

Table 4.3: Registry of transformation. Source: self-elaboration

For each transformation, the following dimensions must be defined:

- **Final data**: Represents the indicator to which the transformation refers;
- **Component**: Indicates the data to which the transformation refers;
- **Transformation description**: Contains a brief summary of the type of transformation;
- **Type of transformation**: indicates the type of transformation chosen from one of the following:
    - **Aggregation**: Sum or ratio of two or more information (without using a calculation algorithm);
    - **Disaggregation**: Simple disaggregation of two or more information (without using a calculation algorithm);
    - **Mapping**: Transcoding of information through the use of tables for restoring the values;
    - **Shortcut**: Determination of standard values in case of missing data or entered incorrectly;
    - **Computing**: Complex operation performed by calculation algorithm;
    - **Manual Adjustment**: Manual operations carried out directly by the Data User;
- **Methods of execution**: Indicates whether the transformation takes place in an automated or manual way;
- **Responsible structure of the transformation**: Indicate the structure that takes charge of the transformation in object;

4.3.2.3 Registry of systems

The objective of this analysis is to define every single system through which information is generated, transits, undergoes processing or is stored / deleted.

Before explaining the composition of a registry of systems, a typical table is shown below (Table 4.4), thanks to which collect in detail all the systems highlighted in the logical path of data (Graph 4.2)

REGISTRY OF SYSTEMS					
Final data	Component	Name of the system	Description of the system	Type of system	Responsible structure of the system
α	β	System A	Collect all the data coming from the accounting department	Internal system	Department X
α	θ	Excel	EUC platform for manual checks or transformations	EUC	-
α	γ	Access	EUC platform for manual checks or transformations	EUC	-
α	ε	External provider	Provides stock exchange ratios	External info-provider	-
α	ι	System B	Collect all the data coming from the sales department	Internal system	Department Y
α	η	External provider	Provides macroeconomics ratio about the country	External info-provider	-
α	λ	Excel	EUC platform for manual checks or transformations	EUC	-
α	ζ	System C	Collect all the data coming from the purchase department	Internal system	Department Z

Table 4.4: Registry of systems. Source: self-elaboration

For each system the following aspects are defined:

- **Final data**: Represents the data referred to by the information transiting in the systems;
- **Component**: Indicates the data transiting in the system;
- **Name of the system**: Indicates the exact name of the system, external provider or EUC software used
- **Description of the system**: Contains a brief summary of the type of system that contextualizes its work.
- **Type of system**: indicates the nature of the system, (e.g. Information provider, internal system or UEC);
- **Responsible structure of the system**: Indicates the structure responsible for the system in question.

4.3.2.4 Registry of organizations

The objective of this analysis is to define every single structure that is involved in the process of creating the indicator.

Before explaining the composition of a registry of organization, a typical table is shown below (Table 4.5), thanks to which collect in detail all the organization highlighted in the logical path of data (Graph 4.2)

REGISTRY OF ORGANIZATIONS			
Component	Organizational structure which send the data	Organizational structure which receive the data	Presence of documentation attesting the quality of the data
β	Department Y	Department X	Yes
θ	Department Y	Department X	No
γ	Department Y	Department X	No
ε	Department Y	Department X	Yes
ι	Department Y	Department X	Yes
η	Department Y	Department X	Yes
λ	Department Y	Department X	No
ζ	Department Y	Department X	No

Table 4.5: Registry of organizations. Source: self-elaboration

For each organizational structure involved in the data creation the following aspects must be defined:

- **Component**: Indicates the data passing from one organizational structure to another;
- **Organizational structure which send the data**: Indicates the structure that send the information;
- **Organizational structure which receive the data**: Indicate the name of the structure that receive the information;
- **Presence of documentation attesting the quality of the data**: Indicates whether or not there is documentation that certifies the level of information quality.

4.3.3 Business glossary

In order to standardize and unify the rules for interpreting corporate information, it is necessary to map and catalog information. Due to the complexity and heterogeneity of the information, this activity is articulated and burdensome, therefore it becomes necessary to equip itself with functional tools for an effective mapping that is in line with the organizational context such as the Business Glossary. The Business Glossary serves for the cataloging of logical data and related metadata[15] (the one that in the previous chapters we have called "identity card" of the data), describes at the level of functional semantics the information of the company, through the "Logical Map of Information ", Integrating it with additional notions relevant for proper management of the same (i.e. information object, data owner, etc.).

The Business Glossary and the Logical Map of the information allow to improve the communication favoring a complete integration between the business functions and the IT functions.

In particular, the linking of terms to IT resources allows to effectively foster the dynamics of collaboration and communication.

The ability to access the definitions relating to business terms ensures a better understanding and knowledge of the information assets and, therefore, greater efficiency and better use of resources.

The Business Glossary tool, therefore, enables data governance through the creation of an information sharing environment within which a common language and shared representation logic are used on the Business side and on the IT side.

The Business Glossary is able to detail the definitions and characteristics of the business terms, but also to associate them with each other and to the metadata of the technical assets (i.e. chart, table, physical data).

Business Referrals, with the support of the Data Governance Unit, feed the Business Glossary. The Data Governance guarantees consistency and updating of the logical map

[15] The metadata is the identity card of the data. They are all the data we previously collected through the logical path of data and the registries.

of the Information. The alignment between the logical map and the physical map of the Information is controlled by the Data Management structure.

Before explaining the composition of a Business Glossary, a typical table is shown below (Table 4.6[16]), thanks to which collect in detail all the components highlighted in all the logical paths of data (Graph 4.2)

BUSINESS GLOSSARY									
Final data	Component	Description of the component	Level	Recipient of use	Data Owner	Data User	Origin system	Transit system	System of destination
α	α	Data used to inform a supervisory body about...	Final data	Regulatory	Department X	Office y in the department X	Excel	Excel	Excel
α	β	Data used to inform the sale department about...	1°	Regulatory	Department X	Office y in the department X	Excel	Excel	Excel
α	γ	Data used to compute the α	1°	Regulatory	Department X	Office y in the department X	Excel	Excel	Excel
α	ε	Data used to compute the β	2°	Regulatory	Department X	Office y in the department X	Excel	Excel	Excel
α	ι	Data used compute the γ	2°	Regulatory	Department X	Office y in the department X	Excel	Excel	Excel
α	θ	Data used to compute the β	2°	Regulatory	Department X	Office y in the department X	Excel	Excel	Excel
α	η	Data used compute the γ	2°	Regulatory	Department X	Office y in the department X	Excel	Excel	Excel
α	ζ	Data used to compute the ι	3°	Regulatory	Department X	Office y in the department X	External provider	Excel	Excel
α	λ	Data used to compute the θ	3°	Regulatory	Department X	Office y in the department X	Data Warehouse	Internal system	Excel
α	κ	Data used to compute the θ	3°	Regulatory	Department X	Office y in the department X	Data Warehouse	Internal system	Excel

Table 4.6: Business glossary. Source: self-elaboration

For each component involved in the final data creation (final data included), the following aspects must be defined:

- **Final data**: indicates the main information, which contains the information that is being dealt with specifically
- **Component**: indicates the specific information processed
- **Description of the component**: brief description of the functions and applications of the information. (eg the information is necessary for the registration of securities)

[16] Recipient of use: we imagine that the final data we are analyzing is contained in a report that has as its recipient of use a supervisory body. For this reason, all the data necessary for its formation are assumed to be also regulatory.

- **Level**: As anticipated in the section relating to the data lineage, the information is decomposed into several elementary components, which correspond to the data reported in this section of the glossary, on several levels. This section shows the level of the data recorded
- **Recipient of use**: consistently with what has been done in the report prioritization activity, this field shows the intended use of the report / information object (Operational, management, board, market, regulatory unit)
- **Data Owner**: Indicates the Data Owner who is assigned the Accountability of the information being analyzed as a knower of his life cycle. The Data Owner has the function of guaranteeing the quality of the data (through the tools provided by the Data Governance system) of which it is the end user and this responsibility remains unchanged regardless of whether the same data, in whole or in part, is produced and / or updated by other actors, internal and / or external to the reference company structure.
- **Data User**: Indicates the Data user of the information being analyzed and is the operational manager of the data quality. It supports the Data Owner in determining the data quality indicators (KQI) and in defining and executing the related controls.
- **Origin system**: Indicates the system from which the data in question comes
- **Transit systems**: indicates the systems on which the data in question transits
- **System of destination**: indicates the final system on which the data arrives

4.4 Definition and implementations of the Key Quality Indicators

The Key Quality Indicators represent the main metrics for measuring the performance and quality of the Data Governance processes. In particular, the KQIs make it possible to measure the results of the checks (grouped into families represented by the different quality control frameworks) and to assess the relative impacts, in order to increase the quality of the data processed in the company.

The KQI measures the quality of a data or a group of data according to predefined metrics aimed at ensuring the completeness, correctness and timeliness of the data and information represented.

In order to make the results of the controls and the severity of the anomaly comparable, it is necessary to develop tolerance thresholds for each unit of measurement. Through KQIs it is possible to measure the impact of Data Quality anomalies in terms of exposure or volume. Through this measurement it is possible to identify the most serious anomalies and intervene promptly for the purpose of their resolution.

The ultimate goal of KQIs is to provide information about the quality of the indicator being analyzed and its main components outlined in the bill of materials. Starting from the KQI level associated with the single component (lower level information on the logical path of data), the DGO defines the cumulative quality of the higher level.

The quality of an information is therefore a function of 2 parameters:

- **KQI of the data production plant and the control system**;
- **KQI related to the controls applied on the information itself**.

In the process of processing KQIs, the Data Owner must:

- Map all the checks carried out for the production of the single information
- Aggregate the controls for each component of the related information of which the quality is being measured
- Aggregate every single component of the type of checks that is carried out;
- Calculate the KQIs using the following formula:

$$KQI_j = \sum_{i=1}^{n} \frac{E_{i,ok}}{E_i} * \alpha_i \qquad i = 1,2,\dots,n\ ;\ j = 1,2,\dots,m$$

Where is it:

KQI_j: represents the value of the KQI associated with the j-th information;

$E_{i,ok}$: represents the number of positive outcomes related to the i-th type of control over j-th information;

E_i: represents the total number of outcomes (positive and negative) related to the i-th type of control over the j-th information;

α_i: indicates the weight associated with the particular type of control.

The outcome of the KQI will be analyzed in order to evaluate the quality of the data and to implement a remediation plan where it will be deemed necessary.

4.5 Gap analysis and remediation plan

The last phase of the Data Governance process concerns the study of all the material collected in the previous paragraphs, in the light of the Key Quality Indicators, and put in place corrective actions.

In this phase, the gap analysis of the as-is framework is carried out with respect to the to-be model required by the Supervisory Authority (BCBS n.239) and by the Standards contained in the main reference regulations to which it was inspired the company. In particular, the Data Governance Office carries out, with the help of the Data Owner and Data User, a gap detection study in relation to three areas:

1) **Controls area**: for each information present in the perimeter, the DGO identifies, firstly, the difference between the number of checks carried out as part of the as-is framework and the total number of types of data quality control identified as necessary during the planning period. Once the control gap has been identified, the DGO, with the collaboration of the Data User, defines if one is in the presence of:
    - controls not applicable to the data / information in question;
    - controls that must be implemented in order to produce an improvement in the quality of the data and, therefore, an increase in the related KQI. In the latter case, the Data Owner, with the support of the DGO, proceeds with the implementation of the new controls in order to conform the as-is framework to the to-be model. In parallel, the DGO carries out a simulation analysis aimed at making a timely estimate of the existing gap aimed at measuring the value of the KQI prior to the implementation of the new controls, defined with the Data Owner, and the subsequent one.
2) **Systems**: the data quality perceived by data receivers depends, among other things, on the systems on which the Data Users apply the different types of transformation. For this reason, the Data Owners, with the support of the DGO, analyze the quality of the systems used through the following phases:
    - identification of all EUCs present in the logical line of data
    - qualitative assessment of the EUC according to the level of risk associated (high / medium / low) and determined on the basis of the number and

complexity of the manual elaborations declared by the Data Users through a specific questionnaire provided by the DGO.

3) **Organizations**: Data Users, as data receiver, with the support of the DGO, identify the Data Providers of all data / information received in the perimeter, and perform an analysis of the checks carried out by the same Data Providers in order to identify any gaps in line with the provisions of point 1).

The Data Owners communicate to the DGO the results of the gap analysis and define the interventions and the possible best practice to be undertaken in order to implement the remediation plan.

The objective of the remediation and improvement plan is to define, list and direct the actions to be taken to close the gap that emerged during the Gap Analysis. The implementation of remediation activities consists mainly of four main moments:

1) **Detection**: in line with the arguments in the gap analysis phase, the Data Owners detect anomalies, in terms of data quality, within their own processes. In particular, it identifies further controls to be implemented in order to improve data quality;
2) **Preparation and validation of the plan**: The Data Owners prepare an action plan to implement the necessary new controls as identified in the gap analysis phase. The Data Owner proceeds with the validation of the plan;
3) **Resolution**: The Data Owners, with the support of the DGO, implement the corrective actions envisaged in the action plan set out in point 2) by including the new controls at the various levels of lineage on which the gap was found;
4) **Monitoring**: The Data Owner monitors the progress of the work to resolve the detected anomalies and measures their quality by comparing the KQI before the implementation of new checks and the next one.

The last phase of the data governance process, monitoring, is a cyclically repeated phase whenever new data are included within each company structure. The monitoring phase is the one that triggers the awareness of securing new data through all the steps listed above, from mapping to the definition of Key Quality Indicators, gaps and implementation of the remediation plan.

Criticism and conclusion

Through the thesis it was possible to understand the importance of data within a business context and consequently to understand the attention given by the companies about their management. The case study, as previously announced, takes as an example an abstract and generalizable business case given the initial data management level of the extended company. This gave us the opportunity to represent a path of Data Governance that, with the due differences, unites all the companies that have to go into such a project starting from an almost primordial stage. Given the standardization of this process described, it is also possible to outline what are the main challenges and critical issues that a company faces in the implementation of this project.

Before entering into the merits of the project and its specific problems, there is a great point of attention that unites the whole process of Data Governance and every company has to face. As we have noted in previous chapters, the process of Data Governance is deep and widespread within the company and in a world made up of limited resources the determination of the personnel dedicated to such a project becomes fundamental. The general critical nature of this project comes into play here. It is due to the fact that the evolution of today's companies is very fast and the implementation / disposal of data takes place at a very high speed and if a company devotes a modest amount of resources, it runs the risk that the evolution of the data assets in ownership of the company grows faster than the Data Governance process that would not be able to compensate for such growth and fall into inefficiency. For this reason, every company must necessarily have a quantity of resources that allows it to be always efficient. This aspect obviously changes from company to company depending on the size and type, but certainly not an aspect to underrate.

Going on the merit of the project, the first critical issues we find in the assessment phase. The assessment phase is a starting point in which the people dedicated to the project leave the boundaries of their areas of competence and face the complexity of the company in its entirety. The only "weapon" they possess is the census sheet with the attached questionnaire. As we have previously said, the analyzes carried out during the assessment are nothing more than the collection of the information provided by the data Users, verbally through an interview, of each structure about the data handled by them. As it is

easily conceivable such a type of information collection involves, by its nature, a failure to cover all the necessary components and a superficial census, certainly not objective, of what has been collected. This creates two types of problems: the analyzes that will be carried out on the census will be carried out on a material that does not represent the objective reality of the company because of their superficiality; the second problem will arise well beyond the assessment phase, in fact the components that have not found coverage in the census will result in the gap analysis phase in which the Data Governance Office will be forced to restart the entire Data Governance process for missing components.

As for the lineage phase, on the other hand, the problem just mentioned about the assessment becomes more widespread. The lineage, which we remember to be the deep study phase of every single data circulating in the company, takes place almost completely through repeated interviews with the operational staff in charge of creating the data in question. As we know during the lineage the collection of all the information about the data created by the operating personnel takes place and the latter when it is interviewed knows that it is not only censoring a data but is recording its work. Precisely for this reason every operator will struggle to be objective about the production of the data that he creates himself. Most likely he will be forced to say that he does all the necessary checks, that he never finds problems in the transformations he performs and that the final data he produces is always correct. Here comes what is called “corporate culture” and therefore the perceived need for the project. From this we deduce how, in the implementation of the Data Governance, the diffusion within the company of the awareness of the innovative concept that we want to bring is important. Only when Data Governance is strongly perceived by the whole company is it possible to have a real mapping of reality, thus preventing huge gap in gap analysis and greatly improving the efficiency of the process.

About the data lineage there is a high criticality affecting the mapping of the data of the first structures in the prioritization raking. After mapping and census the personal data of a given structure, it is possible to completely define the quality level of a data only if this data was produced without the help of external sources to the structure under examination. Recall that the KQI definition of a given final data is due to the cumulative sum of the KQIs of the previous components. For this reason, if a final data had in its lineage also

components from external structures not yet mapped (and therefore do not have a KQI) we would not be able to define a KQI that is completely reliable due to some missing data.

Always keeping in mind KQI, according to my personal opinion, another critical issue is to be found in the very definition of KQI and its method of calculation that changes from company to company producing a completely self-referential quality vision. We remind you that Key Quality Indicators are fundamental within the company to ascertain the quality of the data produced and exposed, but they are also fundamental outside the company for third parties as they could be investors. For this reason, the calculation of KQIs should take place according to standard parameters defined and equally applicable in order to have a symmetry of information that allows third parties to understand the quality of the data produced in each company without having to calibrate their own judgment according to the parameters chosen by every single company.

All this kind of critics confirm that the Data Governance is more than a simple project; it is a circular process in which the monitoring phase is only the beginning. In fact, when we talk about “project” of Data Governance we expect it to have a beginning and an end, but in reality, a Data Governance project never ends for the following reason. All components (in a broad sense) within a company, especially nowadays, change constantly and it is impossible to imagine a situation in which gaps are completely filled. In fact, the frequency with which new needs arise in a company, from every point of view, is very high and if it is true as defined in the first chapter, “data are the new asset on which the value of a company is based “, then to meet these new needs there will always be new data that will require to be mapped, examined, controlled and monitored.

Bibliography

"BCBS 239, January 2013" (Basel Committee on Banking Supervision) of the Bank For International Settlements

Basel Committee on Banking Supervision 268 "Progress in adopting the principles for effective risk data aggregation and risk reporting" (December 2013).

Cheong, Lai Kuan, and Vanessa Chang. "The need for data governance: a case study." *ACIS 2007 Proceedings* (2007): 100.

Cohen, Shoshanah, and Joseph Roussel. *Strategic supply chain management: The 5 disciplines for top performance*. McGraw Hill Professional, 2005.

Crié, Dominique, and Andrea Micheaux. "From customer data to value: What is lacking in the information chain?." *Journal of Database Marketing & Customer Strategy Management* 13.4 (2006): 282-299.

"Directive 2013/36 / EU" of the European Parliament and of the Council of 26 June 2013 on access to the activity of credit institutions and prudential supervision of credit institutions and investment firms;

European Central Bank - "Asset Quality Review" Phase 2 Manual (March 2014).

Felici, Massimo, Theofrastos Koulouris, and Siani Pearson. "Accountability for data governance in cloud ecosystems." Cloud Computing Technology and Science (CloudCom), 2013 IEEE 5th International Conference on. Vol. 2. IEEE, 2013.

Giampiero Carli Ballola, 4 October 2010: https://www.zerounoweb.it/techtarget/searchsecurity/la-definizione-della-data-governance-il-primo-passo-per-la-visione-unica-della-realta/

Groß, Stephan, and Alexander Schill. "Towards user centric data governance and control in the cloud." Open Problems in Network Security. Springer, Berlin, Heidelberg, 2012. 132-144.

Kaufmann, Daniel, Aart Kraay, and Massimo Mastruzzi. "Measuring governance using cross-country perceptions data." International handbook on the economics of corruption 52 (2006).

Kaufmann, Daniel, Aart Kraay, and Massimo Mastruzzi. Governance matters VIII: Aggregate and individual governance indicators 1996-2008. The World Bank, 2009.

Kaufmann, Daniel, Aart Kraay, and Pablo Zoido-Lobatón. "Governance matters." Finance Dev 37.2 (2000): 10.

Kooper, Michiel N., Rik Maes, and EEO Roos Lindgreen. "On the governance of information: Introducing a new concept of governance to support the management of information." International Journal of Information Management 31.3 (2011): 195-200.

Otto, B. (2011). Organizing Data Governance: Findings from the Telecommunications Industry and Consequences for Large Service Providers. CAIS, 29, 3.

Otto, Boris, and Kristin Weber. "Data governance." *Daten-und Informationsqualität*. Vieweg+ Teubner, 2011. 277-295.

Otto, Boris. "A morphology of the organisation of data governance." *ECIS*. Vol. 20. No. 1. 2011.

Panian, Zeljko. "Some practical experiences in data governance." World Acad. Sci. Eng. Technol 38 (2010): 150-157.

Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016.

Regulation (EU) n. 575/2013 of the European Parliament and of the Council, of 26 June 2013.

Supervisory Provisions for banks with the "Circolare n. 285 of 17 December 2013 "with the 11th Update of 21 July 2015 of the BANCA D'ITALIA

Thompson, Nik, Ravi Ravindran, and Salvatore Nicosia. "Government data does not mean data governance: Lessons learned from a public sector application audit." Government information quarterly 32.3 (2015): 316-322.

Tihanyi, L., Graffin, S., & George, G. (2015). Rethinking governance in management research. Academy of management journal, 1015(1), 1-9.

Watson, Hugh J., Celia Fuller, and Thilini Ariyachandra. "Data warehouse governance: best practices at Blue Cross and Blue Shield of North Carolina." Decision Support Systems 38.3 (2004): 435-450.

Weber, Kristin, Boris Otto, and Hubert Österle. "One size does not fit all---a contingency approach to data governance." *Journal of Data and Information Quality (JDIQ)* 1.1 (2009): 4.

Wende, Kristin, and Boris Otto. "A contingency approach to data governance." (2007): 163-176.

Wende, Kristin. "A model for data governance-Organising accountabilities for data quality management." *ACIS 2007 Proceedings* (2007): 80.

Summary

Everything that surrounds us today is governed by data. The data as such, however, are worth very little; what gives value to them is the ability of those who possess them to know how to read, aggregate and manage them. This is the source of the great value of data: from the irrefutable potentials and their ability to determine who wins and who loses in every business sector.

Nowadays the evolution of the business world has reached unimaginable rhythms up until a few decades ago. The war between companies and scholars, in predicting trends and understanding the drivers of tomorrow's success, is becoming more and more "bloody", and the one thing that unites all current and future trends and drivers are data.

Data on the one hand, if well managed, bring great benefit to the company and on the other hand, if mismanaged, they expose to a great risk both the company and third parties. Precisely for this reason, to accompany companies towards a process of data management, regulations and reference principles have emerged that dictate the perimeter and the constraints of a correct management of data. The relevant normative drawn up by the reference supervisory body are the "BCBS 239, January 2013" (Basel Committee on Banking Supervision) of the Bank for International Settlements and the Supervisory Provisions for banks with the "Circolare n. 285 of 17 December 2013 "with the 11th Update of 21 July 2015 of the BANCA D'ITALIA.

Them give the general principles to strengthen banks' risk data aggregation capabilities and internal risk reporting practices, and these principles are expected to support banks to improve the infrastructure for reporting the most important information, in particular those used by the board of directors and senior management for identify, monitor and manage risks; improve decision-making in the entire banking organization; improve the management of information among the various legal entities, favoring at the same time a comprehensive assessment of level risk exposures global consolidated; reduce the likelihood and severity of losses linked to deficiencies in the management of risks; reduce the time taken to prepare information and therefore speed up the process decision-making; improve the quality of an institution's strategic planning and its capacity to manage the risks inherent in new products or services.

Briefly, Principles norm:

- Traceability: the persons involved in the data governance process must guarantee, each for the part of their competence, the traceability of the activities and documents related to the process, ensuring the identification and reconstruction of the sources, information elements and controls carried out at support of activities.
- Segregation of tasks and activities: the process of data governance involves the segregation of tasks and responsibilities, between different organizational units or within them, in order to avoid that incompatible activities are concentrated under common responsibilities.
- Compliance with laws and consistency with the general regulatory framework: the data governance process is defined in compliance with applicable regulations, in line with the internal reference framework and national and international best practices.
- Confidentiality: without prejudice to the transparency of the activities carried out and the information obligations imposed by the provisions in force, the persons working in the data governance process ensure the confidentiality required by the circumstances for each news item / information learned on the basis of their work function.
- Conflict of interest: the people involved in the data governance process act towards their counterparts according to relationships marked by the highest levels of behavioral ethics, in compliance with the Code of Ethics, avoiding decisions and carrying out activities, in conflict, even if only potential with the interests of the Company or in any case contrary to its official duties.
- Approach based on risks and processes: the process of data governance, inspired by a process logic, is based on a preventative approach to risks, contributing to the assumption of informed decisions, and, where possible, the translation of the main risks into opportunities.
- Responsibility management (Accountability): management, within the scope of the functions covered and in achieving the related objectives, ensures the application of the data governance process for the activities of competence, actively participating in its operation.
- Communication and information flows: the information necessary to fulfill its responsibilities, including those regarding data governance, is made available to every corporate body and structure.

Given the guidelines, the companies have to study and build a framework which can allow the application of the principles in the company's tissue. The application method changes from company to company depending on the company's structure and IT systems. Due to that each of them elaborates a custom framework, publishing an own Standard of data Governance.
Despite the different methods of application, every firm starts to build its own Standard from the same starting point: definition of objectives, scopes, tools and perimeters.
The object of a Governance Data Standard is the definition of a data management system that, in compliance with the regulatory provisions, continuously pursues the completeness, correctness and timeliness of the data stored and the information represented.

The data management system is the model that establishes the rules, the actors, the responsibilities, the control models with which the information managed by the information system must be treated, throughout their life cycle, guaranteeing its accountability.

The risks associated with data management can be mapped into the following main types:

1) Operational risk, which can be divided into:
    - Risk events that generate penalties - for example quality defects found on data intended for mandatory reporting purposes that provide for a sanction regime;
    - Risk events that generate losses other than sanctions - quality defects on those data whose erroneous management may involve any other type of operational loss (for example: wrong strategic decisions taken by the summit, judicial or extra judicial reimbursements, devaluations, etc.).
2) Reputational risk - incorrect data management that can cause effects on the negative perception of the company image by customers, shareholders, supervisors and other stakeholders.

In general, the potential risk related to a data item is expressed by the product of the probability that it does not comply with the requirements defined for its use (for example, it has a higher percentage of anomalies than that considered acceptable) multiplied by the

impact of the damage that it follows that it is a function of its recipient of use (for example the payment of an administrative penalty for irregularities in reporting to regulators).

Before go deeper, it's important to name the main actors in a Data Governance model:

- Data Governance Office: The Data Governance Office is the center of competence of Data Governance for the company, with the aim of supporting the various processes of Data Governance, by virtue of specific expertise on the subject. It is entrusted with tasks of direction and control and operational tasks in collaboration with the other roles involved in the Data Governance.
- Data Owner: The role of Data Owner is assigned on the basis of functional responsibilities that perimeter a set of services / applications and therefore data. The Data Owners are identified in the managers of the company functions involved in the Data Governance process, on the basis of the identified perimeter. The Data Owner is a figure who is given the Accountability of a given datum as a connoisseur of his life cycle. The assessment activities allow each Data Owner to identify the main data managed, with a known life cycle and for the use destinations for which they are responsible.
- Data User: The Data User is a figure with in-depth professional skills related to a specific business operation of which he knows the functional logic and the first level interrelations with the other related areas; in particular, it has a specific knowledge of the value of data and their treatment in this operating context. He is therefore the primary interlocutor of the Data Owner and, by virtue of his specific skills on the subject he oversees, he is directly responsible for the data quality control process.

"Data" is defined as the information managed and processed by an IT tool that the company uses and manages for its strategic, reporting and operational purposes.

For each data it is necessary to identify two main aspects: its type (depending on its life cycle) and the uses for which it was generated (recipient of use);

Data must be governed regardless of their source or destination; the company must ensure that, data provided or disseminated in its own name, comply with the rules in force to guarantee regulatory compliance and to protect any consequences (economic, regulatory or reputational) of inadequate management.

However, in order to identify an intervention priority, it is essential to establish a relevant perimeter for data governance purposes, in the sense that the perimeter data will be assigned an entry order in the program and, if necessary, the appropriate corrective measures will be defined and adopted. The quality level should not be adequate.

Data that fall into at least one of the following two categories are defined as relevant: high risk use destination and type of data at risk.

The high-risk use destinations are as follows:

- reporting or information purposes for top management;
- periodic financial reporting purposes for shareholders and stakeholders in general;
- valuation purposes on company processes and systems for the Control Functions;
- reporting or information purposes to the Supervisory Bodies;

In determining the perimeter, an approach is therefore adopted, which enhancing what is already present in the company at the level of description of processes, applications, information and risks, will introduce the Services dimension and data which will be accompanied by an identity card compiled on the basis of their life cycle and integrated with the Key Quality Indicators.

The Key Quality indicators aim to measure the qualitatively most relevant aspects with respect to data management. They also make it possible to extend the control domain to all phases of the life cycle by relating the results of the checks carried out in several phases / processes.

These indicators measure the quality level according to the three main directions:

1) completeness;
2) accuracy and integrity;
3) timeliness;

The measurement of the defined Key Quality indicators will be periodically reported, with periodicity and level of detail, different to the individual Data Owner, the control functions, the Company Manager and the management body.

Going deeper, the implementation of the Data Governance process starts from the Assessment, in which a functional perimeter is defined, called “priority”, in which services and data will be collected from the relevant Data Owners.

For the remaining data considered as non-critical, the level of original risk is accepted without specific mitigation (also called “inherent risk”).

The assessment phase aims to create a complete picture of the relevant information managed within the competence of the Data Owners of a company (remember that the data owners are usually the managers of the structures, those who have full responsibility of the data produced by their structures), as well as tracing the means by which this information is transmitted / communicated inside or outside the company (eg report, information flow, printouts), which we have identified as “Report”.

The final aim of the assessment phase is to obtain a ranking of the company organizational areas based on two main drivers: the relevance of the product reports (recipient of use) and their management methods. Information about these two drivers are collected thanks to two tools: census sheet and questionnaires.

The synthesis of all the information collected is done through an instrument called “prioritization matrix”. It builds a ranking of the organizational areas in such a way as to define which of these needs an immediate intervention. The identification of this ranking is fundamental in the planning of a Data Governance project. In fact, this project requires an important effort from every point of view (costs, personnel and so on) and it is not possible to apply it to all the structures at the same time, but it is necessary to implement

structure by structure (each structure will have what we will call its own analysis “wave”). For this reason, thanks to the ranking identified (in addition to the census that allowed us to understand the amount of components that will be improved), it is possible build a plan in terms of time, costs, personnel employed and applications to be used / integrated.

The next step, after the collection of all the information regarding the various structures, we move to the individual wave, that is the structure-by-structure analysis in order to understand the life cycle of each data, the transformations it undergoes and the controls on it. This phase is called “Data Lineage”.

In the lineage phase, each single data, according to the order established in the prioritization phase, must be analyzed individually through a process of logical lineage, which allows to highlight the entire process of building information.

The lineage can be defined as a synoptic map of the data path in the reporting chain (Data lifecycle), including ownership and tracking of all the checks performed in the steps within the company information systems.

The term Life cycle of a data identifies the set of all the phases that describe the existence of a data within the company, from its creation to its elimination / archiving.

The data lineage occurs through the use of three main tools:

- **The logical path of data:** In order to correctly represent the entire logical line, the data must be decomposed graphically into all its components, highlighting, by levels, each single decomposition, transformation, control and support systems.
- **Data registry:** The objective of this analysis is to define every single control, transformation, system and organization that through which the data pass through during the composition of the final information.
- **Business glossary:** The Business Glossary serves for the cataloging of logical data and related metadata (the one that in the previous chapters we have called “identity card” of the data), describes at the level of functional semantics the information of the company, through the “Logical Map of Information “, Integrating it with additional notions relevant for proper management of the same (i.e. information object, data owner, etc.).

Thanks to all the information collected in the previous phases, it is possible to proceed with the detailed study of the processes in order to implement quality controls through the Key Quality Indicators.

The Key Quality Indicators represent the main metrics for measuring the performance and quality of the Data Governance processes. In particular, the KQIs make it possible to measure the results of the checks (grouped into families represented by the different quality control frameworks) and to assess the relative impacts, in order to increase the quality of the data processed in the company.

The KQI measures the quality of a data or a group of data according to predefined metrics aimed at ensuring the completeness, correctness and timeliness of the data and information represented.

In order to make the results of the controls and the severity of the anomaly comparable, it is necessary to develop tolerance thresholds for each unit of measurement. Through KQIs it is possible to measure the impact of Data Quality anomalies in terms of exposure or volume. Through this measurement it is possible to identify the most serious anomalies and intervene promptly for the purpose of their resolution. The identification and resolution of anomalies occurs through “gap analysis” and the “remediation plan”.

The gap analysis of the as-is framework is carried out with respect to the to-be model required by the Supervisory Authority (BCBS n.239) and by the Standards contained in the main reference regulations to which it was inspired the company. In particular, the Data Governance Office carries out, with the help of the Data Owner and Data User, a gap detection study in relation to three areas:

- Controls area: for each information present in the perimeter, the DGO identifies, firstly, the difference between the number of checks carried out as part of the as-is framework and the total number of types of data quality control identified as necessary during the planning period;
- Systems: the data quality perceived by data receivers depends, among other things, on the systems on which the Data Users apply the different types of transformation;
- Organizations: Data Users, as data receiver, with the support of the DGO, identify the Data Providers of all data / information received in the perimeter, and perform an

analysis of the checks carried out by the same Data Providers in order to identify any gaps in line with the provisions

The Data Owners communicate to the DGO the results of the gap analysis and define the interventions and the possible best practice to be undertaken in order to implement the remediation plan. The objective of the remediation and improvement plan is to define, list and direct the actions to be taken to close the gap that emerged during the Gap Analysis. The implementation of remediation activities consists mainly of four main moments: Detection, Preparation and validation of the plan, Resolution and Monitoring.

The “last” phase of the data governance process, monitoring, is a cyclically repeated phase whenever new data are included within each company structure. The monitoring phase is the one that triggers the awareness of securing new data through all the steps listed above, from mapping to the definition of Key Quality Indicators, gaps and implementation of the remediation plan.

As we can see in the end of the last chapter of the thesis and we can deduce from the criticism chapter, does not exist a real ending phase when we talk about a data Governance project. In fact, it is a circular process in which the monitoring phase is only the beginning because there will always be new data that will require to be mapped, examined, controlled and monitored.

As we can see in the end of the last chapter of the thesis and we can deduce from the criticism chapter, does not exist a real ending phase when we talk about a data Governance project. In fact, it is a circular process in which the monitoring phase is only the beginning because is impossible to imagine a situation in which gaps are completely filled due to the introduction of new data or the modification of the existing data, that will require to be mapped, examined, controlled and monitored.