POLITECNICO DI TORINO

Corso di Laurea Magistrale
in Ingegneria Gestionale
(Engineering and Management)

Tesi di Laurea Magistrale

Financial Services embrace Data Governance: a real case study on a domestic Bank Group

Relatore

prof. Claudio Giovanni DEMARTINI

Candidato

Nicola IERINÒ

Anno Accademico 2019/2020

Table of Contents

Introduction

The world we live in is going through what is called the ‘digital era’, dominated by one and undisputed protagonist: Data. The volume of circulating data is growing dramatically, day after day, so much that it is estimated that every second more data travels on the network than there were on the whole Internet just 20 years ago. Technological progress evolves at an almost unpredictable speed and innovations are increasingly radically based on the production and use of data: you can think, purely by way of example, to social networks, voice assistants or intelligent cars. And this is only the tip of the iceberg of an evolution that is bringing, or more likely has already brought, the Data to the center of global attention.

The increase in the data produced is reflected on the organizations in two opposite ways: if on the one hand they are called to a substantial effort to adopt corporate systems and frameworks suitable for the new information needs, on the other an efficient data management can transform into an unprecedented competitive advantage. Data have become the basis of competition, productivity, growth and innovation and their rise in volume is adding new challenges and opportunities for harnessing the power of information, so much so that a new branch has seen the birth: Data Governance.

The ability to benefit from the possession of data and transform them into a solid base on which to build your business, is inevitably linked to the organization's ability to manage and control them. Inaccurate and disorganized management can only lead to disadvantages: the data, from a source of competitive advantage, can turn into a source of vulnerability, for example by making the correct estimate of the risks inaccurate. The argument is particularly valid for organizations that manage extremely sensitive data and on which the peaceful dreams of a large part of the world population are based: we are talking about Financial Providers. Faults in the data management framework or poor quality of the processed data, would have catastrophic repercussions on the entire economy. It is therefore no coincidence that data-related topics have fallen under the attention of Regulatory Bodies, especially for organizations that supply services with a great public interest, just like the aforementioned Financial Providers.

As a consequence, the issues of Data governance and information technology, which previously were treated only superficially, now begin to be

included in regulations, directives and circulars, accompanied with insights and detailed specifications. First and foremost, at Community level, we find the Basel III Regulation for bank institutes and Solvency II Directive for insurers.

The topic of Data Governance is a relatively new one and not sufficiently debated in the literature. In particular, the major shortcomings are found not so much at a theoretical level, where a discreet number of studies have been carried out, but under the practical application aspect. What is most lacking, in fact, is the concrete empirical application of theoretical concepts within organizations, even more if we consider its relationship with Regulatory Compliance. Yet, it is acquiring always more considerable importance since Financial Providers are subject to increasingly stringent regulations requirements.

The purpose of this thesis paper is therefore to focus on the practical solutions and the implementation of Data Governance issues within the Financial Services sector, applying the theoretical context in practitioners' world and analyzing what are the benefits of this real application. Thanks to my first-hand participation in the implementation of an internal Data Governance Policy for a Primary Italian Banking Group, it will therefore be possible to give a real feedback to the theoretical framework that will be introduced and to Regulatory Bodies requirements, in particular Circular 285 of Bank of Italy. The objective consists in the creation of a Data Governance model such as to be pervasive within each area of the company studied, scalable and flexible, that is capable of adapting to internal and external changes and based on a literature that, although still incomplete, justifies the activity carried out and the choices made.

The paper is structured in a set of seven chapters divided into two macro areas. The first four present a literal review of the topics, in order to introduce the theoretical foundations and concepts useful for understanding what has actually been done during the project activity. The last three relate to the project and show the activities carried out, the project deliverables and the objectives achieved.

Going into detail, *Chapter 1* is dedicated to the introduction of information systems, namely the business tools in which the data resides, along with their evolution and the main components that make them up.

Chapter 2 analyzes the data more closely. In particular, it deals with its management within organizations which have realized that this is one of the most valuable assets available to them. The theme of Data Governance is also introduced, giving a first definition and vision from a literature perspective.

Chapter 3 is entirely dedicated to Financial Providers and to the deep relationship between finance and technology, which gives life to Fintech. In particular, new Fintech trends that will likely dominate and guide the near future are shown. At the end of the chapter, an explanation of the importance of Data Governance in the Financial Services sector is provided.

Chapter 4 is dedicated to Data Governance, first presenting its relationship with the regulatory framework, then moving on to more technical aspects such as the various available organizational models and, above all, the introduction of the DAMA Framework, depicting international best practices as of today.

Turning to the practical part, *Chapter 5* presents an overview of the project and of the specific organizational reality in which it has been developed. A benchmarking analysis at national and international level is reported, having been used as a starting point for the creation of the Data Governance model. In this chapter we also find a clear list of the planned project activities together with the deliverables agreed.

Chapter 6 represents the real core of the project: it presents in detail the activities performed and the way in which the project has been conducted, starting from the interaction with the members of the organization up to the selection of the software to support the Data Governance model adopted. Given its importance for the development of a solid Data Governance framework, an in-depth focus will be made on Data Quality theme.

Finally, *Chapter 7* closes the paper by analyzing the results achieved so far (the project is still ongoing), the next planned implementation steps to conclude the project and a brief presentation of my direct contribution to the activities.

Chapter 1
INFORMATION SYSTEM: A HOME FOR DATA

Clive Humby, the famous British mathematician and entrepreneur, used to say: *"Data is the new Oil,* and as oil*,* it's valuable, but if unrefined it cannot really be used. It has to be changed into gas, plastic, chemicals, etc. to create a valuable entity that drives profitable activity; so must data be broken down, analyzed for it to have value". The popular quote, cited by many authors in literature, for example (Lin, 2014), highlights how data can be seen as a source of unimaginable power, but to provide such great utility, they must be managed, analyzed and, in some way, transformed into information.

In the 21st century, in fact, data have become one of the major assets of an organization, as important as machines and plants, and, for some organizations, way more important than such physical assets. This has a direct reflection on the structure of the company which has the ownership over data and information: the organization Information Systems (IS). As the information needs of the companies grew, IS evolved themselves to support the new business requirements, allowing companies to handle the increasing amount of data produced and used during the day by day activities. Thus, IS represent the primary allies for what concerns data management and this chapter is dedicated to briefly present the main features that characterize them.

But first, it is necessary to take a step back, in order to clearly understand the distinction among Data and Information, two highly interconnected concepts but with different weights and meanings. 'Data' represent the *atomic substance* of IS, a neutral element devoid of any intrinsic meaning (De Nito, 2008). If not inserted in a specific context, they are not able to convey any message at the reader, which will see them only as a set of letters, numbers, observation, symbols, without any sense. Hence, we can consider data as the raw input of the IS, over which elaboration will be made to give them a significance.

'Information' is the result of the extraction and subsequent processing carried out on raw data: this is the moment in which data, transformed in information, appear meaningful for those who receive them in a specific domain (Vercellis, 2009). The processing could be an aggregation, an elaboration, or simply a contextualization of data, which, combined with a correlation to business

elements, become useful, relevant, and focus on a specific purpose: in short, information is the *output* of the process.

Contextualization is at the core of the difference between the concepts of data and information. Data can be thought as independent from the framework: a list of numbers or names does not provide any information of the context in which it is recorded. The same list should have been created in several different environments. On the contrary, for information, both the context and the recipient are important, since the same set of data could be considered a piece of high relevance information from a person and, at the same time, as totally irrelevant from another (Marakas & O'Brien, 2013).

Strictly connected to the previous concepts, is the one of 'Knowledge'. Debated at length in literature, it has myriad of definition, but for our purpose we will leave apart the more philosophical and abstract definitions to concentrate on the practical meaning of knowledge, viewing it as a *transformation* of information. In particular, information turns into knowledge when it is deployed for the support of the decision-making process and for shape the derivative actions: it can be thought as *applied information.* Such information is applied for a specific scope and is placed side by side to the experience of skilled people able to handle and solve complex problems.

It is finally fundamental underline the importance of the external environment. The transformation of data into information cannot prescind from a deep analysis and understanding of the environment in which the IS is working, since it increasingly affects the processing of data. Clients, vendors, competitors and many other stakeholders must be kept in consideration when information is created, but most of all, what is really important to consider are Regulatory Bodies: they play a key role in the world of data and their presence is very pervasive in the definition and control of creation, processing, management and storage requirements. As we will see in detail in the following chapters, such extensive intervention determines the guidelines of Data Governance in financial services.

Having clarified the key concepts of data and information (and knowledge), we can move to the more interesting concept of Information System. The world of information systems is wide, and a variety of themes can be analyzed. However, given the scope of this paper, it will be shown only the main components, which will be more than sufficient for a clear and complete understanding of what is reported in the following chapters.

An 'Information System' is an *organized combination* of people, hardware, software, communication network, sources of data, policies and procedures built with the aim to storage, retrieve, transform and share the information throughout the entire organization (Marakas & O'Brien, 2013). Note that the definition never

refers to digitalization. This highlight a basic point which most of the people ignore, i.e. the fact that 'Information System' and 'Information Technology' are two different things, with the second being a subset of the second, one of its components[1]. 'Information Technology' represents the *merely digitalize part* of the information system, like computers or software. However, an information system can exist independently from such technological components. Think, as an example, at paper-based files in a library, or the ledger, used for accounting purposes for long times before the advent of computers. Nevertheless, although the specification, it is necessary to note that technology has been adopted widespread and nowadays it is almost impossible to find an IS not based on a predominant digitalized part. And indeed, the IS is no longer regarded as a simple tool for transactions recording purpose, but as an organizational asset in which to invest: it has definitively become a possible source of competitive advantage. Given such a premise, the term information system will be used in the following to refer to system based on modern digitalized technologies.

1.1 Information System evolution within organizations

Information systems, whose birth dates back to the middle of the last century, have been through several evolutions, occurred concurrently to changes in the way entrepreneurs and markets think about organization. This, in turn, depends on the emergence of new customers' needs. Historically, two kinds of structures were opposed, the vertical structure and the horizontal one.

The *vertical* is the more traditional structure and it is associated with the Fordism, in which every member of the organization performs a highly defined and focused task. This type of structure, commonly referred to as 'functional', provides for a clear separation of the different organizational departments in a way that every worker is specialized and efficient in the task assigned to him, like if he is allocated in a sort of "silos": the production department is detached from the sales one and both are separated from the procurement. The only point of contact among them is the top management, which hierarchically coordinates and directs every structure. Although the great efficiency, the model is characterized by a big problem which is leading to his disappearance: the effective coordination of the different 'silos'. In fact, the modern world context is unstable, with sudden markets changes and hyper competition (D'Aveni, 1995) and organizations must be flexible and dynamics to adapt themselves in short time. In the functional model every decision is taken by managers, who do not know the particulars of

[1] See www.floridatechonline.com/blog/information-technology/information-systems-vs-information-technology

their organization and thus they cannot reply in best way to external threats and opportunities. Consequently, the vertical structure is making way to the horizontal one.

The less rigid *horizontal* structure does not provide for a clear subdivision of departments and tasks but promotes dynamicity and continuous learning, from which the denomination 'learning organization' usually given to it (Sange, 2006). As opposed to the vertical structure, in the 'learning organization' everyone is asked to learn and operate as if he is independent from the formally assigned role. Since it is focused on the outputs and not on the process, this type of model can better face a market in which what is new today is almost old tomorrow. Horizontal structure quickly adapts to market and environmental movements, as, by not being assigned to a specific functional 'silos', members are less resistant to changes.

To manage such transformation from vertical functional to horizontal learning organization, the leading purpose to follow is the *integration*: developing a horizontal structure without being able to integrate every single component in an interconnected and unique system will lead to a complete failure, not reaching any of the desired advantages.

Therefore, the integration requirements must be at the top of the priorities list, given that, without a cohesive integration among organizational elements, information could be dispersive and of little use. And here the key point: being one of the most important components of a company, information systems must evolve in parallel with the organizational structure.

Bearing in mind what just exposed, in the following the main evolutions which has characterized the information systems world are presented.

1.1.1 Anthony's Framework

Leaving aside non digitalized information systems, way far from the actual reality, like the old paper archives of the banks, the analysis can start from the sixties. It was in those years that people start to talk about IS in the way we nowadays consider them. A useful tool that can support the description of the evolution over time of the systems is the celebrated *Anthony's Framework* (Anthony, 1965), from Robert Newton Anthony. According to the author, the activities performed in a company can be subdivided into three macro categories and IS were born to satisfy the information needs required to complete such activities. The groups identified are:

1. Operational activities;

2. Employees and managers' decision-making support activities;
3. Strategic decision-making support activities.

Figure 1 shows a representation of Anthony's model, which has been adopted in literature as the main reference in the definition of an information system, being cited a multitude of times, for example in (David J. Teece, 1997), (Shim, et al., 2002).

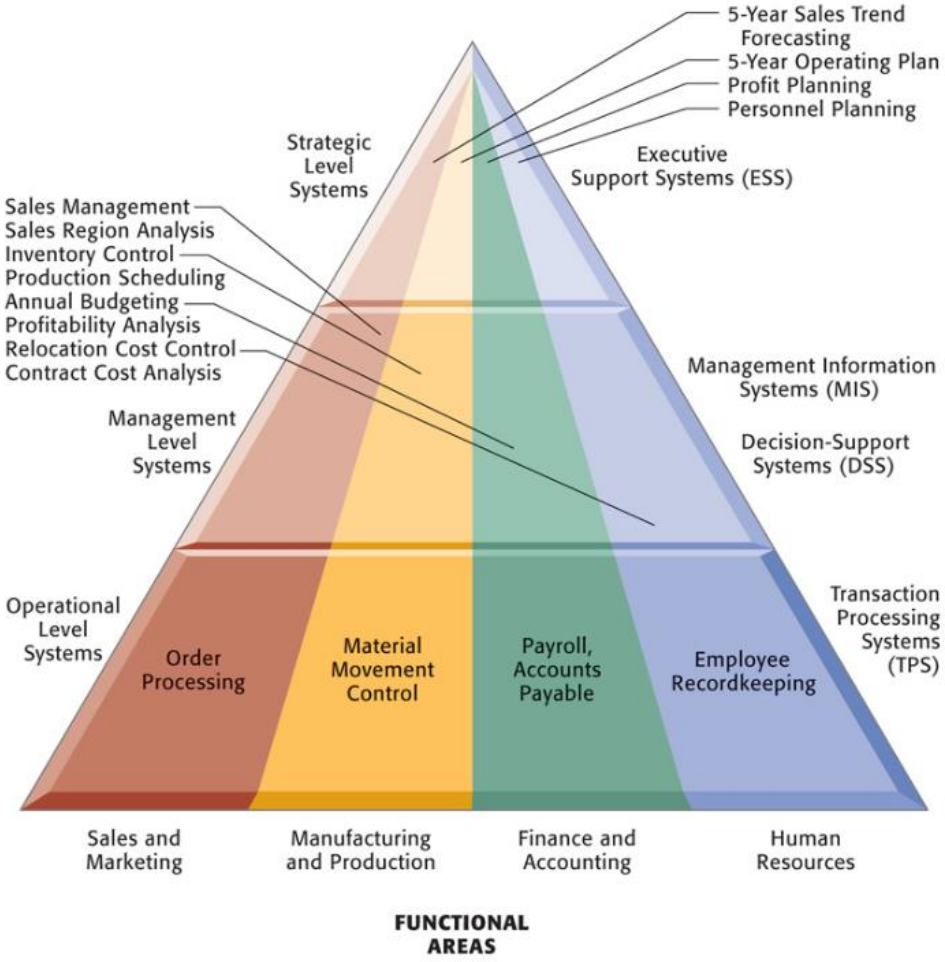


***Figure 1*: Anthony's Framework in (Laudon, 2012)**

Starting from the bottom, you find the *operational activities*, the more practical tasks carried out by workforce. Since standardization and routine hold sway and because of their mechanism and repeatability, operational activities are the first to be embedded in ISs.

Going up to the middle level you find *tactical activities*. These activities involve a lower level of granularity than the operational ones and are usually entrusted to function managers. Tactical activities refer to the programming and control process and therefore relate to goals which are still practical, concrete and

short-termed, such as budgetary objectives or resource allocation. A key aspect of tactical activities is that they must be aligned with the strategic objectives defined by the company's top management, constituting the top of the pyramid.

The *strategic activities* aim to define the strategy of the company in the medium long term in order to guarantee the persistence on the market and the attainment of a competitive advantage. In particular, they involve decisions that impact on the entire company structure, defining or modifying the management modalities, and are characterized by intermittent and unpredictable temporal manifestation (Candiotto, 2013). The granularity is therefore minimal since the activities concern, for example, the market areas in which to operate or the investment policy to be followed.

Although Antony's Framework is recognized as a cornerstone in the construction of information system, it should be noticed that it is a three levels extremely rigid structure which hardly fits to modern organizations. Two are the main reasons: first, every company has its peculiarities and no one of them can therefore be packaged inside a standardized tripartite model. The second reason, but only in order of listing, is that it is very unlikely to find companies with a clear separation between the three different types of activities. This is due to what has already been exposed, namely the need to be responsive to quick destructive changes in the market, which are becoming ever closer in time. Organizations, therefore, seek to eliminate this clear distinction between levels in favor of close and fruitful collaboration between business areas.

Despite the limit just exposed, the model allows to evidence an existing relationship between the various business activities and IS: to different activities correspond different information needs and, as a consequence, different IS able to meet those specific needs. As will be seen shortly, IS were born to meet the information needs required by the activities laying at the base of the pyramid and have evolved over the decades to meet the needs of the activities placed at the top. Finally, in the new millennium, they have gone even further, increasing companies' possibilities exponentially.

1.1.2 The steps toward modern Information Systems

Electronic Data Processing System (EDP)

The first systems to see the light are called EDP System, which stands for *Electronic Data Processing System*. As suggested by the name, these systems deal with data processing, with a focus distant from current IS whose aim is to

disseminate information. EDP, also called TPS (*Transaction Processing System*) are relatively simple and aimed at carrying out those repetitive and extremely standardized activities with the aim of drastically reducing the execution time compared to human execution. These are clearly the activities that populate the base of the pyramid of Anthony: we see how the first activities that benefit from a support in their execution are those operative. There is no doubt that the automation of mechanical activities has led to a huge improvement in terms of cost, time and efficiency within companies, favoring a massive development of the same. On the other hand, however, a huge problem arises in companies, which slowly erode the advantage brought by EDP. In fact, the primordial information systems represent almost an isolated entity within the company, in the hand of very few specialized workers, who are hardly able to interact with the heterogeneity of the requests coming from the different business actors. A harmful gap therefore arises between the effective willingness to bring an informative advantage to the company and the actual possibility to pursue it: the deep groove that divides the IT specialists and the subjects in charge of the operating activities, does not allow the former to know the dynamics of the automated activity, with a consequent difficulty to resolve in concrete way the problems that the latter puts to the attention.

Management Information System (MIS)

Between the 1970s and the 1980s, a first major revolution struck the world of information systems, a revolution that changes the way we think about IS. The focus goes from simple data to information, and many authors try to review the definition of information system. According to Rugiadini (Rugiadini A, 1973) in particular, the information system is defined as "a set of information flows, produced with various methodologies, designed to support the company's decision-making system and meet the information needs of third-party economies in relation to the company".

Considering the pyramid of Anthony, we move to the middle layer and the new systems take the name of MIS, *Management Information System*. They support the tasks of functional managers in the definition of short-term goals, ensuring that their planning decisions are effective and efficient. Thus, they support technical activities.

Francesconi (Francesconi, 2011) claims that "MIS find great use in managerial activities to the generation of structured reports of accounting or extra-analytical nature, with periodic deadlines, through access to historical data of the company". Analyzing what has been said by the author we find some fundamental

points on which to focus our attention. It first confirms that the primary users of MIS systems are functional managers, who now have a more comprehensive view of the activities under their control. Then, he refers to accounting and non-accounting reports. The MIS systems are in fact created to automate the collection of both accounting and analytical information, allowing functional managers to monitor deeply their area, thanks to more targeted and complete reports. Moreover, within each function, the actors are more responsible and involved within the information system, with a hierarchical division of information-related responsibilities among the members of the functional area. Users are allowed to carry out, process, retrieve and communicate information to the top of the organization, allowing the top management to receive communications in a more logical and useful way, not having to intercede directly with technicians as was the case with EDP. It is therefore clear that there is an advantage over old systems, with a clear improvement in the purity, quality and timeliness of information, which, once it reached the management, does not need to be reworked but is directly usable.

However, MISs have limitations in terms of flexibility and analytical capabilities. They do not include sophisticated mathematical tools and use mainly internal data, taking an orientation directed to the past or at maximum to the current situation inside the company, as supported by Francesconi. Not to mention that their development, targeted to meet the information needs of a specific function, has led to the creation of "information islands" with little integration of data and information flows across different structures.

The MIS, therefore, are more useful and complete information tools than the previous EDP, able to analyze what has happened historically within the company and what concerns the present, but some concepts typical of modern systems are still far away: integration, flexibility, decision making support for forecasting purpose.

Decision Support System (DSS)

Towards the end of the 1970s it was clear that standard off-the-shelf information systems could not provide sufficient performance to meet the IT needs required by managers for decision-making operations: *Decision Support Systems* (DSS) were born there.

DSS, as suggested by the name, was thought to interactively support the decision-making process of managers. One of the features that differentiate them from the MIS is their great flexibility and customization: DSS could be adapted to the specific decision to be taken and to the decision style of the managers,

different from one to the other, also due to the variety of real-world problems to be faced. Reporting Martinez's words (Martinez, 2004), "MIS supported decision-making processes too, however there are significant differences between the MIS and the DSS. The MIS provide information and standard knowledge processed on the basis of traditional schemes and procedures, while the DSS are much more flexible and focus on customizing decisions, which change from time to time according to needs". It is therefore no coincidence that DDS developed after EDP and MIS, given their necessarily more complex implementation with respect to previous systems.

The term DSS refers to a large family of different systems which have a common fundamental feature: they all share, besides the aim of decision-making support, the necessary integration of the human component with the information system. The task of the information system is to facilitate the decision-making process, providing all the necessary information and allowing a series of useful analyses, but the final decision remains always and in any case in the hands of the human being. DSSs deals with those decisions that are defined as poorly structured, in contrast to the completely structured ones. For the latter, human intervention is not required, as the system, applying standard rules and algorithms arrives at the decision in an automatic way, leaving no decision-making power to the user and even forcing him towards the final decision; an example is the process of inventory management in which a reorder policy based on a minimal level of stocks is fixed. Conversely, for DSS, the human component is necessary, the last word always belongs to the user of the system: the information machine has the sole role of guiding the human towards the best possible decision, but without ever replacing him.

Depending on the decisions and the recipients, five categories of DSS can be identified. The *Executive Information Systems* (EIS) were the first to appear and were thought for top management figures. Born to support senior management, they focus on 'success factors', focusing the decision maker's attention on those factors and performance indicators most relevant to an organization. EISs are more flexible than MISs, allowing managers to take unstructured decisions and adapt reports to their needs. Moreover, top management asks for less detailed report, with a low level of granularity, request not satisfiable by a MIS system which produces heavy and detailed reports. However, the attention is still on report based on past or, at best, actual data.

The turning point is the introduction of the *Executive Support System* (ESS), evolution of the EIS that enhances their reporting capabilities. But the real improvement was in the decision-making support: the ESS is connected to several information technologies which allow the user to make decision that would be impossible to take based on pure information alone. In fact, ESS no longer has the

sole reporting function: with the introduction of important innovation integrated in the system, like large computation capacity, spreadsheet, communication software and organizational tool, they improve significantly the value of managers' activities.

The activity carried out by these systems are strategic planning, long term resource allocation, scenario analysis: we are clearly in the summit of Anthony's pyramid. Thus, with these new powerful tools, the top management is able to make forecast for the future, finally revolutionizing the way of doing business.

Two other categories are the *Group Decision Support System* (GDSS) and *Organizational Decision Support System* (ODSS). Such tools are used when the decision process requires to involve, respectively, many members of the organization or even external actors. In particular, GDSS systems are used when a decision that has to be taken, for example, will impact on different organizational structures: to ensure a unique way of conduct, dictated first by the summit and then adopted by the various parts of the organization, the support of a GDSS is fundamental. They are therefore intended to bring together different subjects who may be in different geographical locations or even hold quite different hierarchical positions, so as to combine the numerous ideas of groups of subjects, achieving a more productive decision-making process.

The ODSS, like the GDSS, support the decisions made by groups within the organization, but they also have the peculiarity to support the decisions that the organization must make with the external environment, that is, with third parties who are engaged in economic relations with the company. The fundamental point of these systems is certainly the communicative aspect that supports the building of a solid relationship with the external environment.

The last category of DSS is represented by Expert Systems. Their birth is attributed to the enormous increase in activities that at the end of the eighties weighed on the top figures of organizations, who found themselves unable to intervene in all decisions. Expert systems were born with the aim of transferring into them the experience and the decision-making capacity of experienced users, i.e. the company managers. In this way they were raised by day-to-day problems and greater autonomy was left to the more operational members when taking decisions. The aim was to store all the know-how of the organization that is largely placed in the company's top, and share it throughout the organization, from the top down until the operating level, in order to transfer to lower staff the reasoning methodology of those who helped to educate the system.

Enterprise Resource Planning (ERP)

It was only in the late 1990s that the concept of information system began to be accompanied with the one of integration. Until that moment the way of conducting the business was to try to optimize the processes of every single organizational area, seeing each of them almost as a silo separated from the other. Companies were looking on the market for the best off-the-shelf computer products for inventory management, for sales management, for logistics and so on.

Two clear problems arose from this type of procedure: first, off-the-shelf information systems were not designed to satisfy the peculiarities of each company, thus failing to satisfy all the information needs it is looking for. Second, but no less important, for different processes and different business functional areas the organizations utilize a myriad of automated systems extremely heterogeneous without the possibility to interact one another.

The improvement of technologies and the awareness that great advantages could have been obtained from a closely interconnected organization led to the development of *Enterprise Resource Planning*, wider know with the acronyms ERP. These tools now spread from corporate to small businesses, allow every member of the organization to recover immediately and independently any event or information he needs. Their task is to manage every single process present in the company, automating it and reducing the errors edible by the user. But the real innovation is the integration of all business areas so that the activities of the various functions are closely connected and repetitions of the same actions between different areas are avoided, eliminating dangerous duplication of data.

The ERP may be seen as a real ally in the management of organization's activities, but it is necessary to highlight how a poorly designed system that does not fit perfectly with the IT needs of the company brings more disadvantages than advantages. The ERP must be highly personalized and customizable and, first of all, must provide the integration of all information system' s components within the company: programs, mechanisms and automatisms underlying information flows, devices, hardware components, technical instruments, media and communication tools. The ERP should be thought as a system composed by a variety of process modules that can be integrated one another: you can implement just the modules required for solving your specific problems.

Quote Martinez (Martinez, 2004), "ERP systems are characterized by the following specificities: ability to achieve an informational integration thanks to a single database, modularity, configurability". You notice that, as well as the modularity and the possibilities to meet the intrinsic characteristics of the company, another feature provided by ERP is the uniqueness of the data source.

The total integration is in fact made possible only by the creation of a unique database that constitutes the source for all the organizational processes, in such a way that they are closely interconnected and jointly fed between them.

Needless to emphasize the enormous difficulty in implementing these systems, given their huge complexity, the wideness of the areas touched and, above all, the resistance to change that characterizes organizations. The change involves large economic and adaptive efforts, but what you need to focus on to be successful is the involvement of all the business actors, to prevent the operative members from seeing the information system as an enemy to avoid: the economic and organizational consequences in this case would be disastrous.

The new millennium: Business Intelligence and cloud computing

The first two decades of the 2000's, in addition to being characterized by a continuous development and improvement of ERP systems, see the birth of Business Intelligence and Cloud computing**.**

Business Intelligence (BI) refers to all those applications and technologies that are focused on the collection and analysis of data and information that can be used to support strategic decisions. BI technologies allow the organization to have an overall view of the critical internal and external success factors that impact on the business and competitive advantage. Their main task is therefore to extrapolate reliable indications and strategic forecasts from the data collected, in order to allow the top management to direct the organization towards the way of success. Such systems, therefore, require great simulation skills: the more accurate the simulations, the greater the ability of the company to benefit from the data analysis in the future and, consequently, the greater the chance of success. It can therefore be said that Business Intelligence allows you to look "into data" in order to find relationships and opportunities for new profits.

Obviously, the BI technologies must be personalized in order to meet the requirements of every company and, above all, a complete integration and coordination with the ERP management systems is necessary.

Cloud Computing is a modern way for implementing an information system. In the past, companies entrust to external providers the building of personalized IS. Undoubtedly the costs are exorbitant and not all the companies are disposed or have enough resources to make such an investment: the solution to these problems is the Cloud Computing. The precise definition of Cloud computing dictated by

the American National Institute of Standard Technology (NIST)[2] is as follows "Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction".[3] In common jargon, you often hear about Cloud technologies, which are technologies that allow thousands of people to access virtual platforms to manage, store and share their information, programs, services, and software. The same principle can be applied to IS: the organization no longer has the obligation to have an internal information system but can use the one provided by a service provider, simply through a network system. In this way the supplier can distribute his software system directly through the Internet, giving the organization in question the possibility of access, processing and storage directly on the Cloud, which acts as a kind of working platform. Many approaches to the Cloud is developing, with the consequent birth of a myriad of dizzying anagrams, like, for example, Saas, Paas, Daas, Iaas, and so on with a long list. However, what should be emphasized are the achievable results: cost reduction and the guarantee of a reliable and innovative service are ensured[4].

1.2 Information System main components: a conceptual model

To conclude the chapter, it is necessary to spend some words speaking about the main components constituting an IS. *Figure 2* below shows a conceptual model of components and activities characteristic of an IS. It is based on five key elements (Marakas & O'Brien, 2013):

1. Data resources;
2. Hardware resources;
3. Software and procedures;
4. People and human capital;
5. Network resources.

[2] The National Institute of Standards and Technology (NIST) (originally the National Bureau of Standards (NBS)) is a government agency in the United States of America dealing with technology management. It is part of the Department of Commerce and its task is to promote the American economy through collaboration with industry in order to develop standards, technologies and methodologies that promote production and trade. See wikipedia.org

[3] See www.nist.gov

[4] See www.zerounoweb.it/cloud-computing/cloud-anzi-saas-paas-daas-e-iaas-significato-e-guida-ai-vantaggi-dell-on-demand

These five components build up a complex system able to perform input, processing, output, storage and control activities that transform the input data resources into final information, ready to use.

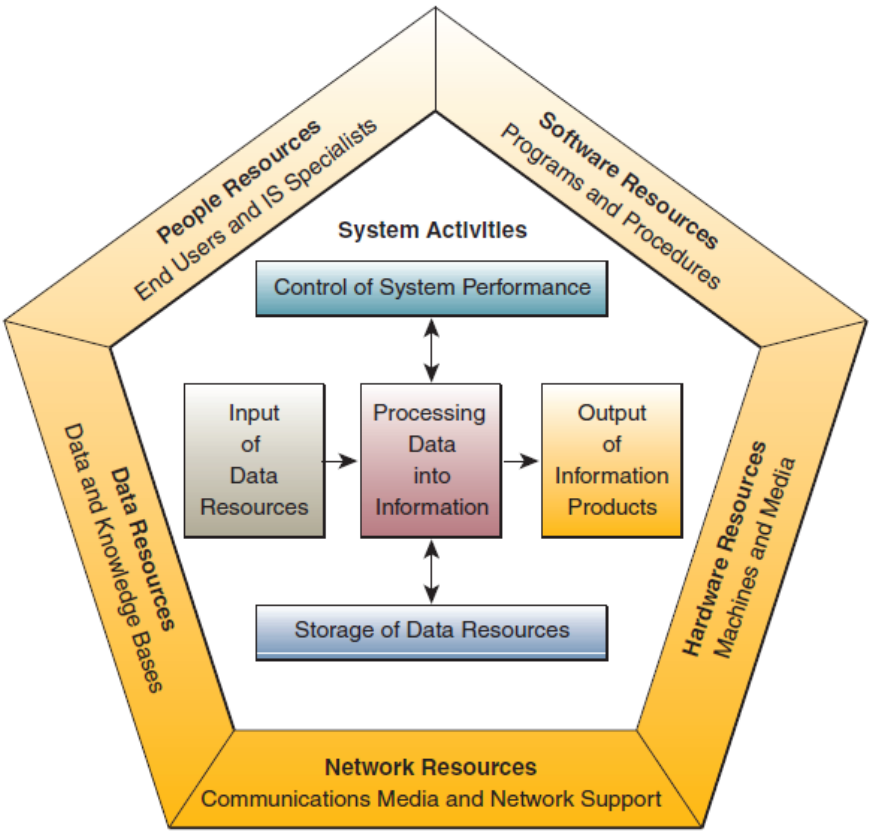


***Figure 2*: information system model in (Marakas & O'Brien, 2013)**

Data resources

Every expertise in the data field, from managers to IT specialists, has realized how data has begun one of the most important resource in a business patrimony: data should be thought, seen, and valued as any other asset from which to draw the major benefit for all business stakeholders. In the moment in which this new role is attributed to data, several changes occurred in the organizations. Many data that were previously only collected as a result of a transaction are now stored, processed and analyzed in order to extract previously unimaginable relations among them. Data resources have thus become a fundamental pillar of companies, and their effective management is considered an integral part of the strategic vision, as we will see later when we will go deeper into the topic of Data Governance.

Data collection must represent a fluid and streamlined process, therefore the selection of which data a company needs represents the first real step of the information process. This means the organization must be able to discern which events and transactions are worthy to be kept in consideration: a carpet collection, mostly if data collected are of little value for business objectives, would create disorder, management difficulties and enormous waste of resources.

Hardware resources

In the common imagination, when we think to an information system, hardware resources are perhaps the most obvious, consisting of computers, smartphones, printers, routers, switches, databases, etc. However, it is useful to point out to the reader how often the attention focuses only on the digital components, forgetting that an information system exists independently of digitalization and that tools apparently not worthy to be notice, like a pen or a sheet, are in all respects constituent elements of the hardware part.

Specifically, hardware resources include all physical devices and objects used for information processing. Machinery such as computers, and “media” on which information is recorded, such as CDs, keys or sheets of paper, are included in the category.

Software and procedures

Software resources include the whole set of operating instructions most commonly called “programs”, which directly control the physical hardware of the system, and the so called “procedures”, that are process instructions needed by system users. It is worth pointing out that the concept of software is separated from the one of computer, so even systems that do not use digital computers have a certain level of software component. From the most ancient systems to the present day, all systems need software resources in the form of information processing instructions and procedures to acquire, process and disseminate information to all users: such instructions and procedures are not embedded in computer component but rather resides in human capital.

It was stressed that procedures are an important component of the software resources of a system. They are the vehicle through which the means, people and organization as a whole are coordinated to achieve the information objective. At the base of the procedures is the organizational culture: members must feel part of a single and cohesive entity, having a logic of thought well targeted, shared by all.

A strong corporate culture, the “commitment”, is a key point on which the success of the company depends. In fact, if every actor puts his own goals ahead of those of the company, the efficiency of the organization will suffer heavily.

It is quite clear the enormous importance which resides in human capital when it comes to information systems. This significance will be the protagonist of the next paragraph.

People and human capital

People are the real essential ingredient for the success of an information system. However complex it may be, an IS can never be fully automated and dependent on IT tools. Thus, perfect coordination of technical and human variables should be at the center of the attention of companies, as a complete computerized information system is not able to replace human attitudes, such as experience, social relationships and values. Through interpersonal relationships, such as participation in meetings, reunions and assemblies it is possible to extrapolate information that it would not be possible to know and find by automating an information process or with the help of IT tools and means. Hence, the human component and relational is absolutely essential to catch specific information or particular signals. The characterizing element of human capital in IS finds voice in various authors in literature, one of whom is Buckingham, who defines an IS as "a system that collects, processes and distributes information concerning an organization so that the information is accessible and useful to those who wish to use it, including among them: managers, employees, customers and citizens” (Buckingham, 1986).

Undoubtedly, the information system must be built on the needs of corporate groups, in order to make it usable and functional. As already said, it is indisputable that members must perceive it as an ally. To this end, user-friendly interfaces can determine part of the success, as a complex and unintuitive system could result in staff rejection.

Network Resources

Modern IS allow companies to implement successful e-business and e-commerce activities, on which they base a large part of their day-by-day activities: this would be impossible without the support of network resources. The Internet, intranets and extranets allow companies to make information travels inside and outside the company instantly and to open to a world of opportunities

that were nothing more than a utopia a couple of decades ago. The telecommunication networks are made up of computers, communications processors and a myriad of other devices managed by appropriate software.

Specifically, we can identify two main elements constituting network resources:

• Communications media: refer to the supporting structure of the networks over which the electrical signals constituting data and information literally travel. Examples are twisted pair cables, coaxial cables, optical fibers, wireless technologies.

• Network infrastructure: includes all hardware, software and technology resources for data processing necessary for the operation of telecommunications networks. Examples are communication processors, such as modems or routers and communications control software such as Internet browsers packages.

Chapter 2
INTRODUCTION TO DATA GOVERNANCE

In the so-called 'Data Economy', data represents a strategic asset for organizations. The advent of social network phenomena, mobile technologies and lastly IoT, has led to the daily production of an impressive amount of data and information. This patrimony, often unstructured, offers great opportunities to companies that know how to exploit it, but at the same time adds to the databases already present in organizations, exponentially increasing complexity and risks of government and management. Despite this awareness, however, it is necessary to highlight that what it means to treat data like any other resource is a topic not sufficiently discussed in the literature, and most companies do not reserve the same importance to data as it does to other resources.

2.1 The new Valuable Asset

Assets valuation has evolved over the centuries, up to the most modern times, in which there is a growing tendency to attribute an accounting value also to intangible assets, such as patents, copyrights, trademarks and brands (Fleckenstein & Fellows, 2018). An interesting study regarding the correlation between intangible assets and the overall value of a company was conducted by Blair, of the Brookings Institution. The study shows that in 1978, only 20% of the corporate value was attributable to intangible assets. Exactly 10 years later, in 1988, this value had grown up to 55%, thus exceeding the value determined by tangible assets (Blair, 2001). And, not surprisingly, this value has continued to grow up to the present day. Several authors have expressed their positions on the importance of these assets in determining the company value, such as (Harrison & Sullivan, 2006). Also, analyzing the guidelines drawn up by the IRS[5] for valuing intangible assets[6], you see that "technical data" are listed under this voice.

To date, there is no agreement about how to evaluate data, but some authors have tried to make their proposals. Ono example is the paper by Daniel Moody

[5] IRS stands for Internal Revenue Services, the US government agency responsible for collecting taxes

[6] Internal Revenue Manual—4.48.5 Intangible Property Valuation Guidelines, Web. Autumn 2014. See www.irs.gov/irm/part4/irm_04-048-005

and Peter Walsh (Moody & Walsh, 1999) showing different accounting valuation models based on cost, market value, and revenue potential and concluding that the best cost approximation of data is based on future cash flows. Another example is Tony Fisher's book (Fisher, 2009). Actually, the book does not propose an approach about how to evaluate data, but rather highlights how well-applied data quality and data governance lead to greater profit, with higher revenues, lower costs and risk mitigation, thus making the difference between successful organizations and the rest. Another author who highlighted how properly implemented data management leads to a decrease in risks and costs is Peter Aiken (Aiken & Billings, 2013). Also interesting is the contribution proposed by Douglas Laney, who introduced the concept of 'Infonomics', in order to take you beyond thinking and talking about information as an asset to valuing and treating it as one. 'Infonomics' provides the foundation and methods for quantifying information asset value and tactics for using information as your competitive edge to drive growth. One of Douglas's most interesting proposals is to draw up an internal balance sheet to keep track of the value of data (Laney, 2012).

Despite the difficulty in achieving a common vision about the evaluation of the data as an asset, as anticipated in the previous chapter, data represents the basis of the information systems, a necessary element for the decision-making process and the consequent persistence of the organization on the reference market. The survival can be declined, in this seat, in a twofold order of reasons: on the one hand, the achievement of strategic and business corporate objectives, on the other, a correct estimate of the risks and the structuring of an information system compliant with the regulation (Cheong, et al., 2007).

Without data, in fact, strategic decisions are nothing more than hypotheses. It is only thanks to data, and above all to their transformation into information, that strategic decisions are made based on solid ground, becoming targeted and precise. Henri Poincaré[7] said: "Science is made of data like a house is made of stones. But a pile of data is no more science than a pile of stones is a real home. " This to point out that collecting data is only the first phase of a process which must then continue with the interpretation of them and their translation into information of value for the company.

To this end, it is necessary to build a framework that allows the company to manage the data in its entirety, in every aspect, in order to be able to transform it into useful information. This structure is called Data governance, which will be discussed later in this chapter and, in more detail, in Chapter 4.

[7] Jules Henri Poincaré was a French mathematician, theoretical physicist, engineer, and philosopher of science (1854 –1912).

2.2 Inside the World of Big Data

By now, when it comes to data, especially if referring to the banking world, one cannot fail to mention the phenomenon of Big Data.

We are at the beginning of a new era of economic and social development, the result of a continuous and unstoppable innovation process, which has characterized the development of the Internet over the past two decades. Innovative technologies, platforms and systems such as Cloud Computing, IoT, Big Data & Analytics, Blockchain, Artificial Intelligence, Augmented Reality & Virtual Reality, Advanced robotics & 3D printing and 5G constitute the new enabling tools of the digital economy which, thanks to its pervasive diffusion in all sectors, promises to give rise to a new era of economic and social development and more generally to a new and more evolved phase of human existence. One of the leading factors guiding the development of the digital economy is represented by data: to exploit this new resource companies must equip themselves with adequate analytical tools. But this is not enough, the real turning point is in fact a necessary change in the economic and cultural structure of digital companies.

The prevailing literature tends to focus on the quantitative aspect and therefore concerning the volume of data, from which the term Big Data comes from. However, as we will see in a while, volume isn't the only important feature. The speed at which data is generated and made accessible is equally impressive. The phenomenon of big data is also defined because of the ability to analyze a variety of unstructured data sets from different sources. Pervasive digitization places data as a catalyst for innovation and it is therefore necessary for organizations to transform themselves aiming at becoming more and more data driven. Data can positively affect the improvement of a company's products or services; they can allow companies to take advantage of new business opportunities; finally, they can also be used to better target potential customers, providing them with personalized services or products (ItMedia Consulting , 2018).

At the basis of the birth of Big Data three conditions are substantially intertwined. The first is an increase in the availability of information, which in turn depends on IoT and ‘datafication’. As for IoT, there would be a Pandora's box to open, but it eludes from this elaborate, so I will limit myself to underlining how the interconnection of more and more machines and devices on the network that exchange information, will increase dramatically the information to be processed. As regards *datafication*, first care must be taken not to confuse it with digitization, which consists in the process of converting analog signals into digital format. Instead, datafication consists in the process of converting a given phenomenon into a quantitative form, so that it can be tabulated and analyzed.

Examples can be found in social networks, where Facebook is transforming relationships into data and Twitter converts feelings into quantitative data instantly.

The second factor that led to the explosion of big data is related to the exponential growth of the processing and storage capacity and the development of artificial intelligence systems.

The third and final reason, strictly connected to the other two, is the decrease in costs for the collection, storage and processing of the data.

Emerging disciplines, as the one of Big Data, often experience a lack of agreement regarding the definition of core concepts. Indeed, the level of consensus shown by a scientific community when defining a concept is a proxy of the development of a discipline (Ronda-Pupo & Guerras-Martin, 2012). The quick and chaotic evolution of Big Data literature has impeded the development of a universally and formally accepted definition for Big Data. In fact, although several authors proposed their own definitions for the term, like (Beyer, 2012); (Dijcks, 2013), (Mayer-Schönberger & Cukier, 2013), (Schroeck, et al., 2012), none of these proposals has prevented subsequent works from modifying or ignoring the previous definitions and suggesting new ones. Such lack of agreement and homogeneity, although justified by the relative youngness of Big Data as a concept, limits the proper development of the discipline (De Mauro, et al., 2016). However, what is certainly found in all the definitions is that Big Data represent "that set of data that cannot be collected, obtained, managed and analyzed with traditional information technologies and databases".

After a profound literature review, in 2016 De Mauro A., Greco M. and Grimaldi M. (De Mauro, et al., 2016) tried to give a concise and complete definition of Big Data, in order to take into consideration the various previously proposals made, and the result was the following definition:

“Big Data is the Information asset characterized by such a high volume, velocity and variety to require specific technology and analytical methods for its transformation into value”.

The definition shows what have been defined as the 4V of Big Data, i.e. Volume, Velocity, Variety and Value, which are widely shared within the scientific community[8]. They will be briefly discussed later in the paragraph.

As anticipated, the use of Big Data has the potential to transform dramatically traditional businesses since it may offer them even greater opportunities for competitive advantage. The big data of this revolution is far more powerful than the analytics that were used in the past. We can measure and therefore manage more precisely than ever before. We can make better predictions and smarter decisions. We can target more-effective interventions and can do so

[8] See, as an example, https://www.gartner.com/en

in areas that so far have been dominated by gut and intuition rather than by data and rigor. As the tools and philosophies of big data spread, they will change long-standing ideas about the value of experience, the nature of expertise, and the practice of management. Smart leaders across industries will see using big data for what it is: a management revolution. But as with any other major change in business, the challenges of becoming a big data– enabled organization can be enormous and require hands-on or in some cases hands-off leadership. Nevertheless, it's a transition that executives need to engage with today (McAfee & Brynjolfsson, 2012).

A fundamental thing that business executive must understand, is that "analytics and "Big Data" is not the same things. Although they era strictly connected, since both have the aim to extract information and relations among data and transform that into business advantage, three main differences differentiate them. Those differences can be found in the definition of Big Data give above: Volume, Velocity and Variety.

Regarding the *Volume*, it is worth noting that as of today, about 3 exabytes of data are created each day and in the last two to three years have been created 90% of the entire available data[9]. Ever more impressive, more data cross the internet every second than were stored in the entire internet just 20 years ago. Using the internet traffic reports of Cisco[10] we can estimate that the entire digital universe is roughly 44 zettabytes. If the estimate is correct, it means that we have 40 times the number of observable stars in the universe available in bytes. And, following the trend, 463 exabytes of daily information could be estimated for 2025[11]. Just to make some clarification, and probably leave the reader open-mouthed, an exabyte is a number with eighteen zero and a zettabyte is equal to a thousand of exabytes!

Clearly, such impressive amount of data gives companies an opportunity to search for previously unimaginable relationship among data that allow them to customize product and service for the single person.

Figure 3 shows what happens on the Internet every minute which gives an insight about the amount of data that are produced every minute by the human being. Looking at the comparison between 2018 and 2019, it is also evident how this trend is increasing rapidly.

[9] See https://www.cefriel.com/it/
[10] See https://www.cisco.com/
[11] See https://www.infodata.ilsole24ore.com/

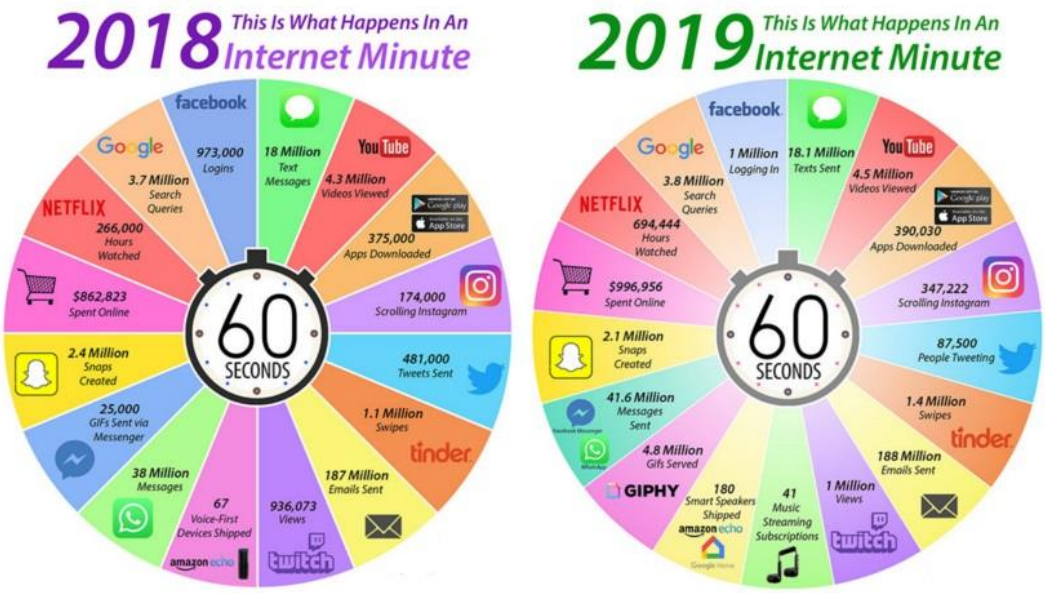


***Figure 3*: what happens in an Internet minute in 2019?**

For what concern *Velocity*, you see that for many applications, the speed of data creation is even more important than the volume. Real-time or nearly real-time information makes it possible for a company to be much more agile than its competitors. As we will see in the next chapter, one of the major trend for the future of services is the willingness of customers to receive them on demand and be satisfied at the exactly same time in which they request for them: the only way to reach such paradigm is to extremely improve the velocity of acquisition, elaboration, analysis and transmission of data.

Finally, the *Variety*. Big data takes the form of messages, updates, and images posted to social networks, readings from sensors, GPS signals from cell phones, and more. Many of the most important sources of big data are relatively new. The huge amounts of information from social networks, for example, are only as old as the networks themselves: Facebook was launched in 2004, Twitter in 2006. The same holds for smartphones and the other mobile devices that now provide enormous streams of data tied to people, activities, and locations. These devices are already ubiquitous and in the future such trend will increase exponentially due to all the new technologies that have been previously listed. Thus, the structured databases that stored most corporate information until recently are ill suited to storing and processing big data. Some insight about how to manage this new form of data can be found in (Castanedo, 2018), which explains why NoSQL multi-model documents-oriented approach is the best to deal with Big (unstructured) Data given its great flexibility. At the same time, the steadily declining costs of all the elements of computing, namely storage, memory, processing, bandwidth, and so on, means that previously expensive data-

intensive approaches are quickly becoming economical. As more and more business activity is digitized, new sources of information and ever-cheaper equipment combine to bring us into a new era: one in which large amounts of digital information exist on virtually any topic of interest to a business. Mobile phones, online shopping, social networks, electronic communication, GPS, and instrumented machinery all produce torrents of data as a by-product of their ordinary operations. Each of us is now a walking data generator.

The other V that has been highlighted is *Value*. It was already present in analytics, but with Big Data its significance has become even greater. As widely said, Big Data contains inside them an unimaginable amount of information that is waiting to be discovered and that will richly reward the companies that will be able to extract such great source of power.

In conclusion, it is worth to highlight that in the last years many other Big Data dimensions are coming out like Virality or Volatility, but only one them is shared by the expert in the field, like for example (Williamson, 2015), bringing the "Vs" of Big Data from 4 to 5: this is the Veracity. Simply talking, data *Veracity* indicates how accurate a data set may be. In the context of Big Data, however, it takes on a bit more meaning. More specifically, when it comes to the accuracy, it is not just the quality of data themselves but also how trustworthy the data source, type, and processing of them is[12].

Figure 4 collects the 5V of Big Data:

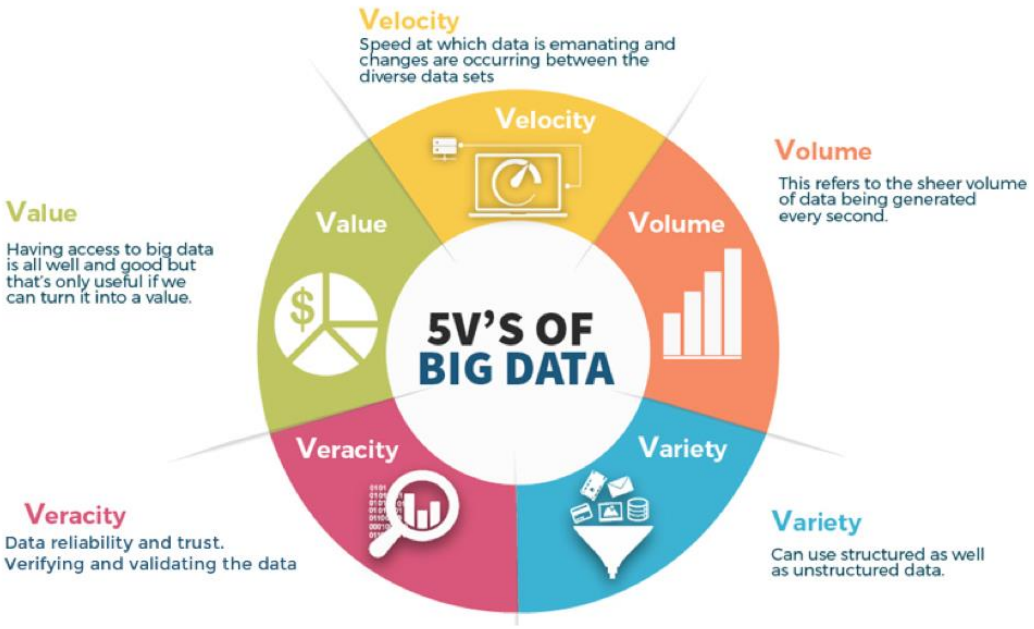


***Figure 4*: the 5V of Big Data**

[12] See https://en.wikipedia.org/wiki/Big_data

2.3 Data Governance: a definition still in progress

At this point is more than clear that organizations in general are facing a new challenge consisting in a smart and proficient management of data. Challenge that will become even harder in the future, but that can lead to profitability business opportunities if addressed in the right way.

Only a careful policy of data management and a well-designed Data Governance model supported by adequate technological tools will allow to make the most of business opportunities while guaranteeing the quality and compliance of an increasingly rich corporate data assets. It is fundamental to understand how a careful and structured Data Governance structure not only leads to being compliant with increasingly restrictive directives by regulatory bodies, but also, and above all, allows you to have greater control over company data with highly positive consequences in term of increase in revenues, decrease in costs and risk mitigation.

A standard definition of the term *Data Governance* (DG) can be found neither in the research community nor in the practitioners' community dealing with information systems. However, proposals to define it have in common that Data Governance refers to the allocation of decision-making rights and related duties in the management and use of enterprise data. According to (Weber, et al., 2009) for example, Data Governance specifies a structural framework for decision-making rights and responsibilities regarding the use of data in an enterprise. Khatri and Brown (Khatri & Brown, 2010) see Data Governance referring to the assignment of decision- making rights with regard to an enterprise's data assets.

It is known, how an absence of a solid government of data can cause major losses to an organization, not being able to accurately determine the value of the data it manipulates. Data, in fact, influence not only the purely operational activities, but, as widely seen in the previous paragraphs, it significantly impacts the strategic choices made by top management (Tallon, et al., 2013). Without taking into account that data are increasingly beginning to be considered an intangible corporate asset of primary value, like physical and other intangible assets, and their governance can only be more critical than ever.

Several authors have dedicated themselves to the topic, such as (Cheong, et al., 2007), (Hagmann, 2012); (Kamioka, et al., 2016) but despite this increase in attention, a study conducted by (Holt, et al., 2015), brought to light as less than half of the participants within the global community of database and data professionals had data governance policies. This only confirms that the studies on the topic are still at a preliminary stage, but above all that the breadth and delicacy of the issues considered leads to a difficulty in the discussion.

So, let's try to give an overview of what the current situation of Data Governance is in both the world of research and in the one of practitioners. To this end, an article recently published by Alhassan Ibrahim in 2018 meets our need (Alhassan, et al., 2018).

The article highlights how both academics and practitioners have developed different data governance models, which allow us to understand what the boundaries of data governance are today (Cheong & Chang (2007), Guetat & Dakhli (2015), Khatri & Brown (2010), Lajara & Maçada (2013), Wende & Otto (2007)), and part of the associated data-related activities (Panian (2010), Weber, et al. (2009)). However, what results from analyzing the literature is that none of the proposed models considers Data Governance activities, i.e. what practically is needed to implement a DG model: as a consequence, practitioners cannot use such research to implement a DG program.

Without going into specific and detailed results, what Ibrahim discovered in his research, is that there is still a great deal of ambiguity as to what it means to draw up a DG model, so much so that in the literature there seems to be no publication explaining in detail the activities required to conduct a Data Governance program. The first phase, the one of defining roles and responsibilities on data, is reflected in a good part of the articles analyzed, confirming the definition of Data Governance given above. However, when we move to the second and third phases, respectively of implementing and monitoring, not only the research decreases considerably, but the concepts and procedures become more ambiguous, a phenomenon found more in the scientific community than among the practitioners. This demonstrates how much still needs to be done in terms of DG research.

Despite the need to deepen the topic, there is a great consensus that DG helps to answers to three main questions (Khatri & Brown, 2010), (Weber, et al., 2009):

- What decisions about company data should be made at company level?
- What roles are involved in the decision-making process?
- How are roles involved in decision making?

Regarding the first question, several answers have been proposed. One of these is that of Khatri and Brown (Khatri & Brown, 2010), which has experienced great success. According to the authors, decisions on DG refer to some fundamental principles of data management, data quality requirements, metadata management, management of data access requirements and data lifecycle management.

Moving to the roles involved in the decision-making process, you see that they are essentially five: (1) Data Governance Committee; (2) Chief Data Officer;

(3) Data Owner; (4) Business Data Steward; (5) Technical Data Steward. Let's briefly analyze them.

First, the implementation of a Data Governance program begins with the appointment of a steering committee, called Governance Committee. The Data Governance Committee must reflect both the interests of IT and those of the Business, thus should be composed by members coming from both the sides. It represents the strategic guide of the DG process, and specifically defines the responsibilities, the ways in which they are assigned to the personnel and the objectives to be achieved. Moreover, it takes care of evaluating any project initiatives necessary to improve the governance of data.

Usually, the Chief Data Officer (CDO) is the head of the Data Governance Committee. It is the responsible for transforming data into information but still does not find a shared allocation by organizations in a single corporate function. The CDO represents the guide and coordination of Data Governance activities by putting into practice the guidelines decided at Committee level and transferring them to the figures placed at a higher level of granularity (Data Owner and Business Data Steward). Lastly, the CDO coordinates the Data Owners. Responsible for the information relating to specific corporate areas, the Data Owners (DOs) are appointed in the manner defined by the Data Governance Committee and the Chief Data Officer. The role of the DO is fundamental in the coordination of the Business Data Stewards, real subjects in charge of the data from the point of view of the Business. The role of the Business Data Steward includes the effective implementation of the data governance metrics on the corporate areas of competence. The choice of the perimeter for which a single Business Data Steward is in charge remains a discretionary choice of the organization and depends heavily on the number of data and information available to the organization itself. The reflection of their role in the IT area is represented by the Technical Data Stewards. The Technical Data Stewards provide support and are associated with specific systems, applications, data stores, and technical processes. Technical Data Stewards are the people to turn to in order to understand how the data is created, manipulated, stored, and moved in technical systems (Plotkin, 2013). It should be emphasized that not all figures are required in every organization, but rather depends on the amount of data processed and the size of the structure. For example, the figure of the Business Data Steward will not always be present, since its role is to support the Data Owner, having the first a vision of greater granularity than the latter.

The third and final question refers to the link between roles and decision-making areas: Data Governance must assign authority and, consequently, responsibility. In order to design DG for individual enterprises, a number of authors (Weber, et al. (2009), Loshin (2008)) proposed to use the RACI notation.

RACI is an acronym standing for Responsible, Accountable, Consulted and Informed. A RACI model is a two-dimensional matrix, also called Responsibility Assignment Matrix (RAM), listing the tasks to be performed along the rows and the roles along the columns. Each cell in the matrix is populated according to[13]:

R = Responsible

Those who do the work to complete the task. There is at least one role with a participation type of responsible, although others can be delegated to assist in the work required.

A = Accountable

The one ultimately answerable for the correct and thorough completion of the deliverable or task, the one who ensures the prerequisites of the task are met and who delegates the work to those responsible. In other words, an accountable must sign off, and thus approve, work that responsible provides. There *must* be only one accountable specified for each task or deliverable.

C = Consulted

Those whose opinions are sought, typically subject matter experts and with whom there is two-way communication.

I = Informed

Those who are kept up to date on progress, often only on completion of the task or deliverable; and with whom there is just one-way communication.

To conclude the chapter, it is worth noting the difference between two term that are usually used interchangeably but they represent quite different things: those terms are Data Governance and Data Management. The main difference between the terms ‘governance’ and ‘management’ is that governance refers to the decisions that must be made and who makes these decisions in order to ensure effective management and use of resources, whereas management involves implementing decisions (Fu, et al., 2011). A representation that can help us is the one from Otto (Otto, 2011):

[13] See https://en.wikipedia.org/wiki/Responsibility_assignment_matrix

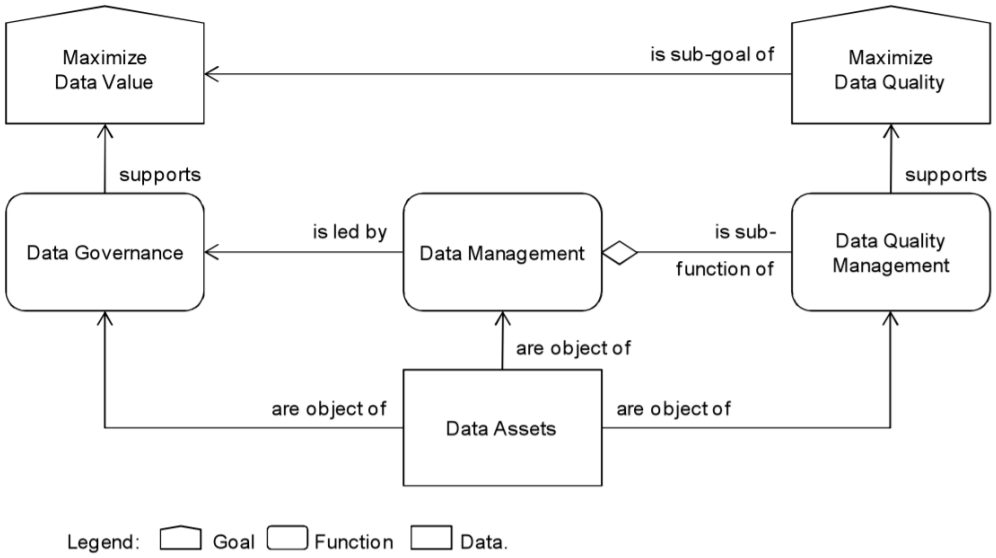


***Figure 5*: fundamental concepts of Data Governance and Data Management, (Otto, 2011)**

Looking at *Figure 5*, you see how Data Governance has the objective of maximizing the value of the data as an asset in the company. In fact, as stressed, data are increasingly considered an asset like other intangible and physical assets. Data Quality Management has the objective of maximizing the quality of data, where with term Data Quality you usually mean the ability of data to be suitable to meet the recipient's needs. Wang (Wang, 1998) for example, defines Data Quality as data's "fitness for use". Data Quality Management is part of the wider Data Management, which, as said, should not be confused with Data Governance. The second, in fact, represents the guiding function of the first, specifying which decisions must be made in Data Management and who must make these decisions. Instead, Data Management ensures that decisions are made and that appropriate actions are implemented.

In the following, the concept of Data Governance will be discussed in depth, but first in Chapter 3 an overview of the Financial Services industry will be conduct. In fact, since the case study that will be later analyzed has been implemented in a bank I think that introducing the banking industry, the tendency of the sector and its relationship with data could be useful to have a wider view on the issue that will be exposed.

Chapter 3
THE FINANCIAL SERVICES INDUSTRY

Analyzing the literature, a unique and widely shared definition of financial services does not exist. Surely, this is due to the breadth of the topic and the different points of view from which it is possible to investigate it. What is common to many definitions, however, is the fact that the concept of financial services includes all the economic services provided by banks, insurance companies, consumer credit companies and investment funds.

Yet, for the purposes of the following elaborate, the definition just given is quite useless. The focus will instead be on the interaction that sees financial services and technology as protagonists. The massive and pervasive use of technology in the financial sector is in fact the basis for the increasingly necessary introduction of Data Governance in the organizational models of banks, insurance companies and financial providers in the more general sense of the term.

3.1 Fintech: a new term for a long time relationship

The world of *Financial Services* has historically been very open to the use of technologies to provide the customer with a 360-degree experience and allow him to facilitate his interaction with finances. However, in recent years this relationship has seen an explosion, which is literally taking the way of financial services into a new era: this revolutionary phenomenon is called FinTech (Gomber, et al. (2018), Chisti & Barberis (2016)). FinTech is heavily affecting the business models of companies, the ways in which they interact with the customer and the regulatory bodies. So, let's try to give a more concise definition to such a term.

Analyzing the literature, what it turns out is that *FinTech* is a term with a broad meaning that refers to the use of a series of advanced technologies in the financial and banking sector. What is missing, however, is a univocal and generally accepted definition, even at the regulatory level. Several Authorities and Bodies that have addressed and analyzed the characteristics of this branch have

tried to provide a definition. The Financial Stability Board[14], for example, defines FinTech as "a technology-enabled innovation in financial services that could result in new business models, applications, processes or products with an associated material effect on the provision of financial services"[15]. ESMA[16] defines it as "a type of financial innovation that relies on Information Technology to function, e.g. internet, cloud etc. and that can result in new business models, applications, processes, products or services with an associated effect on financial markets and institutions and provisions of financial services" (Armstrong, 2018). The European Commission[17], on the other hand, provides a more generic definition of FinTech, understood as " the impact of new technologies on the financial services industry, which includes a variety of products, applications, processes and business models that have transformed the traditional way of providing banking and financial services"[18]. These are specifically vague definitions, in that the attempt to identify with certainty the boundaries of a phenomenon still in progress appears premature.

Despite a unique definition, it is not yet present, surely what all definitions have in common is the fact that FinTech is the branch of technology at the service of financial industry. This means that, although the term FinTech has only recently entered the common language, mainly due to the disruptive revolution that is upsetting the banking world, its birth is dated a long time ago. In fact, the relationship between these two branches was born two centuries ago and over the years has seen exponential growth, up to the present day, where FinTech represents one of the driving forces of the global market.

3.1.1 The evolution of FinTech

Three evolutionary eras referring to FinTech can be distinguished (Arner, et al., 2016). The first of them, defined *Fintech 1.0*, was placed between 1866 and 1967, with the introduction of the telegraph and the subsequent laying of

[14] The Financial Stability Board (FSB) is an international body that monitors and makes recommendations about the global financial system. See wikipedia.org/wiki/Financial_Stability_Board

[15] See www.fsb.org/2019/02/fsb-report-assesses-fintech-developments-and-potential-financial-stability-implications

[16] ESMA is an independent EU Authority that contributes to safeguarding the stability of the European Union's financial system by enhancing the protection of investors and promoting stable and orderly financial markets. See esma.europa.eu/about-esma/who-we-are

[17] It promotes the general interest of the EU by proposing and enforcing legislation as well as by implementing policies and the EU budget. See ec.europa.eu/info/index_en

[18] See ec.europa.eu/info/business-economy-euro/banking-and-finance/fintech_en

transatlantic cables, which, thanks also to the maximum development of railways and steamships, allowed a first real revolution in the world of finance. The information could now travel from one part of the globe to another, resulting in a disruptive change for the time, as can be seen from the words of J.M. Keynes (Keynes, 1920):

> *"The inhabitant of London could order by telephone, sipping his morning tea in bed, the various products of the whole earth, in such quantity as he might see fit, and reasonably expect their early delivery upon his door-step; he could at the same moment and by the same means adventure his wealth in the natural resources and new enterprises of any quarter of the world, and share, without exertion or even trouble."*

During the 1950s, it was the turn of the first credit card, which was initially used for travel and entertainment. Its characteristic was to be a debit card, as the payments were totally postponed to the end of the month. But it was only in 1967, with the introduction by Barclays[19] of the first ATM in Britain, that a new thriving era for FinTech was born.

Located between 1967 and 2008, year of the well-known modern economic crisis, the second era takes the name of *Fintech 2.0*. Together with ATM, the other pivotal innovation of the period is the launch of the first computers on the market. The parallel work of these two new forces made it possible to move from an analogue to a digital industry concept. The revolution occurred in all areas of financial services, starting from payments, where the UK and the USA saw the birth of two major Clearing Services, respectively BACS[20] and CHIPS[21], passing through the securities area, with the birth of NASDAQ[22] in 1971, which made it possible to completely digitize securities trading, up to the consumer area, with the introduction of online banking (or internet banking)[23]. However, what really emerged as the driving force of the period was the advent of the Internet, starting

[19] Barclays plc is a British multinational investment bank and financial services company, headquartered in London. See https://en.wikipedia.org/wiki/Barclays

[20] Bacs Payment Schemes Limited (Bacs), previously known as Bankers' Automated Clearing Services, is the organization with responsibility for the schemes behind the clearing and settlement of UK automated payment methods Direct Debit and Bacs Direct Credit. See https://www.bacs.co.uk/Pages/Home.aspx

[21]The Clearing House Interbank Payments System (CHIPS) is a United States private clearing house for large-value transactions. See https://en.wikipedia.org/wiki/Clearing_House_Interbank_Payments_System

[22] The Nasdaq Stock Market also known as Nasdaq, is an American stock exchange located at One Liberty Plaza in New York City. See https://en.wikipedia.org/wiki/Nasdaq

[23] Online banking, also known as internet banking or web banking, is an electronic payment system that enables customers of a bank or other financial institution to conduct a range of financial transactions through the financial institution's website. See https://en.wikipedia.org/wiki/Online_banking

in 1995 when Wells Fargo began providing an online account checking service via the World Wide Web (Riggs, 2015). The maximum development of technology applied to payments and trading management and the consequent explosion of interconnections between different countries, brought to light a clear increase in risk: it is in this period that important e-banking regulations began to emerge.

The beginning of the third era does not have a well-defined date, as it is connected to the change in consumers' perception about what are the institutions empowered to provide financial services. What is certain, however, is that the 2008 financial crisis marks a watershed between *FinTech 2.0* and *FinTech 3.0*. In that year, in fact, the world of financial services underwent a real disruptive revolution, due to financial, economic and political changes, which allowed a new and fresh generation of competitors to establish the modern paradigm that we now call FinTech. Primarily, consumers lost confidence in banks by starting to consider other types of providers who are able to manage their savings and the financial transactions they needed. Secondly, the financial crisis caused a breakdown in the system and a million people lost their jobs: this resulted in the creation of a large mass of highly trained and jobless people, who dedicated their skills to the development of FinTech 3.0. One major consequence was a tightening of regulations that required greater protection against risks from the banks and credit services, which on one hand led to a safer market, but on the other took away further dynamism from the banks, limiting their ability to compete and favoring the development of technological start-ups. An example of regulation is Basel III[24], but the directives that saw the light during the post-crisis period and which have contributed to defining what modern financial services industry looks like are numerous and the most important will be shown in the Chapter 4.

3.1.2 FinTech today

Today the term FinTech is on everyone's lips, and not only on a theoretical level. According to a recent study by Hampleton Partners[25], investments in companies and fintech startups by venture capital groups more than doubled last

[24] Basel III is an international regulatory accord published in 2009 by the Basel Committee on Banking Supervision that introduced a set of reforms designed to improve the regulation, supervision and risk management within the banking sector. See www.investopedia.com/terms/b/basell-iii.asp

[25] Hampleton Partners is a leading M&A and corporate finance consultancy for companies with technology at their core. See www.hampletonpartners.com

year reaching its all-time high[26]. In 2018 they touched the record of 31 billion dollars compared to 15 billion in the previous year. And if we consider also PE and M&A (Private Equity e Merge &Acquisition), the amount rises to 112 billion (KPMG, 2019). The exponential growth of the economic resources invested in FinTech is the result of the increasing trust that individuals and companies, from SMEs to corporations, place in the FinTech world. According to a 2019 study by Ernest & Young (Ernest & Young, 2019) , considering only customers, the global adoption, albeit with significant variations from one country to another, stands at 64%, starting from 16% in 2015 and passing from 33% two years later: the number of adopters has substantially quadrupled in the space of 4 years. Also with regard to SMEs there is an important surge in adoptions, which stands at 25% but have a clear upward trend.

Such high statistics are also due to the entry in the FinTech market by historical financial providers, such as banks or insurers, which in order to survive the destructive force of innovation have been forced to develop digital offers similar to those of the new market players, namely the 'challengers'. However, many FinTech startups can no longer be considered as such, since they are now very solid and structured players with important market shares. It is noted that the trend is growing rapidly, especially in emerging countries, with China and India clearly leading the ranking with an action level of 87%. The 64% of adoption is already significant, but if you think that in 2017 the forecasts for the 2019 were 52%, you can notice a staggering + 12%. *Figure 6* shows data relating to the countries where adoption has reached the highest levels.

[26] See www.forbes.com/sites/trevorclawson/2019/02/11/fintech-startups-suck-in-31bn-as-the-big-finance-players-step-up-rd-by-proxy

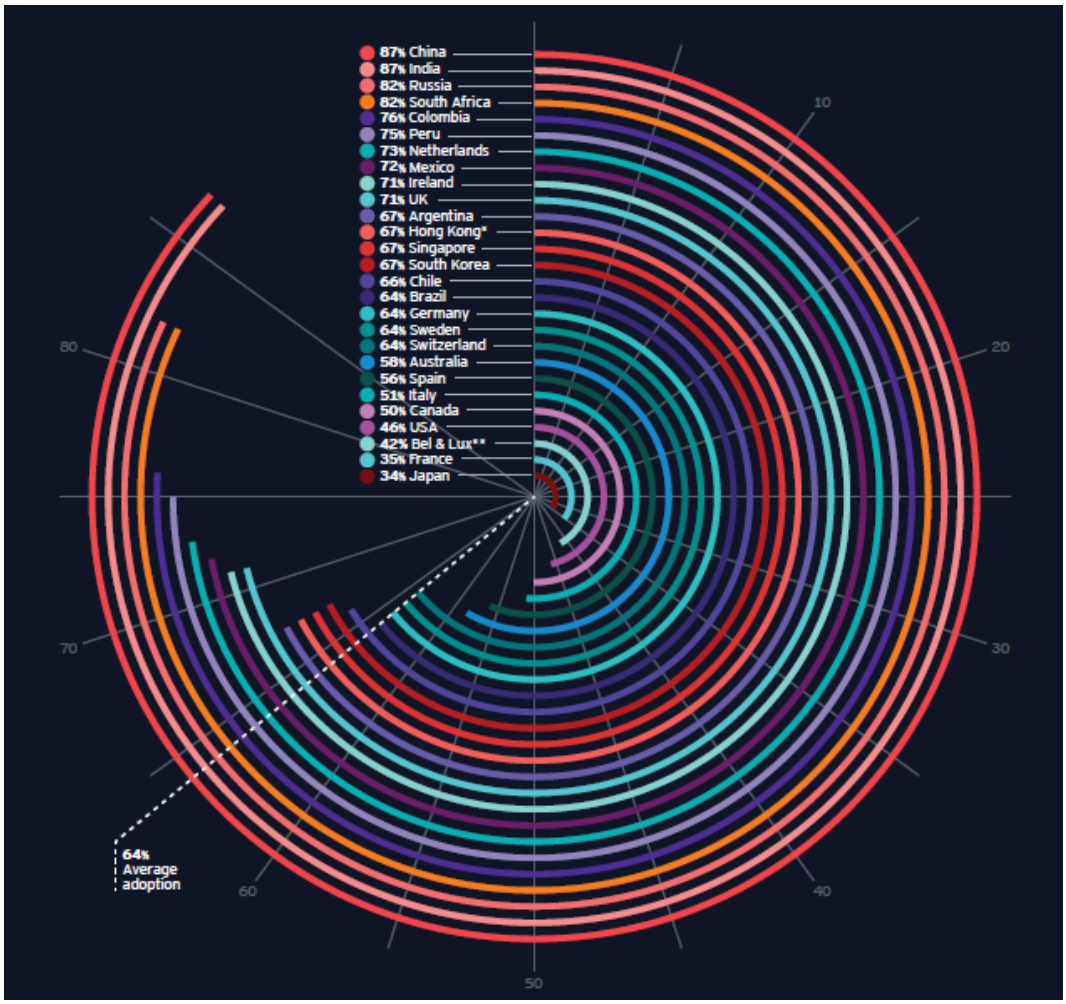


***Figure 6*: consumer FinTech adoption across 27 markets from Ernest & Young' s Global FinTech Adoption 2019, (Ernest & Young, 2019)**

An analysis of the level of knowledge about FinTech services among consumers gives further insight about the spread of FinTech. Considering the 5 macro-categories into which financial services can be divided, *i)* money transfer and payments, *ii)* budgeting and financial planning, *iii)* savings and investment, *iv)* borrowing and *v)* insurance, it is noted that by now the vast majority of the population is aware of the offer of services existing on the market. As for the first category, money transfer and payments, the best known one, the level of non-knowledge stops at just 4%. Also for activities still a bit far from customers, like budgeting and financial planning, the registered level of non-knowledge is low, 29%.

It is even more interesting to evaluate what is the level of market penetration by new technologies. Comparing the data from 2015, 2017 and 2019, we see the great upward trend that is characterizing all 5 categories, leaded by digital payments and insurance services. *Figure 7* shows the temporal comparison of the adoptions for each of the categories listed above.

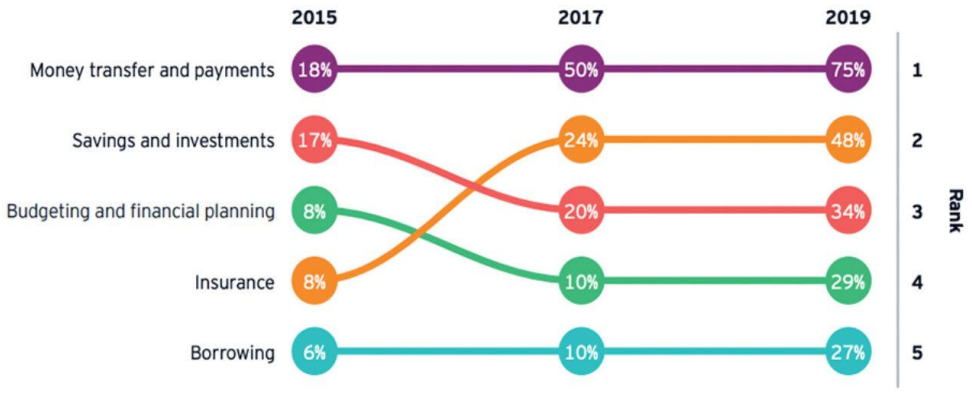


***Figure 7*: comparison of FinTech categories ranked by adoption rate from 2015 to 2019 from Ernest & Young' s Global Fintech Adoption 2019 (Ernest & Young, 2019)**

The high percentages recorded this year must not at all make us think that the market is saturated, for any of these categories, not even for the money transfer and payments. In fact, it is expected that in the future, not only will the penetration of existing services increase significantly, but also that new services will be continuously released on the market. This is also thanks to the most modern technologies such as artificial intelligence and the blockchain, which will be able to create totally revolutionary FinTech services.

Compared to the results coming from the 2017 survey, consumer needs have significantly evolved. Several services that were considered new and innovative just two years ago, are now perceived as basic features, a 'must-have': these have become necessary features to reach the client which no longer values them as 'something more'. A great evidence of this is the fact that in the old survey the most important requirement for catch up to a new customer was the ease that was experienced during the opening of the account, while today this requirement has dropped in positions, leaving in first place the attractive price. Basically, customers expect onboarding without friction, a sign of the fact that it has become a hygiene factor, no longer important for the purpose of attracting new adopters.

When it comes to financial services, a topic that has never been discussed enough is the trust that consumers place in providers. A service is something totally immaterial and as such it causes high uncertainty in the customers at the time of purchase. The only way to overcome this issue is to create a solid trusting relationship with the customer, influencing in a pervasive way the perceived quality of the service (Fiocca, 2013). Analyzing the 2019 survey results, you notice how consumers put little trust in challengers. If we consider the reasons why a person decide to remains a non-adopters, at the top of the list we find the

lack of knowledge in the product offered or the not totally understanding of it, but immediately after it is the lack of trust in new players: those who decide not to enter the world FinTech do so largely because they considers incumbents more reliable providers than the challengers. Actually, among adopters and non-adopters, there is a great tendency to turn to their own bank when they decide to try a new service.

However, another relevant trend is emerging stronger and stronger among adopters: those who use FinTech services tend to increasingly diversify the providers for different services. This is clearly expressed in the words of Alan McIntyre, senior managing Director Banking of Accenture and widely followed thought leader: “Banking is no longer monogamous. The idea that you have a single bank that you run your financial life through is going to become less and less common "[27]. So, it happens that a substantial part of the adopters entrusts the management of their services to five or more different providers, which can be challengers, incumbents, or organizations outside the financial services, which we will talk about shortly. This leads to a tough competition and, in turn, to lower costs and a better provision of services. The cause of this trend is, in large part, the word-of-mouth: many consumers are in fact influenced by family and friends in choosing a certain provider for a certain service. Phenomenon accentuated by the disproportionate use of social networks which obviously facilitates the conversation between various consumers: FinTech uses virality as its primary marketing strategy.

Given this complex and disruptive scenario, incumbents have been forced to enter the market. They must take into account the characteristics that consumers consider as product prerequisites, such as simplicity, transparency, frictionless, personalization and omni-channel. And once products with these characteristics are made, those players can leverage their massive and historic presence on the market and the confidence that consumers place in them to conquer a privileged place in FinTech industry. Indeed, there is no doubt that incumbents enjoy long-standing relationships with customers, connections with other peers and a global brand that challengers certainly cannot enjoy.

As already anticipated, and as we will see in greater depth in the next paragraph, a new competitive threat has emerged on the financial services market in recent times: non-financial services companies. Companies such as retailers or machine manufacturers are implementing a series of financial services, starting from their relationship with the customer, to create a wider and more complete product, which also adds complementary financial services to what they already provided to the customer. The threat to incumbents is evident, mainly because these companies are often very technological, and are able to provide a simple

[27] See thefinancialbrand.com/91736/fintech-challengers-banking-legacy-community-bank/

offer, at low costs and frictionless. Not to mention that they can also leverage the established long-term relationship with the customer. However, a very important factor in the digital age does not play in their favor. Much of the success of a Fintech service is based on the quantity and quality of data available through which it is possible to create a personalized product that fits with customer's needs. Not having this type of data, therefore, results in a service of little attraction for the customer and a clear competitive disadvantage. This is exactly the situation in which non-financial companies find themselves today. If in fact it is true that the willingness of consumers to share their information in exchange for better offers has never been so high, when it comes to sharing with this type of companies, consumers are more reticent, forcing new players to chase.

This continue to confirm the absolutely central importance that incumbents play within the FinTech industry: they represented the main point of contact between consumers and the world of FinTech. Although challengers and non-financial companies have introduced innovations and products of the highest level, they still do not enjoy the trust that is placed in old banking institutions. Incumbents should therefore take advantage of this positive factor before the innovation takes its course sweeping them away from the market, entering in partnerships with challengers and non-financial companies, ultimately creating an innovative industry focused on them.

3.2 The future of FinTech

According to a recent global market analysis conducted by Accenture (Accenture, 2019), consumers have never been so willing to share their data, provided they receive in return highly customized services / products to meet their needs. The analysis has revealed, in relation to financial services, 5 issues that banks and insurance companies cannot afford to ignore if they care about their survival. Let's analyze them one by one.

The first topic identified is the interest from consumers side in integrated propositions capable of satisfying core needs. This basically means that consumers no longer expect financial providers to provide purely financial services, but rather, they ask them for bundles of services that can cover more complex and articulated needs, so that they can secure a '*complete package'* at once. An example is a package that includes the purchase of the machine, financing, insurance and maintenance during the entire life of the machine. Another typical example is a package relating to the purchase of a house, which includes the home search service, mortgage management and property maintenance. Financial providers must try to position themselves as the fulcrum

within an ecosystem of alliances, suppliers, producers and organizations, each of which is responsible for producing a piece of the integrated proposal. Obviously, the financial providers must understand in depth what are their role and value within the ecosystem, in order to evaluate its advantages and disadvantages, especially in the event that it is unable to position itself as the orchestrator of the ecosystem: in this case brand management is essential in order to not be absorbed by the fame of an ecosystem's partner. *Figure 8* shows an example of the ecosystem built by a financial provider and its partners, where azure arrows represent network effect in which ecosystem interactions provide opportunity to up/cross sell services and acquire new customers:

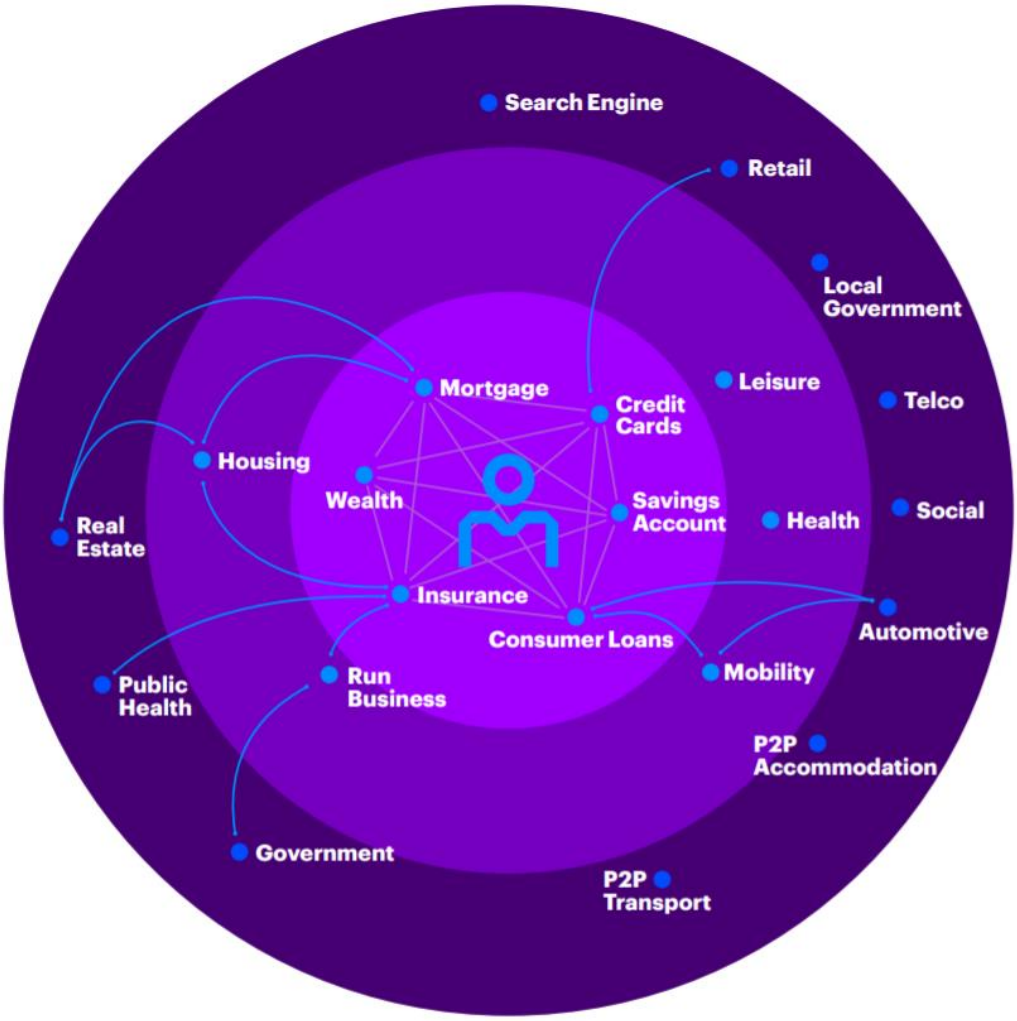


***Figure 8*: Trusted Financial Services Advisor at the center of growth-inducing banking universe from 2019 Accenture Global Financial Services Consumer Study (Accenture, 2019)**

The entire ecosystem rests on the incessant flow of data from consumers to providers, which allows them to receive continuous feedback on their customers'

needs and, in turn, create completely personalized services. Given that one of the reasons why consumers are reluctant to share their information with providers is the fear that they will be shared with third parties, it is essential that providers draw up and communicate clear guidelines for sharing data within the alliances.

And if on one side consumers want highly integrated products, on the other they ask for *customizations* pushed to the point of creating market segments made up of a single person: this is the second major issue that emerged from the survey. This desire includes personalized offers, notifications, notices, customized products, services based on usage or behavior patterns, etc. Again, in order to be able to offer a service of this type to the public, service providers must request a huge amount of data from their customers and this is where the real challenge lies, to process them by extrapolating useful information. A very interesting example of such extreme personalization that has already been on the market for some years is a device connected to cars which records everything that happens during each trip, ultimately generating a score: the better the score, the lower the cost of the insurance premium. Another products category that is having moderate success are wearable devices that can track the wearer's activity: if they follow healthy habits, customers accumulate points that can be exchanged in different stores obtaining discounts. The critical factor for success is the attractiveness of the service together with the non-intrusiveness of the same, but even more important is the safeguarding of customer data: given that a major reason why customers leave banks and insurance companies is precisely a failure in the protection of the data shared, providers must be able and ready to communicate to their customers the security measures adopted to maximize trust, and, of course, implement them effectively.

The third issue that will dominate the future is that of the very high disposition of consumers to *share their information*. Highly personalized financial services are based on data, and the more data a provider is able to collect, the more targeted and customized the offered service will be: consumers seem to have received and accepted this new trend, expressing their willingness to share as much data possible, as long as they receive services that fully meet their needs. Financial providers should therefore devote much of their energy to encouraging customers to share an increasing amount of their data. As already mentioned, some providers have already started offering highly personalized services, for example, applications that help consumers to manage money, or insurance products that are priced based on behavior. But there is a huge 'but': over time consumers will expect ever greater innovation in exchange for the data provided, and only those companies that manage to identify themselves as data-driven organizations by identifying and providing attractive services based on the data collected will be able to gain a competitive advantage. A help to this end comes

from artificial intelligence and in particular from predictive algorithms. By identifying correlations between hundreds and thousands of variables, such as income, age, sex and more complex patterns, like behavioral habits, these algorithms can provide vital assistance in predicting spending habits in the various channels and for the different products.

The fourth trend that suppliers have to commit to is consumer demand for what is called '*omnichannel*'. Before explaining how this trend is gaining a foothold, we need to spend a few words on the concept just introduced. To understand the meaning and impact of an omnichannel strategy, it is important to consider another concept, that is the one of *customer experience*. The costumer experience arises from a whole set of interactions that take place between a customer and a product, a brand, a company, or part of its organization: the response to these interactions is called customer experience. It is strictly personal and involves the customer at different levels (rational, emotional, sensory, physical and spiritual). Hence the definition of omnichannel: "omnichannel is nothing more than the synergic management of the various contact points (or touchpoints) and channels of interaction between the company and the consumer, to the aim of optimize this consumer experience"[28]. The contact points are the assets available to the company to build a relationship along the purchase process (advertising, pre-sale, payment and post-sale). They can be either physical (retail, call center) or online (social media, mobile app, e-commerce site): the integrated management of these contact points is the basis of an omnichannel strategy. This type of strategy starts from a multichannel approach to increase it and take it to a subsequent, more complete and deeply rooted dimension. In a multi-channel approach, the company limits itself to developing multiple contact points, without however taking on an integrated management of all the information, data and behavior of the users who pass through these touchpoints. Omnichannel takes a step forward because it not only puts the consumer at the center but provides for an interconnected system between all the contact points. There is, therefore, a transfer of data between the different channels and coherent content strategies, so that the user can not only interact with the company with a multiplicity of options, but also live the same experience on all those touchpoints and have no interruptions in the move from one to the other. In this way, the user can start an activity on one channel and continue it on another, without having to start it all over again. Many experts in the field confirm the importance of omnichannel strategy, for example this are the word from Daniel Hong, Senior Director of Product Marketing in [24]7[29]: "Omnichannel means supporting all channels and

[28] See https://blog.osservatori.net/it_it/omnichannel-significato-strategie

[29] [24]7.ai (full company name [24]7.ai, Inc.) is a customer experience software and services company that uses artificial intelligence and machine learning to understand consumer intent. It

having a holistic view of the customer regardless of communication method. This is vital as customers are increasingly dictating how they want to be engaged and serviced. Unlike multichannel, omnichannel interactions are not siloed but integrated providing for richer customer experiences that are connected (digital), continuous (consistent across devices, channels, and time) and contextual (relevant) no matter how many times a customer may transition from one channel to another for one task or during an entire journey".

Given what just exposed, companies should try to better integrate their channels. In particular, the goal must be to create for each market segment a customer experience consistent with the needs of those particular clients belonging to such segment, by offering customized products / services through a mix of channels that best fits with the considered consumers profile. Older customers who are attached to the past have an obvious preference for face-to-face contact, while younger customers, eager to experiment are highly interested in channel innovation. An example of the above is the one proposed by Pioneer Federal Credit Union, which has developed an application called "myPioneer Personal Assistant"[30] that allows customers in the United States to participate in video conferences with bank agents. The app simultaneously displays a video of the agent and a series of forms at the bottom of the page, allowing the customer to sign during the video conference, simply using his finger. Another example of a experience halfway between physical and digital is the virtual reality platform provided by the Indian insurance company PNB MetLife India Insurance Company Limited[31], which takes you directly to the office of an insurance expert with whom you can speak in "first person". Once again, artificial intelligence offers important help: predictive algorithms allow to predict customer dissatisfactions, signaling the need to devote more time to those customers who are able to bring high value to the company and who are highly likely to change provider. These technologies also allow to increase the speed and efficiency of the service, assessed as one of the most important criteria relating to the service, automating the process and increasing the number of transactions. In turn, it allows agents to focus on customers, to generate new value and an innovative high-level and first-person customer experience.

Since, as just anticipated, being able to connect to your bank / insurance company quickly and being able to use fast and efficient services is a fundamental need in customers, financial providers should leverage those omnipresent channels, such as GAFA (the term refers to the four most important data-driven

---

helps companies create a personalized experience across all channels. See https://en.wikipedia.org/wiki/(24)7.ai

[30] See https://www.pioneerfcu.org/Personal/PioneerGo/myPioneer-Personal-Assistant

[31] See www.pnbmetlife.com

companies currently in the world, namely Google LLC, Apple, Inc., Facebook, Inc. and Amazon.com), as means of expanding the presence on the market of their products and services. When considering the GAFA channels, the providers must carry out a very careful analysis, weighing the pros and cons of the initiative: if on the one hand this type of alliance allows you to reach the customer with payment services through consolidated social networks worldwide, to increase the interaction with the customer through augmented reality or even to reach the customer with more sophisticated market messages, on the other hand, it can prove to be a destructive move, given the threat by these giant channels owners to increase their financial services bundle and cut financial providers out of the market. Further in-depth information on the methodologies with which providers can exploit GAFA channels to reach consumers can be found in the article *Maximizing Revenue Growth in Retail Banking* from Accenture (Accenture, 2018). This topic however falls outside this document purposes.

Another point to focus on is the fact that almost a half of the customers say that the digital interaction with financial services providers is less satisfactory than their digital experience with companies in other sectors. Moreover, they add that they would trust technology and communications companies such as Apple, Inc. or Google LLC. to look after their long-term financial well-being. This introduce the fifth and last issue to be analyzed, i.e. the role played by *trust*. As previously highlighted, trust in services, and mostly in financial ones, is a pillar on which providers build their relationship with clients. The survey from EY has reported that customers still pose great trust in traditional standard providers, however two considerations should be evidenced. First, trust does not mean loyalty. Nowadays is extremely simple to switch from a provider to another, also given the quite absence of switching costs, thus more and more people are changing their services provider embracing the one that offers best services. Second, but strictly relates to the first, such people that trust banks most tend to trust also other non-financial providers. Consequently, competition increases and standard financial providers are forced to find new ways of creating value for customers: as we mentioned earlier in this report, personalization, a move to providing integrated propositions with non-financial vendors, and a focus on robust data security are promising avenues for strengthening customer loyalty and increasing customer lifetime value.

What said so far, hides between the lines an important concept: the entry of the world into the *post-digital era*. Following an Accenture report (Accenture, 2019), almost all of the IT and business executives in the companies find that technological innovation within their organizations has accelerated, in some cases with a decisive step. The phenomenon of digitization within companies is called *Digital Transformation* and it concerns products, business models or the

organization as a whole. In 2019 alone, the expenditure on Digital Transformation was reported to be 1.25 trillion of dollars (which for lovers of numbers is a number with 12 zeros) and the expectations are nothing short of impressive, with a forecast that touches the 2 trillion in 2022 according to a new update of the International Data Corporation (IDC)[32] Worldwide Semiannual Digital Transformation Spending Guide[33]. The importance of the issues related to transformation is therefore on the agendas of all those who decide the future orientations of companies in the financial sector. What marks a turning point compared to the past is the evolution and change of intensity in the relationship between IT and strategy, which has become mutually defined. The strategy defines where to compete and how to win and guides what are the technological choices enabling the achievement of corporate objectives (Lafley & Martin, 2013). At the same time, however, technology stands as the accelerator of change, making possible new business models, operational and application models from which the strategy cannot be separated. Then, technology is no longer put at the service of strategy as it used to be in the past, but rather the intertwining is such as to make you lose sight of which of the two pushes the other. This creates a holistic vision of the reciprocity between strategy and technology, the combination of which forms the basis for pursuing growth, efficiency and better business (Deloitte Consulting, 2019).

Therefore, given that most organizations are prioritizing the digital transformation, they will finally converge at the same point, where digitization, once a differentiated factor and source of competitive advantage, will be nothing more than a 'must to have' for each business (Accenture, 2019). And that's exactly where the world is today, at the dawn of a real revolution that will impact the way consumers perceive the reality. The digital saturation of reality guarantees more and more organizations the possibility to know their consumers at a high level of granularity. The companies are therefore trying to understand how to shape the world around consumers and exploit the right time to offer the right product / service. Hence, if we consider the collective effort, we realize how a real change in the way humanity will see the world in the next generations is happening: soon, each person will have his or her reality and every moment will represent an opportunity for companies to shape this reality.

But of course, the other side of the coin must also be considered. The more connected we are, the more vulnerable we become. The world that awaits us will be entirely interconnected, and banks will have to manage an ecosystem that sees

[32] International Data Corporation (IDC) is a US based provider of market intelligence, advisory services, and events for the information technology, telecommunications, and consumer technology markets. See en.wikipedia.org/wiki/International_Data_Corporation

[33] See www.idc.com

no longer just traditional partners such as credit agencies or payment networks, but also a new series of customers and collaborators belonging to well-established industries far away from the financial one.

3.3 The reasons behind Data Governance in Financial Services

Got here it is quite evident the importance of Data Governance in Financial Services sector. Yet, in this section we want to give some more insight about the fundamental role that a solid framework of governance of data plays for banks, insurers and the other financial providers.

The reasons for Data Governance are mainly two, i.e. the normative pressure and the effective benefit that can be pursued in term of cost reduction, increase in revenue and mitigation of risks. In the last decades the world has been through economic difficulties, uncertainties and the great economic and financial crisis of the 2008, that have raised attention on the concept of risk. Risk is obviously an old issue, but such events lead regulatory body to pose much focus on the default risk faced by financial providers. Every business in fact is expose to default risk, this is the natural trend of the market. But when bankruptcy occurs to a bank or an insurer the consequences can be catastrophic: they manage the savings and the hopes of hundreds of thousands of people which will be destroyed in case of default. Thus, many directives, policies and regulations were born to reduce such default risk and to regulate the financial world for the benefit of the entire society. Given the wide use of digitalization in banks, regulatory bodies must consider management of data when drafting directives: data are in fact the basis of business processes and are used to take strategic decision about the way to follow and poor-quality data will surely lead to destructive consequences. This relation is widely accepted and shared among expert, so much that it has been indicated as one of the reasons of 2008 financial crisis (Francis & Prevosto, 2010). Moreover, financial providers, as any other organizations, must draft documentation that goes to the public or to the Market, first of all Financial Statements: having to represents the real financial condition of the company, the information contained must be trustable, which, in turn, requires data of extremely high quality as ground for extrapolation process. Given the huge amount of data stored and manipulated by financial services nowadays, a solid Data Governance framework will be the only solution to avoid errors that can cause reputational damages of unimaginable propositions. And if in the past the quality of data was merely related to the business side, in the last years regulatory bodies have understood the necessity to focus on data quality also in term of IT and

elaboration of data. Errors that may seem small can lead to much greater information problem. Think for example at spelling errors that will make the information inaccessible or duplicated data that will make the information difficult, if not even impossible, to be recovered. Other examples are represented by incomplete data that will complicate the aggregation process or, in the extreme case, a loss of data with all the damages consequences that it can have.

This represents a touchpoint between the two forces previously indicated. Considering both the literature and my (short) direct experience with knowledgeable people, I can say that both business and IT workers are starting to understand the importance that Data Governance has in term of performance improvements. Until some year ago, most managers start Data Governance programs just to be compliant with regulations, however the trend is changing quickly. The awareness that high quality of data supported and enhanced through a well-structured Data Governance framework can lead to an increase in revenues and a costs reduction is becoming the first reason to implement Data Governance in the organizations. This will in fact eliminate, or at least mitigate, the famous problem of Garbage In Garbage Out (GIGO), extremely improving the effectiveness of the decision making process. In computer science GIGO is the concept that flawed, or nonsense input data produces nonsense output or "garbage"[34]. Therefore, for what said so far, having trustable data as input allow to extract from them powerful insights that can lead organizations to gain competitive advantage. It is not by change that many companies are adopting Data Governance and Data Quality tools to improve their data management process experiencing enormous advantages. The major suppliers of Data Governance/Data Quality are represented *Figure 9*:

[34] See en.wikipedia.org/wiki/Garbage_in,_garbage_out

***Figure 9*: Data Governance Tools from Gartner, 2019**

Having highlighted the reasons why Data Governance is gaining attention in financial services, the next chapter will be dedicated to study in depth this complex theme, both in terms of regulatory context and more technical features.

Chapter 4
DEEP DIVE IN DATA GOVERNANCE

In Chapter 2 we have introduced Data Governance trying to give a definition of it through literature analysis. Moreover, they have been shown the main roles of a well-implemented Data Governance model and the main question to which Data Governance should answer. In Chapter 4 we will give more information about this relative new branch of study, describing the regulatory framework in which Data Governance is located and the best practices adopted in order to implement an effective and useful enterprise-wide model for the governance of data.

4.1 Community and National Regulatory Framework

The importance acquired by management of data in the last years is underlined by the extremely complex and day by day wider regulatory context in which Data Governance is placed.

A good starting point is *Basel III regulatory framework*, also called ("Third Basel Accord" or "Basel Standards"). Basel III is an international regulatory accord that introduced a set of reforms designed to improve the regulation, supervision and risk management within the banking sector. The Basel Committee on Banking Supervision[35], the body who published the first version of Basel III in late 2009, does not have supranational authority and therefore does not have direct legislative power, however its proposals are usually accepted as binding legislation. Basel III replaced the old agreement Basel II and its introduction where pushed by the awareness that the deeper connection among banks, insurers and other financial services providers will improve risk dramatically. In the case that even just one of such institutes will result not strong and safe enough, the entire system would be threatened. What is worth to focus on for Data Governance purposes are the "Principles for Effective Risk Data Aggregation and risk Reporting" or PERDAR (Basel Committe on Banking Supervision, 2013).

[35] The Basel Committee on Banking Supervision (BCBS) is the primary global standard setter for the prudential regulation of banks. See bis.org/bcbs/

They are also defined BCBS 239 which stands for “Basel Committee on Banking Supervision n.239”.

14 principles are listed in PERDAR and they cover four closely related topics: overarching governance and infrastructure, risk data aggregation capabilities, risk reporting practices and lastly supervisory review, tools and cooperation.

The first category is related to the general management risk framework and highlight how the company’s board has the leading role of manage data quality risk and search for mitigating actions. Moreover, attention is placed on the data architecture and IT infrastructure. As stated by the second principle, a bank must have an IT infrastructure able to support its risk data aggregation capabilities and risk reporting practices both in normal and in stressed conditions. Roles and responsibilities should be established as they relate to the ownership and quality of risk data and information for both the business and IT functions.

Second macro category refers to the capability in the aggregation of risk data and in particular lists the Data Quality dimensions you need to be compliance with. The first principles belonging to this category talks about *accuracy* and *integrity* highlighting how banks should generate data which are reliable and accurate and that are aggregated on a largely automated basis so as to minimize the probability of errors. Controls surrounding such risk data should be as robust as those applicable to accounting data. Also, banks should document and be able to explain all of their risk data aggregation processes when a supervisory authority asks them. Principle 4 introduce the *completeness* which requires that bank should be able to capture and aggregate all material risk data across the banking group, there must not be lack of significant data. Another dimension of Data Quality is *timeliness* that says that 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*. This last one is related to the fact that a bank should be able to generate aggregate risk data to meet a broad range of on-demand, ad hoc risk management reporting requests, which means that it should be flexible and able to produce quick summary reports.

Data alone does not guarantee that the board will receive appropriate information to take effective decisions about risk. Third macro category, which refers to risk reporting practices, contained other principles to follow in order to manage risk effectively. They are *accuracy*, which expand the definition given before adding that risk management reports should accurately and precisely convey aggregated risk data and reflect risk in an exact manner; *comprehensiveness*, for which reports should comprehend all material risk area within the organization and granularity of the information contained should depend on the size and complexity of the bank and on the recipients expectations;

clarity and usefulness, stating that information contained should be clear and concise but detailed enough to help skilled people in the decision making process; *frequency* of the report, that is decided by the board and depend on the type of the risk reported, the speed at which the risk change and the contribution of such report in the decision making activities; *distribution*, for which risk management reports should be distributed to the relevant parties while ensuring confidentiality is maintained.

The fourth and last macro category deals about supervisors, thus is out of scope in our Data Governance - related analysis.

In Italy, such normative is implemented through *Circular No. 285* of 17th December 2013. On December 19, 2013 the Bank of Italy[36] published the new supervisory provisions for banks and investment firms which entered into force on January the 1st, 2014. Circular No. 285 implements the regulatory package, known as the "CRD IV Package", containing rules aimed at strengthening the capital requirements and prudential supervision of credit institutions and investment firms in the European Union. The CRD IV Package, approved on 20 June 2013 by the Council of the European Union, is made up of Directive 2013/36, the CRD-Capital Requirements Directive, and from EU regulation No. 575/2013, the CRR-Capital Requirements Regulation. The Directive and the Regulation incorporate the standards defined by the Basel III framework and replace the previous Directives on the matter, 2006/48 / EC and 2006/49 / EC (which reproduced the detailed provisions of Basel II). The Circular introduces important concepts related to the centrality and availability, for the Banks, of high-quality information, accessible at appropriate times and in controlled ways, and therefore requires them to define an adequate data management system (Data Governance). In this regard, it is considered appropriate to highlight Chapter 4 of Title IV, the 'Information System', and in particular Section V, which refers to the Data Management System. The section opens by reporting the definition and the tasks assigned to the data recording and reporting system, as follows: "The data recording and reporting system is responsible for promptly tracking all company operations and management events in order to provide complete and up-to-date information on company activities and on the evolution of risks. It continuously ensures the integrity, completeness and correctness of the stored data and the information represented; in addition, it guarantees accountability and easy verifiability (for example, by the control functions) of the recorded operations "(Banca D'Italia, 2013). The definition highlights several key points on which it is necessary to dwell in order to implement a system that ensures the total quality of

[36] The Bank of Italy is the central bank of the Republic of Italy. It is a public-law institution regulated by national and European legislation. See bancaditalia.it/chi-siamo/index.html

the data. First of all, it can be seen how the definition requires *timeliness* of the data, which is connected to the time interval between a change in the real world and the resulting change in the information status. *Completeness and correctness* are two other dimensions of the data on which attention is paid. The first is defined as "the recording of all events, operations and information with the relevant attributes necessary for processing". As for correctness it is presented as "the absence of distortion in the registration, collection and subsequent processing of data concerns the degree of reliability of the data and the absence of errors in its production". The concept of *accountability* has been already presented previously during the presentation of the RACI matrix: guaranteeing accountability means installing a system of roles and responsibilities relating to the data produced. Continuing with the examination of the section, in the list of the requirements that a data management system must satisfy, at point 2 we find the definition of a corporate Data Governance standard, not better specified, which therefore leaves full autonomy to the organizations to implement the system they deem most suitable for them. However, the Circular explicitly highlight the necessary presence of at least one responsible for the management of the controls required and the final validation of the quality of the data, namely the Data Owner. In the same requirement, Key Quality Indicators (KQI) are also appointed: they are indicators designed to guarantee and measure data quality. It is in fact necessary to establish some variables able to summarize and disclose the quality of data in order to allow the reporting and sharing of results achieved. Continuing with the analysis, the Circular face the problem of information integration in the case of a banking group: there must a clear, accurate and complete share of information among all the legal entities of the group to be compliant with Circular requirements. In this case it is fundamental to draw up a policy which clearly states sharing procedures and equip themselves with information sharing tools. Finally, it is worth to point again the importance of the reporting capabilities of the system. First, process, procedures and storage of data must be completely documented: it must be created what is called Data lineage of data, which allows to reconstruct the data flow in its entire lifecycle. Second, the reporting requirement. As already seen, it is required that the reporting system is able to produce timeliness high quality information for Supervisor authorities, evaluating the solidity of banks, and for the market: this is strictly related to the Regulatory Compliance, one of the main goal of Data Governance. To reach the goal, however, tools that allow to visualize information at different level of granularity are fundamental.

The counterparty of Basel III in insurance sector is *Solvency II Directive* (officially Directive 2009/138/CE). It is a community-level that replaces the

previous 14 directives and the 28 national regulations, going to constitute a single regulation for the entire area of the European Union. Solvency II was definitively approved by the European Parliament 10 years ago, in 2009, but only from 1st January 2016 it entered into force. A legislative process that has lasted so long is justified by the complexity of the technical problems that had to be faced to develop the new risk metrics for insurance companies. You can imagine the difficulties in finding harmonization between the different national realities. Furthermore, the great global economic crisis of 2008 risked ending the process with nothing done. With Solvency II a sort of new paradigm is introduced that can be defined 'risk based' as it places the quality and quantity of risk that each company takes with its decisions of commitment to policyholders at the center of the attention of the supervisor, the market and the company. Since policies management is done through company information systems, risk management is strictly related to the quality of the data managed by the insurance company. Solvency II is the first legislation to establish strict data quality requirements for insurers. Since many business processes provide data that will then be used within the model adopted by the company, data quality must be guaranteed in every process and at every level. The quality requirements must be applied not only to all data used in the internal model, but also to those used in the management and development of business choices. In particular, Article 48 highlights how "insurance and reinsurance companies must provide an effective actuarial function that assesses the sufficiency and quality of the data used in the calculation of technical provisions". The "Quality of the data" is explained in Article 82, which highlights three fundamental dimensions of the Data Quality, such as *appropriateness*, *completeness* and *accuracy*. From Article 120 to 125, the need for insurance and reinsurance companies to provide for a review and update cycle both of the data series used for the calculation of the probability distribution of technical provisions and of the internal model is reiterated.

As you can see the contents of the Directive is remarkably similar to the ones of Basel: Solvency II is in fact an extension of Basel II (now replaced by Basel III) to the insurance industry.

In Italy, Solvency II is implemented through IVASS[37], the Insurance Supervision National Institute. In particular, IVASS introduced in the piece of legislation No.17 of 15 April 2014, which integrated and amend Regulation No. 20 of 26 March 2008, specific criteria on Data Governance taking up many of the general data quality requirements defined by the Solvency II Directive framework. According to Article 20, "Companies must have accounting and management information that ensures adequate decision-making processes and

[37] See https://www.ivass.it

allows you to define and evaluate whether the strategic objectives have been achieved in order to subject them to any review. The internal control system ensures that the information complies with the principles of accuracy, completeness, timeliness, consistency, transparency and relevance". This concept has been later expanded in Article 12 bis of Provision No. 17: “Companies provide for a registration and reporting system of the data that allows their traceability in order to have complete and updated information on the risk profile and solvency situation”. Thus, while all the “old” Data Quality principles must continue to be assured, the Provision requires also to be able to reconstruct the activity carried out and to identify the relative owners. In order to that, the company defines a corporate Data Governance standard that identifies roles and responsibilities of the functions involved in the use and treatment of corporate information. Regulation 20 has finally been repealed by the entry into force of IVASS Regulation No. 38 of 3 July 2018, whose announced primary scope is to strengthen the qualitative management requirements which, together with the quantitative prudential requirements, represent the pillars to safeguard the stability of insurer companies.

Finally, GDPR. The *General Data Protection Regulation* or *GDPR* (Regulation (EU) 2016/679) is a regulation by which the European Parliament, the Council of the European Union and the European Commission intend to strengthen and unify data protection for all individuals within the European Union (EU). It also addresses the export of personal data outside the EU. The GDPR aims primarily to give control back to citizens and residents over their personal data and to simplify the regulatory environment for international business by unifying the regulation within the EU.
The GDPR has come into effect in all EU Member States on 25th May 2018, without the need for any national implementing laws. From that moment on ‘data controllers’ and ‘data processors’ will be exposed to the full effect of the law, which includes the risk of regulatory action (enforcement notices and fines) and the risk of compensation claims brought by individuals. In fact, the GDPR imposes different requirements depending on whether a party makes material decisions regarding the relevant processing or follows another party’s instructions. In particular ‘data controller’ determines the purposes for which personal data is processed and the way personal data is processed and ‘data processor’ processes personal data on behalf of a controller.

The GDPR brings certain new rights for individuals, such as the ‘right to be forgotten’ (people will have greater power to demand deletion of their personal data) and ‘data portability’ (people will be able to take their data away with them). For this the GDPR ‘privacy by design’ applies, which means that we need to

design new systems and tools to facilitate such rights. In the meantime, it is important that you act according to applicable regulations for processing and storage of personal data. Next to that, it is important to be transparent about the reasons to collect, process, store and delete personal data. This requires better registration and documentation of our data processing activities.

The GDPR applies to the 'processing' of 'personal data of EU citizens' by controllers and processors based in the EU. It also applies to organizations based outside of the EU, if they are offering goods or services to people within the EU or monitoring their behavior for which personal data of European citizens is processed and/or transferred. Note that processing means any operation that is performed on personal data, in the broadest sense, including mere collection and storage. The GDPR applies to both manual and automated processing activities.

A clarification is need about what Personal Data are: information relating directly or indirectly to an identified or identifiable person, which include obvious identifiers (such as name, address and birthday); value judgments about people (as in HR records); online identifiers (such as IP addresses and browsing histories). Some categories of personal data are marked sensitive data (special-), such as sexual orientation, ethnicity, religious beliefs, political beliefs and medical data. Such sensitive data is subject to additional restrictions on collecting and processing. Whether personal information is publicly available is irrelevant. Publicly available information is also in scope (so gathering personal data from social media websites is regulated under GDPR).

4.2 Organizational Models and drivers of choice

During the previous chapters it has been repeated several times how much Data Governance is focused on the definition of roles and responsibilities, and indeed, in chapter 2, we saw how this is the first and most shared phase for the implementation of a Data Governance model. Defining roles and responsibilities in fact allows you to always have references to turn to in case of problems and allows to mitigate the risk related to Data Quality problems. Upstream of this assignment, however, there is a preliminary phase, in which the company must decide which type of organizational model to use to implement Data Governance. Considering a series of distinctive elements such as a clear understanding of the organization's values, objectives, culture and vision, data management can be organized in different ways, declining roles and responsibilities on existing structures or creating new ones. The models that can be adopted are typically distinguished by the level of centralization / decentralization and the distribution of responsibilities, either hierarchical or cooperative approach. The four typical

organizational models represented in *Figure 10* are determined according to the composition of these two parameters.

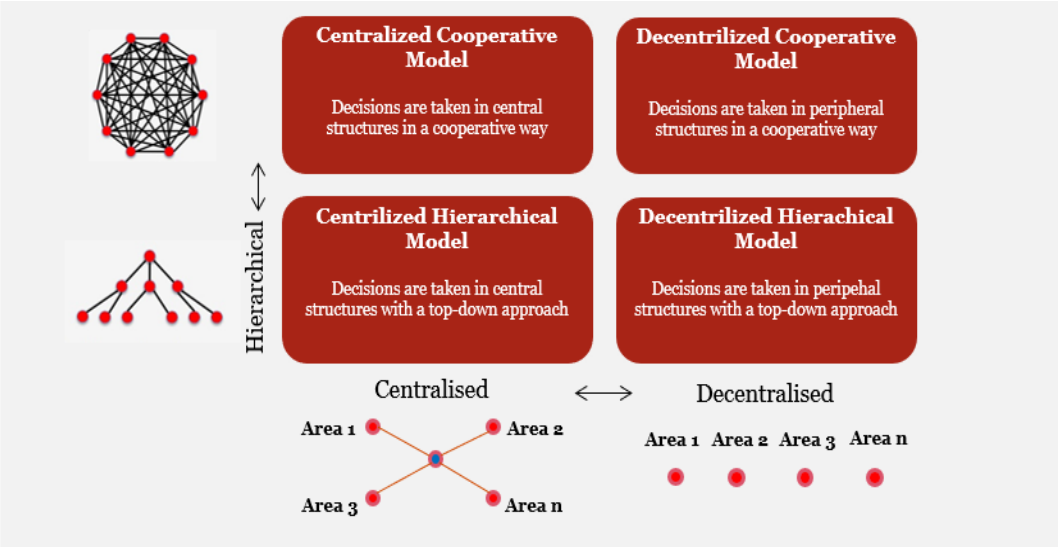


***Figure 10*: organizational models for Data Governance**

Starting from the *centralized hierarchical model* (lower left quadrant), it provides a strong central power, with a single control point belonging to the top management, such as the Data Governance Committee or the CDO. Thus, standardization and optimization are at the highest possible levels. Having control concentrated in a single point allows the possibility to use unique management, measuring and monitoring systems for the enterprise as a whole. Great attention is paid to company's policies and directives, which are spread from above onto the entire structure, in a forced manner, leaving limited freedom and delegation to the lower layers of the pyramid. However, it should be necessary to let go old structure or models previously adopted, in order to embrace the only chosen one. Moreover, the implementation effort both in term of organizational and IT side could be considerable.

Going up, we find the *centralized cooperative model*, which like the previous one focus decision power at central level. Unlike the hierarchical model, however, in the cooperative one, less apical figures are involved in the decision-making process too, trying to include therefore a greater heterogeneity of ideas and individuals, represented not only by the top management. In some cases, it allows to integrate coherent Data Governance/Quality structures already adopted, but generally old structures should be, at least partially, redefined.

The benchmarking analysis at Italian and European level, which will be shown in detail in Chapter 5, has revealed that the centralized model is the least suitable for managing a Data Governance policy due to the cost of development

and maintenance. However, the advantages are not absent, especially in terms of the uniqueness of the data management project, guaranteed by the fact that it has a single control point. In addition, commitment increases with a centralized hierarchical structure, since taxation from above prioritizes the initiative, being perceived as “imposed from the top”.

Let's now move on to the *decentralized* structures, starting from the *hierarchical* one (lower right quadrant). This is the most used structure, and since it has been adopted also in the Data Governance project I participated to, it will be widely presented in the part of the document dedicated to project presentation. Hence the treatment is postponed in Chapter 6.

The latest model is the *decentralized cooperative*. It abandons the concept that policies and directives must come from top centralized top management and favors a profound decentralization. A greater number of roles and responsible are identified, and they are left with a wide autonomy and decision-making freedom. It is suitable to be adopted in the case several Data Governance structures have been already adopted without a strong central guide. However, there is a big issue, since this model can lead to problems of non-uniqueness of data processing and difficulties of centralized reporting, such as in the case where it is necessary to calculate aggregate KQIs in reply to supervisory authority requests. Since each functional unit tends to treat data differently, complex integration mechanisms must be foreseen when there is a need to treat them homogeneously.

Finally, it should be noted that in decentralized model Data Governance activities are not the ordinary operational activities entrusted to the corporate actors, but rather, they are usually on-top activities, which therefore increase the effort required to them. In the case this effort is not recognize by top management, commitment can be adversely affect given the loss of motivation of the staff. In turn, this can create tension and resistance from low levels members, slowing down the adoption and resulting in an increased effort by top management, or in the worst scenario, in a failure.

4.3 Best Practices in Data Governance

It has been widely emphasized that Data Governance is still a relatively new topic in the literature. This means that to date there are no standard models of application useful in order to clearly implement a solid and scalable Data Governance framework. Despite the short introduction, however, it is possible to identify some guidelines that allow to shed more light on what is set out in the previous chapters. Several frameworks have been proposed, such as the DAMA Framework, the CMMI DMM Model, the MITRE DMDF model. In this chapter,

greater attention will be paid to the DAMA Framework, as the basis on which the Data Governance model implemented in the project under consideration has been structured. Finally, some words will be spent on the CMMI DMM Model, given its important in the determination of Maturity Level.

4.3.1 DAMA framework

In 2009, DAMA (Data Management Association)[38] association published the first version of the Data Management Body of Knowledge (DMBOK) (Mosley, et al., 2009) revised in 2017 with the publication of DMBOK2 (Henderson, et al., 2017). The purpose of the association is that of providing widely approved guidelines regarding Data Management, that is the management of the data in all its facets. According to DAMA, the world of Data Management can be divided into a series of domains defined as "knowledge areas", as shown in *Figure 11*:

***Figure 11*: The DAMA-DMBOK2 Data Management Framework (DAMA Wheel)**

[38] DAMA International is a not-for-profit, vendor-independent, global association of technical and business professionals dedicated to advancing the concepts and practices of information and data management. See https://dama.org

Totally ten domains are identified plus one that is placed at the center, the one of Data Governance. The insertion of Data Governance in a prominent position, underlines how for the authors this knowledge area plays a superior role: it is the functional domain without which the others could not work. Therefore, if it is true that all the domains are linked together and that a given issue related to data probably falls into more than one knowledge area, for creating a holistic model the Data Governance functions is essential: it represents the glue of the whole theory. The contents of the eleven domains as presented on the DMBOK will be briefly analyzed below.

Let's start with *Data Architecture*. When we talk about data architecture, we refer to the way in which data is saved and shared throughout the company during their entire life cycle. Attention therefore, is not placed only on systems but rather on the flow of data between the different systems, at a level of detail such as to understand where the data was created, where it is saved, how it is shared, and how it is exchanged between one system and the other.

There are several techniques used to implement the data architecture and they all work in combination with each other, so much so that taken individually, the would not be enough to design the entire data architecture frame ,but taken together they are able to give life to a solid architecture. Among these we find, data asset inventory, data standard and data models. But what we want to focus on is the starting point of data architecture and, probably, of the entire Data Governance implementation, namely the Business Glossary. It may seem strange, but it often happens that within companies, different people attribute different names and / or meanings to the same concept, creating confusion and the impossibility of sharing trustable information. It is therefore fundamental that all people have a shared vision about what data flows within the company and what is the meaning attached to those. This issue is addressed through a Business Glossary which is made up of a list of important terms for a company's business operations accompanied by the related definitions that must be shared by all the people who work within the organization. It will assure a common background for all the members by promoting easier communications between individuals belonging to different functions and a more fluid sharing of information, useful for the company's day by day activities.

Data Modeling and Design is the process through which a data model is created for the data that must be entered in the information system. The data model is a conceptual representation of the Data Objects, the associations between them and the rules in force between them. The data model helps in the visual representation of data and allows better alignment with business rules and regulatory compliance. Note how the data model focuses on what data is needed

and how it should be organized and connected to each other, rather than on the operations that must be performed on such data.

There are different types of schemes that can be used to represent data, such as relational, dimensional, object-oriented. All these models can exist at three different levels of detail: conceptual, logical, physical. The conceptual model defines *what* the system contains, it is therefore a very high-level model containing business concepts and rules. The logical model defines *how* the system should be implemented *regardless* of which type of information system or DBMS will be used. Finally, the physical model describes *how* the system will be implemented on a *specific* DBMS, its purpose is therefore to implement the database.

Moving on to the next slice of the data management wheel, we meet *Data Storage & Operations*. It includes the design, implementation and support of saved data, in order to maximize their value during the entire life cycle and facilitate their disposal (Data Purging). The activities belonging to this category can be divided into two sub-categories. On the one hand, the activities supporting the database, which focus on the data life cycle, from the initial implementation of the database to the purging data. They also include database monitoring and tuning activities. On the other hand, the technological activities supporting the database, which concern the resolution of all technological problems, such as the definition of technical requirements, architecture, etc.

The fourth knowledge area we meet is *Data Security*. In the more general sense of the term, it refers to the most complete data protection and is also known as "information security" or "cybersecurity". Going into more detail, Data Security includes the planning, development, and execution of procedures and policies on security in order to provide correct authentication, authorization, access and control of data. It should be noted that the specifications for security vary from industry to industry, depend on the country in which we are located and change from public to private sector. However, it remains common the need to protect information in accordance with what is established by business requirements, contracts and, above all, privacy regulations, first of all the GDPR.

The fifth functional domain is that of *Data Integration & Interoperability*. It describes the process of movement and consolidation of data within and between data stores, applications and the organization as a whole. Data Integration has the purpose of consolidating data in forms that are consistent, both physical and virtual. Data Interoperability, on the other hand, is the ability that systems have to communicate. This functional domain is newly defined, in the sense that it was not present in the old version of the DMBOK. In version 2, its introduction was necessary mostly due to the huge amount of fundamental data management activities that it allows and that most companies need. Examples are the sharing of

data between different applications, the management of interfaces, the acquisition of external data, the data migration. The need to effectively manage data movement is increasing, given the fact that organizations use hundreds of different databases that have to exchange data with each other. This need is further enhanced by the fact that organizations today are unlikely to develop internal software when they need one, but rather look to the market. And since each application found on the market has its own peculiarities and characteristics, it will be necessary to interface them with those already existing in the internal information systems.

Sixth place is occupied by *Document & Content Management*. The task of this domain is to acquire, store and allow access to data and information not contained in relational databases. Its focus is therefore on documents and other types of unstructured or semi-structured data in order to manage them and ensure easy integration and interoperability with structured data. The indexing carried out by this function is also important: the classification of documents must be done through an index so that a connection is always maintained with the original document, whether it is digital or in paper form.

The *Reference and Master Data Management* has the primary purpose to reduce the redundancy of data within the company information systems. Every organization needs some data that are common to all company business units, the same for all jobs: obviously, if such data is accessible from a single secure source, which is usually called Golden Source, the whole organization would benefit, with a positive advantage also on customers. What happens too often, however, is that, due to the various processes, merges, and business activities, different systems perform the same activities in an isolated manner, without communicating with each other with a consequent deterioration in the quality of the data. This, as we know, leads to increased costs and a dramatic increase of risks. By introducing the Master Data, it can help in the reduction of redundancy and therefore in the reduction of inconsistencies in the data structures and in the data values across the different systems.

Data Warehousing & Business Intelligence essentially allows access to data for reporting and decision-making purposes. A data warehouse allows you to save operational data relating to day-by-day business, thus providing the history of the company. Starting from these data, it is possible to extrapolate trends, information and suggestions that allow analysts to make predictions about the future and support managers in the decision-making process with suggestions based on certain, concrete data. Furthermore, the data warehouse usually integrates data from different sources, which allows you to expand the ability to analyze and forecast future trends.

The ninth functional domain is represented by *Metadata Management*. A simplistic definition of metadata is the classic "data about data". However, there are many types of information that can be classified as metadata, starting from information relating to technical and business processes, passing through the rules and constraints imposed on the data, up to the description of the logic of data. They can also represent the data itself, the concepts that the data represent, or the connections between those concepts. It is therefore clear that a more complete and articulated definition is needed, such as that reported by Gartner (Beyer, et al., 2104): “Metadata is information that describes various facets of an information asset to improve its usability throughout its life cycle”. The purpose of metadata is to make data visible, trustworthy, usable, and encourage their sharing. One of the cardinal principles of data management is Data Sharing: such sharing should lead to a benefit for the recipient, who should be able to interpret the information received. And data experts agree that the data is shared only when it is accompanied by all the metadata necessary to understand and use the data effectively. Therefore, the importance that this knowledge area has in the world of data management is evident.

The last two domains are probably the most interesting and articulated. We start from *Data Quality*, which obviously refers to the quality level of the data received by the user. As we have seen above, high quality data are as such if they are suitable for the uses for which they were created. Taking this perspective to describe data quality has consequences for the investments that an organization has to start, since not all data quality needs inside the company are equal and of equal importance. A classic example is the difference in quality required for financial reporting and that required by analysts to make predictions about the future: if a firm should decide the area in which to invest in order to improve the quality of data produced it will for sure start from the financial one. Data quality is an objective pursued by all dimensions of data management which underline how this knowledge area is of prime importance for company data. Poor quality data have negative results on risks, costs, loss of revenues and reputation. It is a fact that no organization has perfect business processes or flawless data management practices, with the result that everyone experiences data quality problems: a careful policy focused on these issues can really make the difference between a successful company and one on the verge of collapse. Some of the most common dimensions of Data quality are, for example, accuracy, which means that the data does not have any material error, has a consistent definition for all instances and is correct when compared to the reference source (which must be unique and certified); completeness, linked to the fact that the information set necessary for a user to perform his business must be complete, there must not be a lack of datasets or metadata; consistency, related to the fact that the data must be

consistent between the various sources on which it is present, it must be aligned among all the sources both in terms of content and meaning. There are also some data quality elements more related to the IT part, such as integrity, which refers to the level of protection, for example via password or limited access to the data itself so that their modification is possible only with prior authorization, or the audit trail, which considers the level of traceability of changes made to data, in order to always be able to trace the amendments made and the owner of them.

It should also be emphasized that Data Quality is a long-term program, a vision that requires the commitment of all the people involved in the process of creating and using the data.

Finally, the last and most important functional domain: *Data Governance*. As already highlighted, it is placed at the center of the model, indicating its necessary presence, without which the whole model collapses: without Data Governance we can say that there is no Data management. Data governance focuses on Data Management rules and resources and is always required, regardless of the level with which an organization wants to develop its own data management model. An important function of it is to determine who is responsible for making decisions, assigning roles and responsibilities among company figures, also providing indications on the escalation process to be adopted in the event that Data Quality problems arise and the person directly concerned is not able to deal with it. This leads to another task of Data Governance, that is to help people complete their work by providing them with the resources necessary to answer their questions, making publicly available the names of the people allocated in Data Governance roles and giving a way to the people to raise problems that can slow down activities.

It is worth note that all companies make decisions about how to manage the data and which people are involved in data-related tasks, but not all of them do it formally and with a clearly outlined data governance model. The difference between them and those that formalize the model is that the latter are able to extract much more value from the data than the former.

Once again, even the best Data Governance model in the world will not lead to any result, and indeed only to internal frictions, until the moment when commitment is created within corporate actors and they accept the change: this is the only possible way to extract from the data information that can lead to a great competitive advantage.

4.3.2 CMMI DMM Model

Usually a framework implementation begins by focusing on a single or at maximum very few data management domains in which particular and specific business issues need to be solved. Though this can work well to solve such particular problems, it is necessary to think bigger and thus have an eye toward the future: when these issues are solved you need to place the basis for a holistic and enterprise wide Data Management model. Following this way of thinking, you need to understand that Data Management is an incremental phenomenon and must be built on procedure already existing in the company, obviously aligning it the organization's strategy. Thus, in practice, Data Management implementation starts by narrowing the initial scope of application, which can be done vertically or horizontally. Considering the vertical situation, it means that you first adopt a high-level view and solve only the superficial problem and then you deeper the analysis. The horizontal way provides to execute data management at a more local level and in a subsequent phase at the enterprise level, perhaps across multiple departments involved in the same line of business.

When narrowing the scope this way, it is important that the organization follow a disciplined approach to Data Management. One useful resource to help accomplish this is to reference a Data Management Maturity Model as a guideline, the most important of which is the CMMI DMM (Capability Maturity Model Integration Data Management Maturity Model)[39]. The model breaks down Data Management into five high-level categories and one supporting category called Supporting Process and subdivides each category into three to five process areas, as illustrated in *Figure 12* below.

[39] See https://cmmiinstitute.com

***Figure 12*: CMMI DMM model categories and process areas from (Fleckenstein & Fellows, 2018)**

However, the key feature of the CMMI DMM framework is the application of "maturity levels" to certain data management areas, thus allowing an organization to assess its level of maturity in one or more independent data management domains. Moreover, it presents practitioners with the types of procedures and artifacts to be managed at various stages of maturity. Note that the way an organization deals with data-related topic will differ based on its data management maturity level. *Figure 13* and *Table 1* below give respectively a graphic representation of the 5 levels of maturity identified by the model and an overview of the procedures characterizing each one of them:

Figure 13: **Maturity Levels in CMMI DMM model**

Maturity Level	*Description*
Performed	Processes are performed ad hoc, primarily at the project level and are typically not applied across business areas. Company act on a reactive basis, thus repairing the issues and not trying to prevent them. The vision is narrow and focused on the single process.
Managed	Processes are planned and executed in accordance with policy, employ skilled people with adequate resources to produce controlled outputs, involve relevant stakeholders, and are monitored, controlled, and evaluated for adherence to the defined process.
Defined	Set of standard processes is employed and consistently followed. Processes to meet specific needs are tailored from the set of standard processes according to the organization's guidelines.
Measured	Process metrics have been defined and are used for data management. These include management of variance, prediction, and analysis using statistical and other quantitative techniques. Process performance is managed across the life of the process.
Optimized	Process performance is optimized through applying Level 4 analysis for target identification of improvement opportunities. Best practices are shared with peers and industry.

Table 1: **CMMI DMM capability and maturity levels in CMMI Institute, "Data Management Maturity (DMM) Model," Ver. 1.0, 2014**

The CMMI DMM has been used as starting point during the real case project in order to identify the as-is model of maturity of Data Governance in the bank group being studied. The analysis of the Model will be deepened in Chapter 5.

This section closes the literature review and Chapters from 5 to 7 will be dedicated to the presentation of the project in which I have been involved. We will start in Chapter 5 with an overview illustration of the initial situation of the company and the data - related issues it was facing.

Chapter 5
A REAL CASE STUDY ON A BANK GROUP: PROJECT INTRODUCTION

5.1 Project Overview

What exposed in the first four chapters, gives all the necessary background to understand what will be shown in the following, i.e. a real case implementation of a Data Governance/Data Quality model in a Bank Group. In particular, the contents of the following chapters come from my firsthand experience on a project carried out for an Italian Bank Group. Let's start by presenting an overview of the project and the reasons that induced the Group to start the Data Governance activities.

The project under analysis is carried out for a leading Italian Banking Group. The Group, headquartered in Italy, operates throughout Europe by means of an enormous network of subsidiaries. Its main focus is the performance of banking activities, both at retail and commercial dimension, however, through a series of specialized companies, it also deals with insurance, savings and asset management services. The Group is listed on the stock exchange, with all the advantages and disadvantages that derive from it. One of the most painful themes is certainly the reputational aspect, which is closely linked to the quality of the data produced. You can think of the various reports that a Banking Group produces towards the Market or the Regulator, first of all the Balance Sheet: if it contains incorrect values, the economic and reputational consequences can be incalculable. The Group, of course, is absolutely aware of these issues and over the years has implemented several improvements within its information systems and in their processes, in order to guarantee (or at least try) an absolute level of data quality. However, data quality issues raised on a daily basis highlighting the difficulty to manage data-related topic and prepare final reports headed both outside and inside the company. With reference to the last category, high quality aggregated information allows the implementation of a solid decision-making process based on reliable and trustable sources: strategic level choices, the ones

that determine the survival of the organization on the market, must necessarily be based on solid and reliable data. Moreover, the reporting production is slow, due to the fact that they spend much time and effort in controlling and clean up messy data with the risk of failing to meet regulatory requirements and deadlines and delivering wrong data. In addition, an audit inspection had revealed how, despite the level of data quality-related activities is evaluated as 'adequate', some issues are present, requiring the Group to start a process of adaption to the sector regulations. Finally, they are beginning to understand the big amount of opportunities they are lacking not being able to extract from data the right insights due to the poor quality of them. Improving all these issues ultimately means convey value to customers enhancing loyalty and strengthening stakeholders trust and relationship.

The whole set of reasons has led the company to favor the implementation of an integrated Data Governance government, capable to meet all their needs, starting from regulatory compliance passing from strategic decision-making process elevation and ending at supporting for value creation.

The first necessary step for the development of the framework is substantiated in the drafting of an internal *Data Governance Group Policy*, also in order to incorporate the obligations and requirements required by Circular 285 and BCBS 239. Necessarily, the Policy is drafted in a such a way to adapt to the numerous, and somewhat dispersive, policies already present within the Group relating to Data Management, to take into account what has already been established and leverage on it. The Policy, which will be dealt with in detail in chapter 6 declining it in the directives of which it is composed, contains, among the various annexes, the one relating to Legal Entities in scope: it is the list of all the subsidiaries to which the Policy should be extended, following the principle of proportionality. The *principle of proportionality*, states that the costs related to the application of a normative must not be disproportionate to the benefits it intends to achieve. This means that every Legal Entity, depending on size, internal organization, nature, scope and complexity of the activities carried out, must evaluate whether or not to inherit the Policy, asking in the latter case to be excluded (waiver).

Within this broad framework, I have been staffed on the project stream of implementation of the Policy in the Group subsidiary that manages the Real Estate assets belonging to the Group. The portfolio, we are talking billions of dollars, is made up of both historic and recently constructed buildings spread all over Europe. As a Real Estate company, the legal entity deals with buying and selling properties, renovations, purchase of land for the construction and sale of buildings, rental management, as well as all the surrounding activities to enhance the properties owned. Moreover, the company fully owned an asset management

company (SGR) through which manage a range of real estate funds on the behalf of Group companies and third-party investors. In order to clearly frame the activities carried out under the name of SGR, it is necessary to take a step back and define what an SGR is, starting from the concept of OCIT. OCITs (Italian acronym for Organismo d'Investimento Collettivo del Risparmio) are bodies whose purpose is to invest people savings based on a predetermined strategy. The reasons behind their existence are very simple and can be found in savers' ignorance about financial staffs who are unable to invest on their own, in the possibility of diversifying the risk, in the possibility to entrusting the money to people with skills and, not for the latter, in the possibility of entering into investments that are impossible to manage with a small capital. OCITs are divided into *i)* mutual funds (in Italian "Fondi Comuni di Investimento") and *ii)* investment companies, two categories which, while having various characteristics in common, have two fundamental differences. The first difference lies in the fact that mutual funds are not a standalone legal entity but are only investment instruments, therefore they need a company that deals with their creation and management: that company is precisely the SGR. Conversely, investment companies are already separate legal entities. The second difference concerns management of capital: investment companies do not separate the corporate capital from the capital raised by investors, which means that the capital raised is the company's capital. One consequence of this is that investors become full-fledged shareholders, with all the rights deriving from that position. In asset management companies (SGR), on the other hand, the investor becomes a simple holder of the shares in the fund. Focusing now on the asset management companies only, it can be seen how the assets included in the funds can be of different types, such as shares, bonds, monetary or real estate securities. And here we are at the point, since the subsidiary in question, as mentioned, is an asset management company that also manages real estate funds.

Analyzing the local reality of the Real Estate company, first we need to highlight that, as expected, it is facing pretty much the same problems as the holding. Also, the great complexity of the structure, mainly due to the large number of activities carried out and the huge amount of assets managed, it is an additional difficulty to address. The company already has several projects in place aimed at improving information systems, one of which has a significant impact on quality and reliability of data. It concerns the creation of a Data Warehouse within which all the data relating to the assets managed are collected. The insertion of data within the platform is based on a double-level control, carried out first by the IT experts as regards to technical aspects and then by the Business Data Stewards for the Business aspects, in order to assure the high quality of data entered. The aim of the project is to create a 'Golden Source' and to provide the actors

involved with a platform able of carrying out specific analyzes on the loaded data, both through the report generation tool provided by the platform itself and through the extraction of the data in Excel and their subsequent customized elaborations. The implementation of the Policy has to take into account the presence of these parallel projects with interlinked purposes, avoiding overloading the corporate actors by duplicating the activities required to them. However, they will not be shown in this essay, both because they are not part of the project managed by my team and because our team's view on them was extremely limited, being them carried out by other consultancy companies.

With the aim to provide a complete vision over the project, we present in *Table 2* below an overview of the activities carried out along with the correspondent deliverables.

Cluster	*Document Section*	*Activities*	*Description*	*Deliverables*
Cluster 1	5.2	Benchmarking Analysis	Evaluation of best practices, both at European and Italian level, through a Benchmarking Analysis	Data Governance Benchmarking focus on organizational aspects
	5.3	Data Governance Model Definition	Definition of the organizational model to adopt, also based of Benchmarking results	Data Governance Model
Cluster 2	6.2	System of Governance definition	Identification and appointment of Data Governance roles and responsibilities	System of Governance
	6.3.1	Data Governance Awareness Program and Data Quality high-level assessment	i) Conduction of meetings to aware and involved in the project the identified people ii) Collection of outputs/reports produced in each structure and Data Quality assessment of them	i) Spread the contents of the Group Policy all over the subsidiary ii) Evaluation of the initial Maturity Level. Mapping of reporting production
	6.3.2	Relevant Output candidates identification and evaluation	Prioritization of the "Eligible" Outputs according to the criteria and the process defined by the Group Policy	New Relevant Output identification
	6.3.3	Business Data Elements collection, Data Quality Controls collection	Collections of Business Data Elements and Data Quality Controls through templates from Business Data Steward	First collection of Business Data Element and Data Quality Controls
	6.3.4	Business Glossary and Dictionary of Controls	First filling in of Business Glossary and Dictionary of Controls. Support for the analysis of Data Governance/Data Quality software tools	Business Glossary draft, Dictionary of Controls draft, Software Selection
	6.3.5	Gap Analysis	Identification and evaluation of existing gaps in terms of Data Governance/ Data Quality from outcome of outputs collection activities	High-level remediation proposals and activities Roadmap
Cluster 3	6.5	Data Directory filling in and KQIs calculation	Filling in Data Directory only for Already Relevant Output and evaluation of KQIs	Data Directory filling in and KQIs calculation

***Table 2*: project overview with references to the document sections**

Hence, the intended final outcome of the project is to develop a model for the governance of data that will solve all the gaps present in the initial situation, both in term of information system, enhancing the handling of data and managing all the necessary Data Quality controls, and in term of ownership over data and activities, providing the company with Policy, guidelines, practices and standard to be followed. All this should lead the company to a new Maturity Level allowing to reach a double advantage. On one side, the requirements of Data Governance and Data Quality are met, allowing standardization of data-related activities with the possibility to improve reporting production and convey more value to the customers thanks to the insights gathered from data. On the other side, which actually is the one that organizations value the most, developing a Data Governance framework correlated with the right level of Data Quality, means being compliant with regulatory requirements, keeping themselves safe in case of a supervision control. Ultimately, in the long term, the company could create a reliable environment capable to improve customer loyalty.

5.2 Financial Providers Benchmarking Analysis

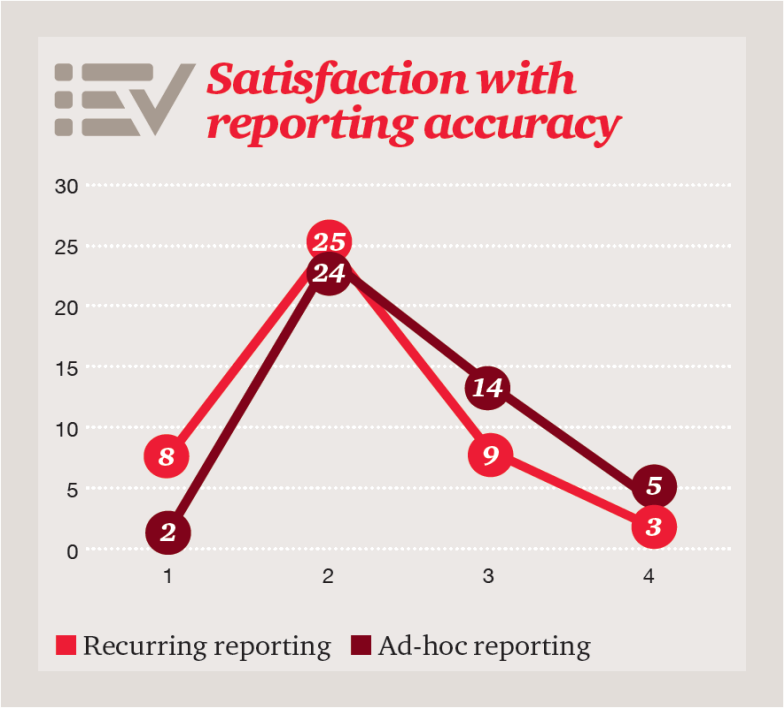
One fundamental activity carried out before to build the Data Governance model, also after explicit request by the Group, is an evaluation of the results coming from a benchmarking analysis. The aim is to extrapolate the best practices as of today and carry them inside the project. The study analyses the national and international framework with regard to the solution adopted by different Financial Services providers (banks and insurers) in the moment they need to build up their Data Governance Framework. In particular, the company I work for has a long experience in the sector and has already been through 14 projects similar to the one presented here, all for domestic banks. All of them were implemented to meet the particular Data Governance requests of the clients, keeping in mind that each of them has its own peculiarities and must be treated in accordance with its features. This means that does not exist a single model suitable for all the organizations, but, instead, the capabilities of the advisory company is in understanding which are the specific characteristic of the targeted bank, starting from the willingness of the organizational members to change their way of work, adopting new procedure, new tools and embracing a more data-driven approach. In addition to the national scenario, we can also count on a benchmarking analysis conducted at a European level through a survey. Such a survey, focused on organizational aspects, was structured in open-ended questions and provided useful insights for tackling the project we deal with.

Let's start form the results coming from the international survey by PwC (PwC, 2016). First, we need to introduce the sample considered. Using a standardized questionnaire derived from the firm Data Governance Framework, key specialist and expertise in 45 banks across Europe were interviewed. Among them there were CROs, CIOs, CEOs and in some case also CDOs. The banks which took part are spread across 12 European territories, mostly in Western/Central Europe and considering the balance sheets amount they cover a wide range, going from few billions to thousands. Looking at the outcomes of the questionnaire, what immediately catches the eye is that, although authority bodies are tightening their requests in term of granularity, volume and timeliness and the new regulatory framework asks bank to implement a clear and well-structured Data Governance Framework, almost 80% of the European banks are not provided with a CDO function. In the few banks in which it is found, it both is not concerned with data management and governance with 360 degrees view and does not have a clear position in the organizational charts. For what concerns the first issue, the interviews with CDOs functions show how they focalize their effort on small pieces of the cake, like for example, assess data relevance or managing data availability, losing the global view and looking at data management with a myopic way. The second issue is the clear reflex of the confusion that affect financial provides regarding the world of Data Management. The opinions about where to locate the CDO functions are wide and different, ranging from a direct report of CRO, CIO or CFO, to a complete independent function reporting directly to the CEO. However, the structure more widely adopted is the one in which he is linked to the COO. The reason behind a such diversified panorama is that, as already stressed, Data Governance is a relatively new topic and as any other new big argument it takes several years to enter in the organizational standards, due to resistance imposed by organizational inertia along with the fear to abandon the comfort of the safe status quo and jumping in the frightening unknown.

More interesting are the results related to the implementation of a Data Governance framework: two thirds of the banks do not have a companywide model for Data Governance. Moreover, and this is probably the worst trend discovered in all the survey, the first reason for addressing such topics is regulatory compliance. It has been in fact underlined how a well-defined Data Governance and Data Quality Management structure not only allows banks to be compliant to regulations, but can also increase business opportunities: better data available means better insights that can be extract from them and, consequently, better decisions, which, at the end, lead to above average performance. Nevertheless, the majority of the organization still decide to enter in data related project just to be adequate internal standards to the normative.

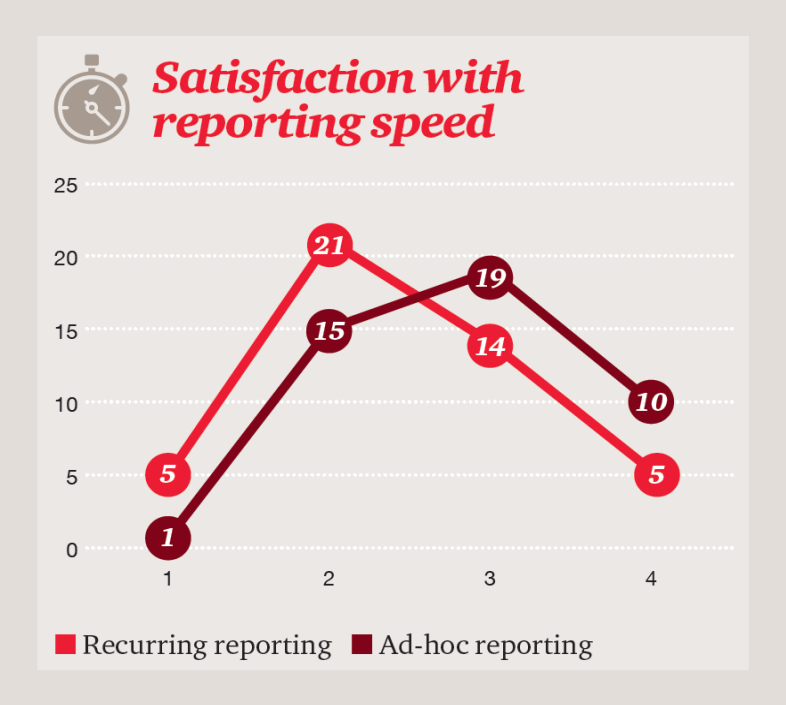
Also, the ownership of data, in the 70% of the banks have not been completely assigned. Data ownership means clearly define the responsibilities for the definition and the quality management of data items. This must be documented in order to avoid the occurrence of issues like duplication or redundancy of data or misleading between functions in cross functional data/report production situations. In addition, if ownership is clearly defined, in case of problems you can immediately go back to the responsible.

Another delicate topic is the reporting task, which required a decisive improvement in terms of quality, time, and effort required. The theme of reporting is in fact a big challenge for banks. According to the study made, the satisfaction of financial providers with regard to speed and accuracy are definitively improvable. The problem is bigger when ad-hoc reports are considered. The scores are represented in *Figure 14* for the accuracy and in *Figure 15* for the speed:



1 = Completely satisfied 4 = Not satisfied

***Figure 14*: Satisfaction with reporting accuracy, both for recurring and ad-hoc reporting, from (PwC, 2016).**



1 = Completely satisfied 4 = Not satisfied

***Figure 15*: Satisfaction with reporting speed, both for recurring and ad-hoc reporting, from (PwC, 2016).**

The major problems come from the ad-hoc reporting, since it implies less standardized procedures. Nevertheless, banks should build procedures also for this type of reporting, since regulators can ask for them in an unpredictable way. Although the statistics shown, four over five banks feel themselves confident for answering to ad-hoc reporting demand: what they did not consider is that the lack of clear responsibilities and standardized procedures means they continuously face the risk of failing to meet regulatory requirements and deadlines, and the risk of delivering wrong data. In addition, 70% of the banks in the sample, produce much more reports than the needed, wasting enormous amount of resources and shrinking the revenues: thus, procedures must be put in place also to ensure that only reports that are relevant for the company's objectives are produced.

Regarding the Data Quality, all the financial providers have understood the damages that poor quality can lead to, and the great majority of them have implemented data quality controls, procedures and KPIs (or, better KQIs, which stands for 'Key Quality Indicators') for measuring purposes, but this is only the tip of the iceberg. Going deeper, as a matter of fact, two major problems come to the surface. Primarily, only few of the respondents have implemented a process of continuously improvements, making the controls useless, since there is no further improvement of Data Quality and business practices based on the outcomes the controls. Secondly, less than 50% use software to measure Data Quality. This

software usually are in-house ones, since it is difficult to find on-the-shelf product able to adapt to the several IT systems used in the same banks. And the trend even decreases if we consider Data Quality reporting tool: less than one third have it. Business Glossary and Data Dictionary, the pillars of Data Quality, are absent in more than half of the banks, highlighting the risk of not being able to satisfy regulatory body requests. All this means that, despite banks recognize the basic importance of Data Quality, yet they are reluctant to build the framework needed to manage it in the most correct way with the consequence of a considerable waste of time and resources.

Moving now to the analysis of the Italian scenario, 14 banks have been included in the sample. Such organizations were previously involved in Data Governance-related projects and thus, thanks to the good relationship created, it has been possible to interview the members of the organizations through open-ended questions. Due to the sensitive contents of the data and non-disclosure agreements made with the respondents, the question and the answer cannot be reported. The only thing that will be shown are the aggregated results. First of all, the survey shows how the 70 % of the banks have adopted a decentralized structure, mostly with a light weighted structure, which means without a strong central coordination.

The level of maturity has resulted quite high for all the companies, with a concentration in the intermediate level, the Defined. Table 3 show the level achieved by the whole set of banks.

Maturity Level	*Bank*	*Count*
Performed	*None*	*0*
Managed	*Bank 2, Bank 4, Bank 9*	*3*
Defined	*Bank 3, Bank 6, Bank 7, Bank 10, Bank 11, Bank 12, Bank 14*	*7*
Measured	*Bank 1, Bank 5, Bank 8, Bank 13*	*4*
Optimized	*None*	*0*

***Table 3*: Maturity level of the surveyed Italian Banks**

Another important point coming out from the analysis is that in the initial phase of the project, an external help was requested by all the respondents. Two are the grounds: on the one side, banks do not have the necessary technical competences in Data Governance field, also given the novelty of the topic; on the other side the initial activities are time consuming and require the involvement of many resources. Furthermore, an external support could be the real enabler of success: when a strong central coordinator is absent, the commitment in the members touched by the project could be weak. You need in fact to consider that

most of the times, and particularly in the case of distributed models, the Data Governance and Data Quality activities are on-top, adding to the strictly operative ones and stooling time and effort to the workers. The presence of an external advisory company might ease the involvement of such actors, through a continuous interaction and supporting them in the project activities.

Last topic that should be point out in the national scenario is the one related to the adoption of software tools which can support Data Governance and Data Quality activities. It turns out that, with a greater average with respect to Europe, almost 60% of Italian banks have adopted both a Business Glossary and Data Quality controls tools.

The two surveys underline several problems that are affecting financial services industry when talking about Data Governance-related topics. The tasks of assigning role and responsibilities is fundamental for Data Governance activities, yet many banks still do not have clear ownership over data and reports and have not establish a dedicated CDO function. The Data Governance framework has been purely established in a low number of banks and the main reason to build it is not searched in the pursuing of better performance but in regulatory compliance. The reporting activities are far than optimized and banks are continually wasting time and resources, obtaining output of questionable accuracy and being not able to reply quickly to regulatory body requests. And finally, the Data Quality, which is at an embryonal level: banks are aware of the importance to have high quality data and many controls are place during the lifecycle of them, but the problems are mainly of two level. First, a continuous improvement process should be implemented in order to obtained better business performance based on the results coming from the controls made. Second, software tools must be used in terms of Business Glossary, Data Dictionary and Data Quality controls tool.

As we will see, the project we engage touches most of these points, trying to deliver to the Bank Group the solid ground of a complete Data Governance model, over which it should construct, day after day, a strong and organized framework.

Finally, it is worth nothing that this survey is 3 years old and since then financial providers have done some further steps towards the adoption of Data Governance model. However, the trends show that still a lot of work should be done. The Group for which I worked for is a manifest of what just said: despite being a leading European provider, it shown tremendous gap in term of data management.

5.3 Organizational Model definition

Following the outcomes of the survey, some decision which poses the grounds for the whole framework have been taken, mostly concerning organizational issues. First, it was decided to adopt a decentralized model with a central light-weighted coordination body, being the most common model by far in domestic and international environment. It is worth to remember that it promotes high flexibility and scalability given the absence of a strong central coordinator and allows the integration of different tools already present in the organization. Also, a first definition of the main roles was carried out and two macro categories of actors were identified, as show in *Figure 16* below. Business actors (Data Governance in the figure), related to the Bank side and IT actors (Data Quality), related to IT infrastructure and information system. At the top, you see the Chief Data Officer and, above him, the Data Governance Committee.

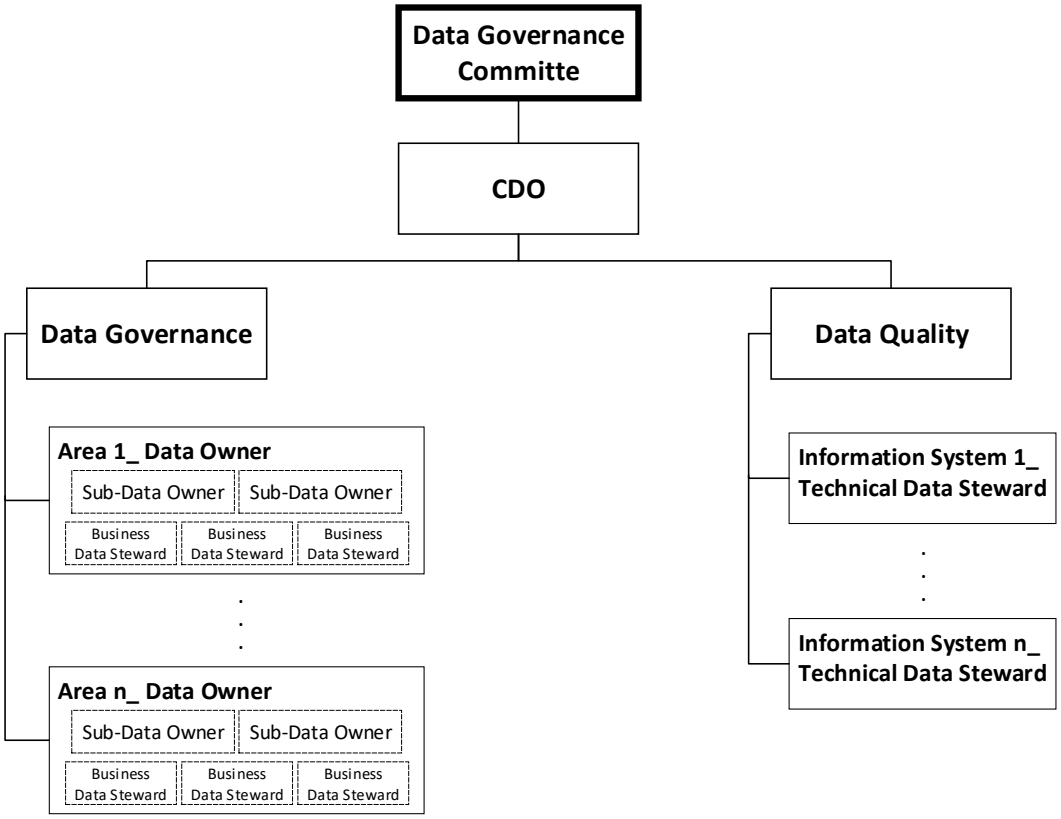


***Figure 16*: organizational model adopted in the real case study**

This same organizational model is replied in each of the subsidiaries, apart from the Data Governance Committee and the CDO function which are present only at Group level. At Group level the CDO is identified as a stand-alone function which report directly to the CEO. For the Real Estate company and the

other subsidiaries, instead, a figure called Data Governance Leader is identified within the CFO function, due to the fact that such a project is a CFO initiative. The Data Governance Leader is constantly supported by the COO, given they have a whole vision about the activities of the company. The Data Governance Leader, moreover, report to the Group CDO for what concern the status implementation of the Policy at local level.

With regarding to roles, the *Data Governance Committee* has the role to analyses the escalation requests on Data Governance/Quality issues and the definition of remediation action and the respective resolution plan. When in fact a Data Governance/Quality issue can cause major problem, an escalation process is activated from the point it arises until reach the Committee which, except in extremely special case, is empowered to solve them. The *Data Governance Leader* has the role of coordinating and monitoring the local implementation of the Data Governance Model according to the Policy and the additional internal regulations. Hence, it is the contact point to which we interface through the entire development of the project. Moving now to more operational level, a first fundamental step is to decide in which way to assign the roles between two possible solutions: they can be identified either with an 'ad personam' appointment or by following the organizational chart and thus respecting, in some way, the hierarchies already present in the company structures. The choice falls on the second options, attributing responsibilities based on preexisting roles. First figures to be identified are the Data Owners, whose responsibilities are assigned to the Function Managers. A *Data Owner* is the person accountable for the data processed and the output produced within the boundaries of its controlled area. A clarification is needed with regard to the concept of 'ownership of data'. The ownership over a data/report arise in the moment in which there is a manipulation of such data, an elaboration: if data are just taken, used, and transmitted in the same form as received, the ownership over them belongs to the first sender, not to the user, which will not be identified as Data Owner. The Data Owner can appoint one or more *Sub Data Owners* depending on the complexity and the vastness of the structure and the necessity to know in detail the report produced: Sub-Data Owners have basically the task to support the Data Owner in its Data Governance related responsibilities. Going deeper, we find the *Business Data Stewards*, identified as the Head of Department. They usually have a great knowledge about one or more reports, given their directly involvement in the day to day operations that lead to the production of such outputs. Regarding Data Governance/Data Quality program, an important task of the Business Data Steward, is the definitions of the business-related controls to be executed on data. Those controls are implemented in the system and are added to the technical-related checks: the *Technical Data Stewards*, counterparty of the Business Data Stewards with

responsibilities over the information system and the software tools, are the figures appointed for the embedding in the system of the controls identified by both sides. Moreover, the Business Data Stewards are the people with actual responsibilities over the output produced by each structure.

Some outputs, as we will see later, are called Relevant Output. Basically, a *Relevant Output* is a report or a family of report that given its importance, deserve a deeper focus. The importance is given either by its public exposure, toward the Markets or the Regulator, or to its strategic relevance for internal purposes and decision-making process. At Group level, 5 Relevant Outputs have already been identified, and each Legal Entity, in accordance with its characteristic, the business carried out and the normative to which it must comply, inherits all of them. Moreover, new Relevant Output can be identified at local level. All the details of such process will be discussed in section three of the following chapter.

Next chapter will be dedicated to the detailed presentation of the second and third project streams, going more in details with activities carried out at an operational level.

Chapter 6
A REAL CASE STUDY ON A BANK GROUP: PROJECT IMPLEMENTATION

After having introduced the project and set the objectives to be achieved, we can move to the practical implementation of the contents stated in the Policy. The aim is to establish a well-structured Data Governance framework allowing the company to both satisfy regulatory requests and improve their data management. The chapter will introduce all the activities carried out and the one still ongoing, going deeper in details when necessary. Furthermore, the tool supporting the model will be presented as well.

6.1 Data Governance Policy

The internal Data Governance Group Policy is composed by a main document and several annexes. It clearly states the objective, the perimeter in scope, the scope of data, all the necessary definitions, the list of the Already Relevant Output identified, the rules through which roles and responsibilities are appointed, the list of general Data Quality principle to which output must be compliant with, the activities to carry out and all other specification and procedures needed to implement smoothly the framework. In addition, Guidelines that deepen each relevant section of the policy and the annexes have been added.

Given that it is quite useless to show the entire content of the policy (and it also forbidden given the confidential information here contained) we will concentrate just on the three directives in which it is subdivided. The three directives are associated respectively to the Data Governance Function, related to the Bank side, to the single vertical functions, associated with Data Owner, and to the Data Quality Function, embedded in the IT system and thus related do the IT part of the Bank. The directives are built subdividing the responsibilities on the basis of the DAMA Framework seen at the end of Chapter 4. The knowledge areas involved in the project are just seven over the eleven laid down by DAMA, but it is absolutely coherent with the fact that, in the first wave of a project, in

order to reach the objectives, you should start from a few sets of data-related issues to be solved. It is only in a second moment that, after having posed the grounds for building a powerful , flexible and scalable model, it is possible to both enlarge the perimeter covered and deepen the management of data, by including all the aspect of a such complicate theme like data management with a 360 degrees view. In particular, the knowledge areas covered by the Policy are: Data Governance, Data Strategy, Data Modelling, Data Management, Data Architecture, Data Protection and Data Quality.

6.1.1 Data Governance Function Directive

Being related to Data Governance, the directive is mainly related to the framework definition. Here, the perimeter of application and the activities to be developed are defined, as well as the other high-level objectives, like the coherence with corporate strategy, which should always be aligned with the way in which data are managed and the attention that must be always kept regarding to the quality of data: it is necessary that every member of the company internalize the concept that data must be always of the highest possible quality. The enlargement of the area covered is another topic contained in this section. Data strategy arguments are related to the top management, which constitutes the guide leading the Bank and taking decisions when major issues are raised. More interesting is the Data Modeling aspect, in which we find the Business Glossary. Data Modelling deals with the management, the functioning and the coherence of the Business Glossary. Furthermore, the evolution of the implementation of the tool is traced, by evaluating the coverage with respect to the whole company's perimeter. Data Management takes care of data lifecycle, by facilitating the interaction between Business and IT members. Data Architecture is mainly an IT side subject, dealing with the architectural model, both at hardware and software level, with the practical implementation of Data Quality controls and being always aware of the theme of model scalability and flexibility. This topic will be resume when the Data Quality directive will be presented, however, the Data Governance function has the important intermediary role between the Business edge who requests improvement and evidence issues and the IT one, responsible of the maintenance and enhancement of the architectural model. Data protection is a such complex topic that it has been assigned to a dedicated team, thus my vision about that is quite limited. The particular attention dedicated to the protection of data is due to GDPR, which has changed the rules of the game: having just one personal data of a customer force you to adequate your security level to GDPR

requirements. Another theme connected to protection is the accessibility to company's data: different members, belonging to different hierarchical level or different structures should have access to a different set of data. For example, an operational member should not see the information that a manager can handle; at the same time, a broker should not have access to HR data, and vice versa. Finally, the Data Quality, whose model and controls are defined by the directive object of study.

6.1.2 Data Owner Directive

The second directive is related to Data Owners and the adoption of a decentralize model with the attribution of responsibilities to the single structures, and, in particular, to the figure of Data Owner. The role of Data Owner and the delegation of power and responsibilities are pillars of the model: the power is assigned to Data Owners, which can delegate it to Sub – Data Owners and Business Data Stewards, if necessary. Also, Data Strategy highlight the fundamental cooperation between Business and IT members in order to reach Data Governance goals. The Data Governance aspects are several and remarkable. First, the identification of roles and connected responsibilities, widely discussed in the last chapter, are highly detailed. Secondly, the definition of Relevant Output and their identification process is explicitly presented. Thirdly, the activity of Business Data Element determination, assigned to Business Data Stewards, is explained. Postponing the detail analysis of them to section 6.3.3, we anticipate that a Business Data Element is one of the many single business terms composing an output. Passing to Data Modelling, it asks the single function to define the features of the Business Data Elements identified, along with the domain and the algorithms use to calculate them. After been validated by the Data Governance Function, they can be entered in the Business Glossary. Data Management refers to requirements that data lifecycle have to satisfy to opportunely evidence the control of the perimeter. Data Architecture states that Data Owner has to test the architecture of data and evaluate its performance. In case of problems or inefficiencies, it has to notify the IT in order to solve them. Moreover, it has the duty to eventually propose enhancements for such architecture gaps: thus, it is the guide for the IT side when the architecture has to be built or changed. Data protection topic resumes the concept of accessibility seen for the previous directive. The Data Owner has to define the visibility its outputs should have, and hence which company's members and structure are allowed to see them. To reach this objective, security measures must be taken and, in case such measures don't

implement what is agreed, the Data Owner has the task to inform IT structure in order to take remediation actions. The last area is Data Quality. Data Owners have much responsibilities regarding the quality of data, being them the owner of the data. The directive states that they have to define the Business Rules and the controls to verify that they are observed. Data Owners responds for all the types of Business controls made on its outputs, both manually and automatic. Finally, they also have the duty to identify new Business Rules and relative controls, which can lead to an improvement of data quality.

6.1.3 Data Quality Function Directive

The Data Quality part is probably the most relevant one in the project we are facing, since Circular 285 from Banca d' Italia requires clearness and transparency in report production, goal reachable only with a great quality of data. Needless to say, the way Data Quality is approached must be in line with Data Governance framework, otherwise non coherence will lead to a fragile model.

The analysis of the Data Quality Function Directive can be carried out along the lines of the two already seen directives. Starting from Data Strategy area, we see how Data Quality provides the information needed to the evaluations made in relation to the strategy of data. Concerning Data Governance, Data Quality deals with technical aspects of data and with tools supporting the governance of data, along with the management of the whole set of controls. Data Modelling aspects, are related to the fact that Data Quality has to support Data Governance during data lifecycle, helping with the Business Glossary filling in, both with the porting activity of terms from the template filled out by Business Data Steward and by allowing the correct association between Business Data Elements and Technical Data Components. About Data Management aspects, Data Quality has the task to share to the whole organization the way in which data extraction, elaboration and storage should be made. Moreover, it has to identify the structures with gaps with respect to the standard and the targets set, trying to delete them: Business and IT side have to work together, since the first identify the issue, the second looks for data flows involved and then they put their effort side by side for solving them. Data architecture depends heavily on Data Quality, since it evaluates the effective alignment between the logic Data Governance model and the physical implementation, correcting eventually misalignments. Also, it has the ownership over the management of governance and quality tools. Data protection, as said before, is a particular area which has been addressed with a deeper focus, but it is not included in the current project. Nevertheless, it has been highlighted the

fundamental role of Data Governance in handling the access rights, defining which functions and member can access to specific data and/or output. Finally, Data Quality Function is responsible for the Data Quality software installation and training of Technical Data Stewards. The directive states that IT figures have to support Business Data Steward in controls identification, having them a more technical view.

6.2 System of Governance definition

In chapter 5 we have seen that a decentralized model with a central low powered coordination body has been adopted, with role and responsibilities been assigned following pre-existing organizational position. Here we deepen the analysis of the role and the tasks associated to any of them and draw the System of Governance.

Two types of fundamental actors are involved in Data Governance processes: on one side the Business members, representing the Bank, on the other the IT members, having control and responsibilities over the information systems and the software tools the whole Bank uses. It is basilar to clearly subdivide the responsibilities and the ownership of each single actors in a company. This is especially true when actors belong to two different yet strictly interconnected worlds and cross functional interactions are a daily occurrence, like in this case. Obviously, the two side have different short-term objectives and way of thinking about data-related topics, but it is necessary to understand that the final objective should be a coordinate management of data. The success does not only depend on a well-defined subdivision of tasks among members, instead lies in a harmonization of visions. Thus, if on one side it is important to appoint ownership in a way that both overlapping and uncovered areas will not occur, on the other an integration between Business and IT areas and a share vision could not be missing if the goal is to manage data effectively. All this consideration lead to a logical model whose main features can be summarize in *Figure 17*.

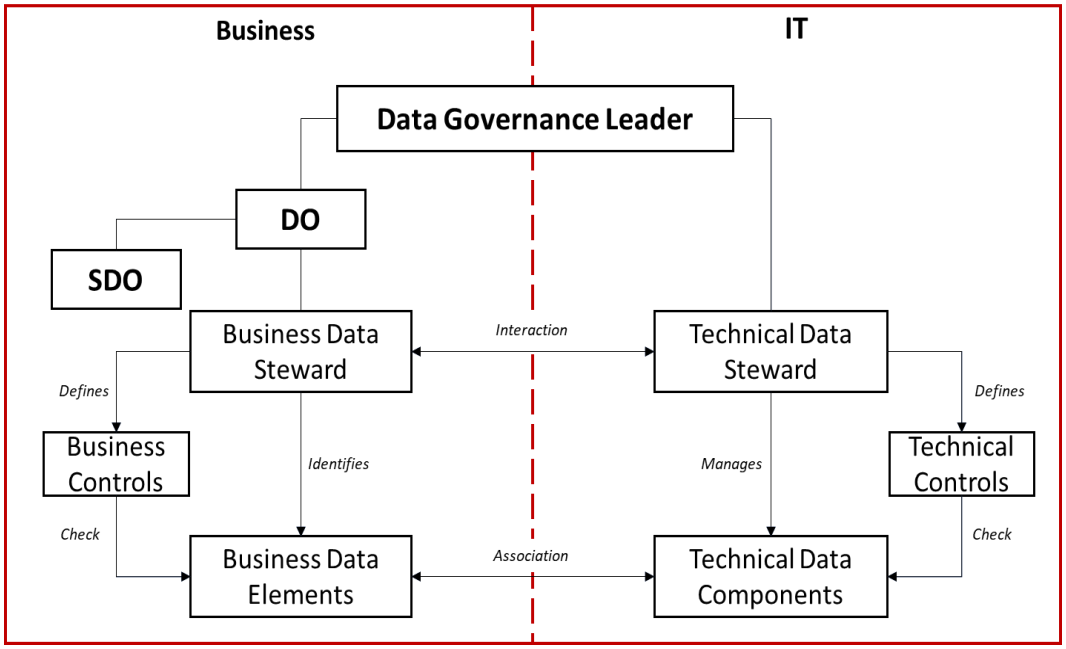


***Figure 17*: Data Governance Model in the real case study**

First, the Data Governance Leader who has direct responsibilities over the project and the coordination role, thus is the figure to which we interface with during the entire development of the project. Then the Data Owner, probably the most important figure of a Data Governance framework. Along with the Sub Data Owner, which is eventually appointed directly by the Data Owner in case of complex structures, it has the duty to identify the outputs under his ownership, first and foremost the Relevant ones. It has been already described a Relevant output as one which goes either to the Market, the Regulatory Authority or the internal Board, thus it is an output whose importance in terms of compliance to regulations, reputation or decision-making support is marked. Another job assigned to the Data Owner is the identification of the possible risks that can occur due to poor Data Quality. This process is carried out through the identification of business rules defining the conditions characterizing the right occurrence of an event and hence the relative controls verifying such happening. In section 6.4 we will see how the monitoring activity over Data Quality is make through the use of Key Quality Indicators (KQI), defined by Data Owner themselves and implemented by the IT functions. Last task of Data Owners is the appointment of Business Data Stewards.

Business Data Stewards have direct responsibilities over one or more outputs. Note that an output is under the ownership of just one single organizational member, but one owner can be the responsible for more than one output identified: this will clearly avoid any overlapping of ownership which would cause internal dispute and difficulty to address in the case that an issue

concerning such outputs would arise. They also have the important duty to determine and list the Business Data Elements and filling in the Business Glossary with the Data Elements identified. Another fundamental task of Business Data Stewards is the identification of the necessary controls to be implemented over each data flow under their ownership, supported by the Technical Data Stewards. This is a fundamental step, since the real ground of Data Quality is placed here: on the one side the Business actors, who have the knowledge and the sensitivity to understand which type of controls on data should be implemented; on the other the IT actors, which have the technical skills to implement in the best way the controls in the information systems. The coordination of the two faces of Data Quality should perfectly coordinate, in a way that when a Business Data Steward asks for the implementation of a new control or a change to an old one, the IT should activate to satisfy the requests, as far as possible. It is clear that Data Owner, Sub – Data Owner and relative Business Data Steward, work together for the develop the best framework possible. For example, most of the times the Business Data Steward, having a much granular visibility and a more operational approach, helped their superiors to establish the business rules. In addition, sometimes it was necessary to involve also people that are placed below Business Data Steward in term of hierarchical level. It is necessary in such areas that manage either a high number of outputs or are highly structured due to the specific peculiarities of the activities carried out.

For what concerns the IT part, the figures identified are the Technical Data Stewards appointed to implement the controls requested by the Business actors and to maintain the information systems. Their role will be better defined in paragraph 6.4 where the Data Quality part of the project will be explained.

We have seen that the decision is to assign the roles based on pre-existing positions. Hence, starting from the organizational charts sent to us by the company, we apply the rules stated in the Policy and identify the Data-Owners and the Business Data Steward. This allow us to define a first draft of the *System of Governance*, the first pillar of Data Governance: it consists of the set of the roles and responsibilities, defined to ensure a proper adoption and implementation of the Data Governance model. In particular, for what concerns the Business Data Stewards, we just made a hypothesis about who to appoint, since as said, it is a Data Owner responsibility to appoint them. Same consideration applies to Sub-Data Owners. Hence, during the introductory meeting with the Data Owner a confirmation of such assignment is asked with the aim to design the official System of Governance. Moreover, the Technical Data Stewards cannot be identified from the organizational chart, but only after a deeper understanding of

company procedures. Given that, they have been identified during the awareness/assessment phase of the project.

6.3 Inside the core of the activities

6.3.1 Data Governance Awareness Program and Output collection

A fundamental activity to do in a project like this, is for sure give an evaluation of Data Governance Maturity Level. Defining an initial level allow both to identify potential gaps and to set a basic level shared by all the actors involved in the project to be compared with the targeted one. At the end of chapter 4 we already introduced a maturity model, the CMMI DMM, which subdivided the spectrum in 5 different levels: performed, managed, defined, measured, optimized. Going from the performed level to the optimized one, the awareness about the significance of Data Governance and Data Quality rise exponentially. In the meanwhile, we pass from a first level in which problems are solved when they occur without mitigating the causes and risk assessment is made only with relation to critical data, to a one last level in which process are completely automatized, standardization of activities are absolute and the value of data is formally recognized and quantitatively evaluated.

Such an assessment has been carried out through a set of interviews made with the representatives of the functions, which is probably the best way to have a clear view about the as – is situation of an organization. The meetings have the aim of both evaluating the initial situation of Data Governance and Data Quality, but also to spread the contents of the Group Policy.

Starting to the Bank side, Data Owners identified in the System of Governance are the first people to be interviewed. During an introductory awareness meeting, we ask them to confirm the Business Technical Stewards under its controlled area and eventually identify one or more Sub Data Owners which should help them in getting the things done. Then, similar meetings are held with the Business Data Stewards to show them the main contents of the Policy and get a first impression with regards to the awareness level about Data Governance related themes.

Moving to IT side, we find a better understanding of such topics, mostly concerning Data Quality. What comes out is that the Bank already uses a few information systems, for different company areas, able to perform controls and collect anomalies and resolution of the ones failed. The information systems are extremely powerful and well designed for the tasks they have to carry out, however, they are quite poor in term of Data Quality, containing just a small set of

controls with respect to the regulations requirements. Various controls are implemented by the single Business functions manually, through an Excel file. One major issue, in addition, is the lack of a standard procedure for translating the requests of a Data Quality control from the Business side to IT. It does not exist a clear and formalized methodology for the identification of the controls by the business members and the subsequent request to the IT for their implementation: this often results in the loss of the possibility to enhance the quality level of data.

As expected, the initial situation is anything but positive. Although most of the respondents are aware of the importance that the quality of the data produced by a Bank must have, for the majority of them Data Governance was a totally new topic. And the situation on the Data Quality side is even worst. The information systems are present just in the most important functions, such the CFO or the Asset Management, but their ability to perform control is quite limited, forcing the Business members to carry out several controls by hand or through Excel templates properly structured. For the other functions, the control process is completely manual with the support of Excel templates. It is clear that this process contains numerous shortcomings and gaps in terms of Data Quality. First, the information systems are not able to perform the whole sets of controls required, mostly the ones related to the new Data Quality principles introduced by the last regulation. Second, for all the functions a consistent part of the control process is based on the experience of the workers. Yet, the human ability in the control of file Excel with thousands of columns and rows is obviously limited. Thus, it is a strong Data Quality gap to base most of the activity on data which are just manually controlled without properly tools: the human operational failure is right around the corner.

Furthermore, given the novelty and the complexity of the theme, they immediately refer us the need to implement the Data Governance model through a step by step approach, since the activities are all “on top activities”, thus adding to the already massive amount of work they have to carry out as daily operational activities.

After the first turn of interviews we define the perimeter in scope composed by 21 Data Owners and 41 Business Data Stewards and only 2 Sub Data Owners. For what concerns the IT side, the decision is to appoint just 3 Technical Data Stewards. Delimited the scope of application, a second turn of meeting is organized. One-to-one meetings are scheduled whit each Business Data Steward, and possibly with the presence of the Data Owner (and the Sub Data Owner where exists), in order to present an Excel template in which detailed information about each output produced in its department is collected. An extract of the Excel template used is shown in *Figure 18*:

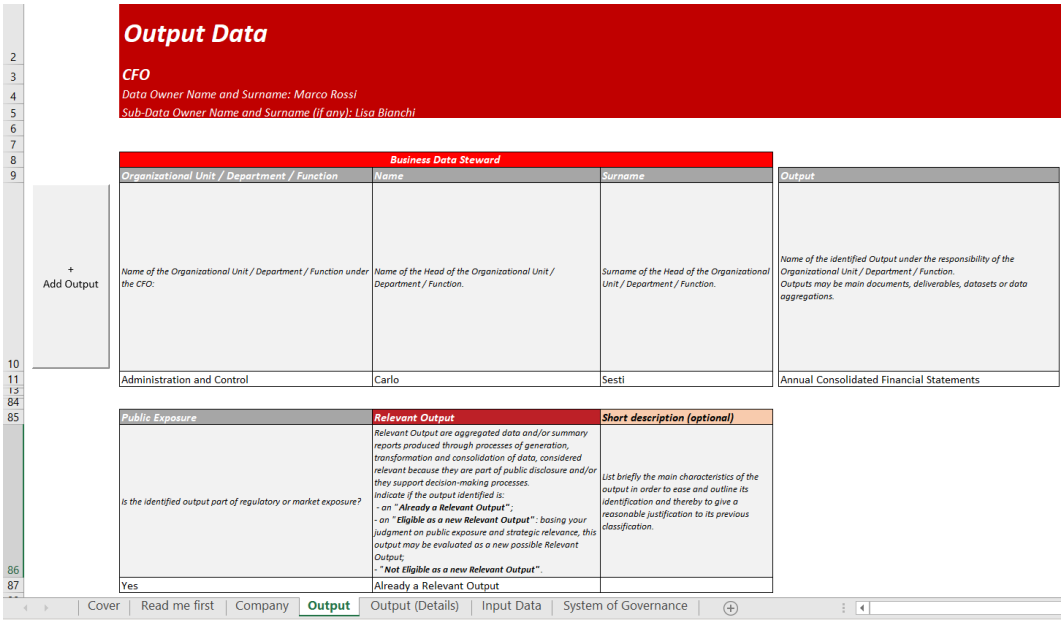


***Figure 18*: template used for the collection of outputs produced and relative detailed information. All data contained are purely indicative for respecting confidential agreement.**

The template is composed by several sheets and embedded with macros that, at end, allow to make a recap of all the information inserted. The Cover and 'Read me First' are purely introductory sheets explaining the objective of the file and guide the user through the filling in, so they can be skipped. The third sheet, Company, contains just some general information of the company in scope, thus can be skipped too. The "Output" sheet is the one represented in *Figure 18* above. At the top, in the red banner is showed the name of the function and the name of both the Data Owner and Sub-Data Owner (if any). Then, you can find the name of the Business Data Steward, together with the department he belongs to and in column F the list of all the outputs produced by such business unit over which the Business Data Steward has responsibility. Finally, some information about its relevance are asked and the output can be tagged either as 'Not Eligible as a new Relevant Output', 'Eligible as a new Relevant Output' or 'Already a Relevant Output'. The concept of Relevant Output has been previously seen. They are important outputs that has been defined as Relevant and undergo a specific activity of Data Directory, that will be briefly explained in paragraph 6.5. Five Relevant Outputs have been identified at Group level and they apply at every single Legal Entity in scope. As stated by the Policy, however, new Relevant Outputs can be elected at local level, in order to give voice to the important peculiarities of each subsidiaries. It is worth to remember that the holding is a bank, while the subsidiaries controlled provides many different services, from insurance, to investment to asset management. It is clear that each of them has

particular regulatory requirements and perform specific activities, due to the business carried on, that can lead to the production of significance reports for which a greater focus is required: those outputs are identified as possible 'New Relevant Outputs'. In this case the report is tagged as 'Eligible as New Relevant Output', and further consideration are necessary to understand whether to promote it or not. 'Not Eligible' outputs are the one that, given a lower significance in term of strategic relevance, are not consider as candidates to be promote. The 'Output (Details)' sheet go deeper analysing each output, asking for example the system used to produce it and a self-assessment from the Business Data Steward about the perceived quality of it. This a key part of the tool, since for each of the Data Quality Principles listed in the Policy, the Business Data Steward gives his self-evaluation about the level to which an output is compliant with it, calculating finally an overall qualitative Data Quality Level, either 'High', 'Medium' or 'Low'. Further details about the Data Quality Principles are show in section 6.4, when Data Quality topic is addressed. The second to last sheet, 'Input', asks for all the data flows concurring to the production of each output listed in the previous sheets. Such information is useful to carry out a first reconstruct of the Data Lineage of the outputs, really useful to find Data Quality gaps and addressing them quickly. Finally, the last sheet, creates a table through a macro that recap all the information entered in the previous sheets returning a compact dashboard.

An initial analysis of the results coming from the meeting has been performed and a first level of non- exhaustive outcomes is represented in *Figure 19*, *Table 4* and *Figure 20*.

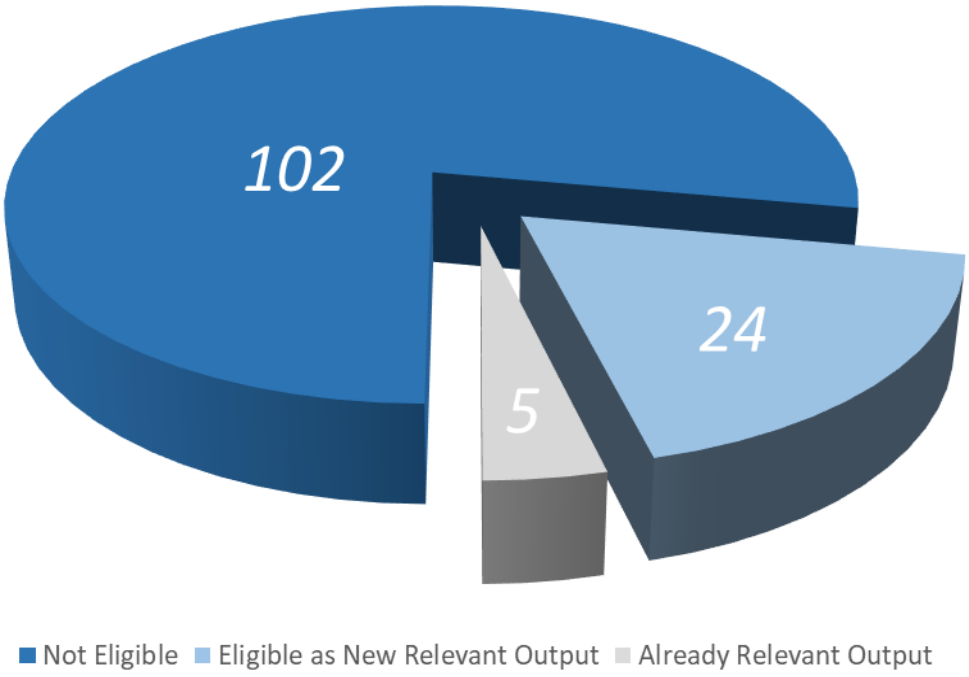


***Figure 19*: outputs in scope subdivided by category**

Function/Area	Data Owner	Business Data Steward (Total)	Output (Total)	Not Eligible Output	Eligible Output	Already Relevant Output
Asset Management	Marco Sesti	11	31	28	3	0
CRO	Giulio Rossi	3	12	6	4	2
CFO	Anna Rossetti	7	20	14	3	3
Transactions	Claudie Dubois	4	6	4	2	0
Other Functions	[…]	16	62	50	12	0
Total	21	41	131	102	24	5

***Table 4*: outputs collection detailed recap**

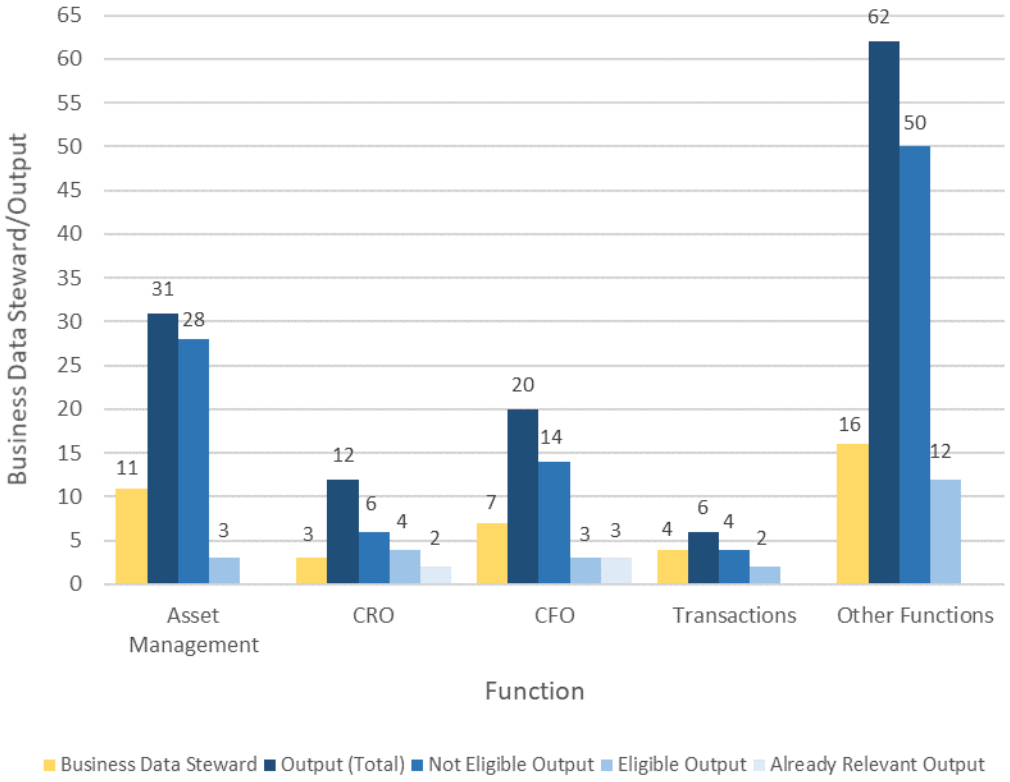


***Figure 20*: outcomes of the outputs collection**

Some evidences came directly out from the activity carried out. First, based on the general awareness and the procedures applied over data in the whole Bank, the Maturity level is equal to Managed, thus a lot of improvement could and should be implemented. For sure, this first awareness activity has improved substantially the sensitive of organization member about Data Governance and Data Quality topics, but much is still to be done. Secondly, a big consistent number of reports produced is tagged as 'Eligible', 24, thus a prioritization is necessary to understand which of them to promote as 'New Relevant', given the non-affordable effort required to promote all of them.

6.3.2 Prioritization of Relevant Outputs candidates

From the set of 24 outputs identified as 'Eligible', a skimming process is necessary to define a small subset that will be suitable to be promote. It is worth to evidence that the 'Eligible' tag is assigned just on the basis of the Public Exposure or the strategic relevance of the output, but further discussion is

necessary to make clarification about their features. The decision, also guided by the Policy, is to organize, for each of the 'Eligible' output, another turn of meetings involving both the Business Data Steward and the respective Data Owner.

After the meetings, only 11 over the 24 outputs survive, while the other will be evaluated periodically for a possible promotion in successive waves. Given the activities procedure applied to Relevant Outputs, 11 'Eligible' reports are still too much and a prioritization is required in order to make a second skimming and obtained a little group composed just by the most significant reports.

Some objectives criteria are used with the aim to make a prioritization as much impartial as possible. They are listed below:

- ***Regulatory constraints***: outputs subjected to normative and regulatory constraints that impose not extendable deadlines and well-defined output structure. Regulatory constraints not yet approved but on which the authorities are actually focusing their efforts are also considered;
- ***Data Quality activities already on goin***g: some structures have already implemented a set of Data Quality pervasive controls which ease the implementation of a structured Data Governance/Data Quality framework. Moreover, usually, different projects addressing the management of data are ongoing in parallel: this will ease the work too;
- ***Availability of resources***: the effort required to company's member is not negligible, mostly in decentralized model. Thus, it must be taken in consideration the availability of them;
- ***Timing***: it is better to start from the simplest areas and leave at a later time the more complex and demanding ones also leveraging on what just done;
- ***Amount of data involved***: as much data are involved in the production, as much the possibility to make improvements.

At the end of the long skimming process, just 4 outputs are saved, and the next step is to discuss them with the Data Governance Leader and the Data Owner accountable for it in order to understand if it should be promoted or not. Such a complex process might seem quite useless, but it is absolutely necessary. The effort in term of time, money and resources required to perform the activities characterizing a Relevant Output are considerable, thus it is in the interested of the organization to promote just the real significant reports, maybe leaving the other to a second wave. In fact, for each Relevant Output the demanding activity of filling in a Data Directory is required, as we will show in paragraph 6.5.

At the end of the long process, just 2 over 24 ‘Eligible Outputs’ have been decided to promote. It means that the local number of Relevant Outputs is now 7: the 5 inherited from the Group and the 2 new identified.

6.3.3 Business Data Elements collection and mapping of Controls

The activities carried out with the Business Data Steward do not stop with the collection of the output produces within their domain. Indeed, as stated by the Policy, they have two other important tasks: the collection of Business Data Elements and of the controls performed on the reports under their ownership.

With the regard to the first activity, a *Business Data Element* “is any atomic unit of data that has precise meaning or precise semantics”[40] within the business field. It is a ‘logical definition of data and, although in practice they are used in an interchangeably way, it should not be confused with the terms ‘field’ and ‘data items’. *Fields* are the actual storage units, and *data items* are the individual instances of the data elements. For example:

- Business Data Element: product
- Fields: code, name
- Items:
    - P100, charger X
    - P178, mobile phone Y

Given the impossibility to assist each single Business member in the identification of the Data Element under its control, a template and helpful guidelines are prepared and sent to them, actual owner of Business Data Elements, to insert all the Data Elements under their domain. The porting of information from the template to the Business Glossary is an activity that is going to start in the next few weeks, thus the results, at today, are not available.

The second activity, time demanding as well as the previous, ask to Business Data Stewards, supported by their technical counterparty, to reconstruct a high-level Data Lineage of the outputs which they are responsible for. Given the big effort required by this assignment, the decision is, at least in an initial phase, to perform it just for the ‘Eligible as New Relevant’ outputs and the ‘Already Relevant’ ones, due to their greater importance. The main objective of the activity is to identify actual and eventually implementable Data Quality controls. Firstly, the Business Data Steward is asked to draw a high-level architecture model for its

[40] See en.wikipedia.org/wiki/Data_element

output, adding all the controls already implemented, both at business and technical levels, and the ones suggested to improve the quality of data. An example is reported in *Figure 21* below, where Business As-Is controls are highlighted in yellow, Technical As-Is controls in blue and suggested implementable control in azure. Furthermore, it is also indicated the Data Quality tool that will be implemented which will handle, among other thing, the whole set of controls related to reporting production.

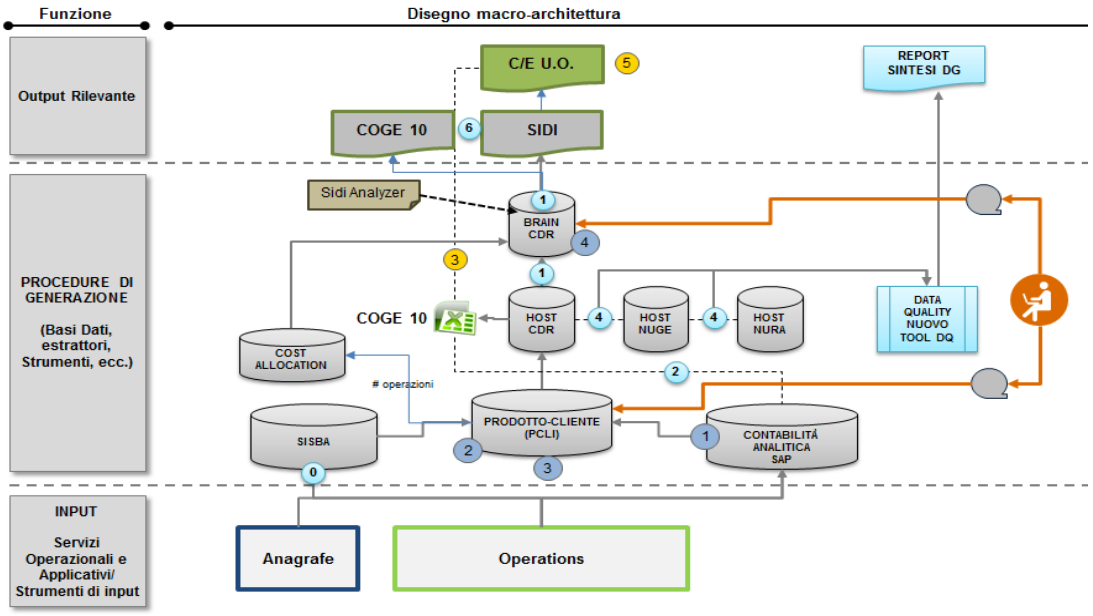


***Figure 21*: As – Is and To- Be architecture with controls identification**

Then, more details were asked relating to the controls identified on the architecture. The Business controls are in fact classified in three different categories, and each of the check is requested to be insert in one of the following three classes: pattern controls, reconciliation controls, consistency controls. *Pattern controls* verify that a trend, for example, is in line with the values measured in past months; *reconciliation controls* check that data coming from different sources but which should have the same values are not discordant one another; *consistency controls* include all the situation not addressed by the previous two categories. Technical controls are procedural controls, that evaluate if a process has been carried out correctly verifying the outcome of the upload.

6.3.4 Business Glossary and Dictionary of Controls

A pillar of a Data Governance framework are the tools supporting the operative implementation of the model. The tools necessary for an effective development, at least for this first phase, are the Business Glossary and the Dictionary of Controls. They are strictly interconnected instruments, which help the organization to have a list of all the Business terms used in order to have a single and clear definition of each one of them shared among all the members. It may seem strange, but it often happens that within companies, different people attribute different names and / or meanings to the same concept, creating confusion and the impossibility of sharing trustable information. It is therefore fundamental that all people have a shared vision about what data flows within the company and what is the meaning attached to those. Needless to say, the major problems that different definitions or names for the same business – related term can create. Such differences, usually, are even more marked between IT members and Business ones, which speak two completely different languages: the first related to info systems, the second to business. In addition, the Dictionary of Controls, contains the controls to be executed over each Business Data Element.

Let's start from the *Business Glossary*. Also called *Data Glossary*, is the complete list of business terms (Business Data Elements) with their definitions. It defines the whole set of business concepts used by an organization, independently from any specific database or vendor. Other information contained are the functions to which the Business Data Element belongs, synonyms, where it is used, the owner, which is a Business Data Steward, the respective Data Owner, the access rights and so on. Note that a Business Element can appear in one or more output and have a single owner. Important to understand is the difference between a Business Glossary and a Data Dictionary. A *Data Dictionary* is a detailed definition and description of data sets (tables) and their fields (columns). This specification includes information such as data type, size, allowed values, default values, constraints, relations to other data elements and meaning/purpose of data set and field[41]: it usually considered an IT tool. Hence, what makes a Business Glossary different is that, although it can contain some information related to the IT side, the main focus of the content collected is information designed to improve business understanding and use of data. It is not the case that in must be filled in by Business actors and not from Technical ones. Moreover, each project should have its own Data Dictionary, containing term and IT information related to such specific project, while the Business Glossary is usually company-wide, and should cover all the terminology used. *Figure 22* show an

[41] See dataedo.com/blog/business-glossary-vs-data-dictionary

example of a page of a Business Glossary related to the element ‘Purchase Orders’.

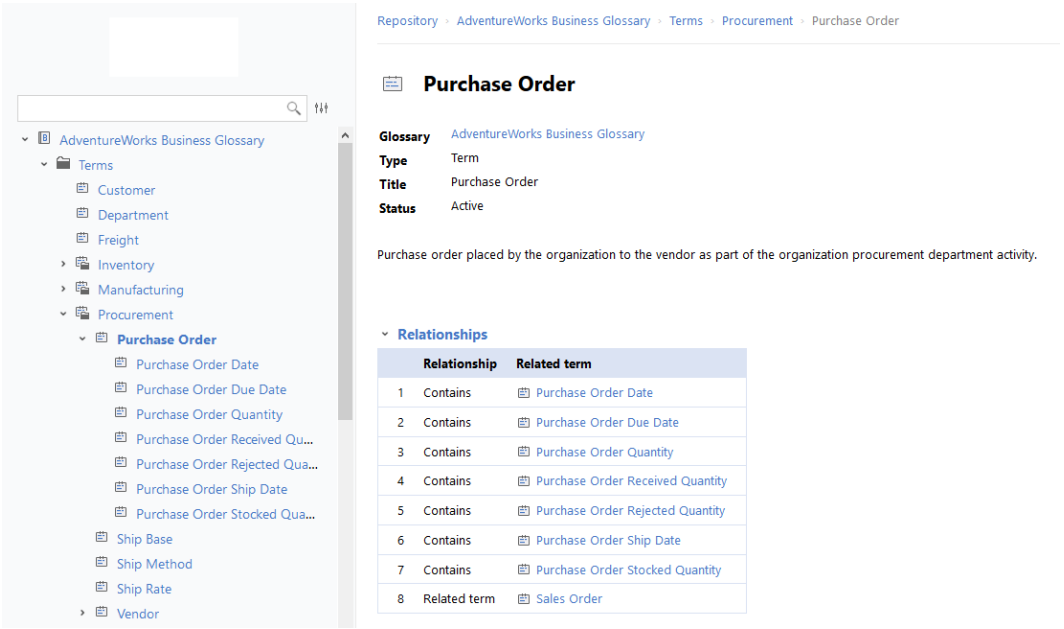


***Figure 22*: example of a Business Glossary page**

Clearly, business term can be linked to specific tables and columns in a Data Dictionary to create a much more integrated system. In such a way, you have a both ways connection allowing you to search for a definition from the Data Dictionary and have access to all the instances of the data related to a specific term when you are on the Glossary. An example of link is reported in *Figure 23*, where a page of a Data Dictionary is shown and the links to Business Glossary are inserted in the red rectangular.

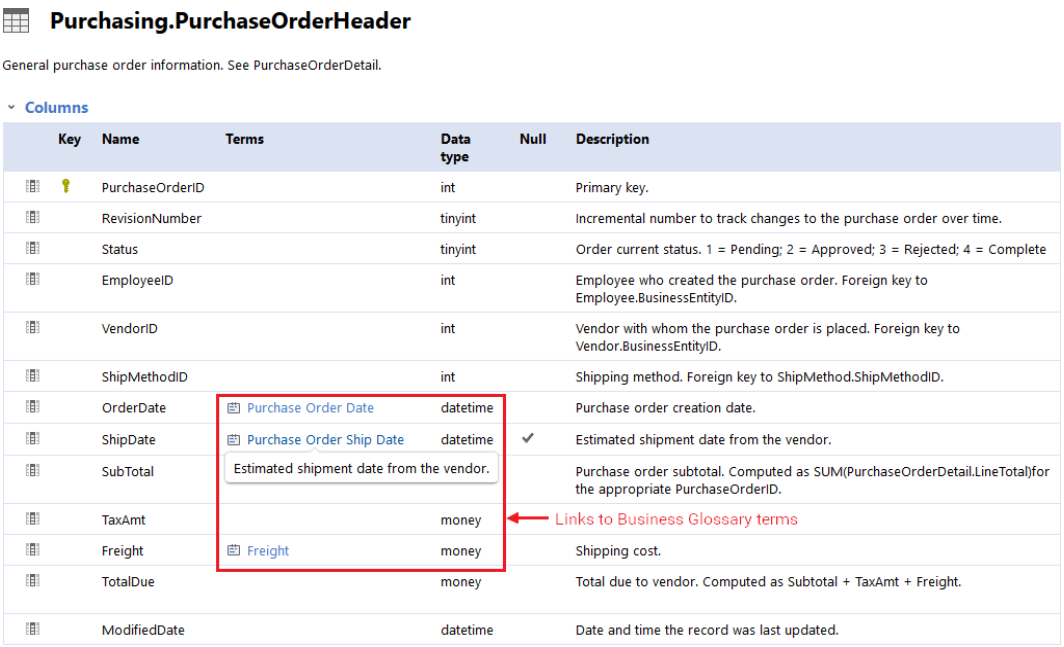


***Figure 23*: example of Data Dictionary with links to business terms**

The implementation of the new Business Glossary replaces the old incomplete one that the Bank was using. It contained only some business terms related to a subset of outputs, without a precise and structured methodology. There was no clear ownership over them, and the information inserted were not detailed enough, with the consequences that the tool was just unusable. Moreover, most of the members did not even know the existence of it. Our job thus is to both broaden the amount of Business terms contained, adding the relative deeper details, and aware all the members of the presence of such tool in the information system. As said before, this task is going to start.

The second tool which is worth to focus on is the *Dictionary of Controls*. As the name suggests it contained all the controls provided for every Data Elements and data flows, along with other relevant information. First of all, every control has an owner. As defined by Data Governance model, the owner of a business control is always on the Business side, which has the tasks to define it and verify that is well implemented. Nevertheless, the execution responsibilities are on the IT side hands. Other information contained are the Data Owner, the IT system supporting control execution (if any), the type of control, either manually or automatic and so on. Then, for each control, is indicated its goal and the risk controlled: a check, in fact, have to address a potential risk and avoid that it occurs, raising warnings to the user when something is not in line with Business rules happens. Imagine for example to insert a value that is outside the domain, a text value in a cell accepting only numeric values or to enter two different values

in two cells requiring the same one. Those are just three examples of the thousands of errors that can be committed. At last, the Dictionary must contain the classification of the control, which means the Data Quality General Principles it covers, the metric used to evaluate the result of the control, the result of it, which can be either 'Ok', 'Warning' or 'Alert', and the remediations action in case of failure. These concepts will be recalled in the next paragraph.

As we have seen, during a previous phase, we asked the Business Data Stewards to draw the Data Lineage of each output, clearly identifying actual and implementable controls. In addition, a template in which inserting the major information related to the controls is prepared and filled by Business Data Stewards supported by Technical Data Stewards. The results coming from this activity are transferred to the Dictionary of Controls, which is incorporated in the new Data Quality tool (highlighted in azure in *Figure 21*), core of the Data Quality Function.

Data Quality controls can be semi-manually implemented through Excel, creating the necessary controls using formulas, data validation and macros. And we have seen how the company under analysis makes extensive usage of this spreadsheet, thought the complexity and huge amount of data treated. But the fact is that this type of approach can only work in small reality. When the amount of data is massive, like in the case of a Bank, and the normative to be compliant with is highly demanding, the adoption of a dedicated software tool is the only possible solution: using Excel and semi manually or full manually control is the ground of possible enormous Data Quality problems given by human limits. Moreover, the information systems used, as stressed, do not performed all the necessary controls that data should undergo.

Hence, the decision is to adopt a Data Quality tool in which to embed the Dictionary of Controls, moving all the controls on it, both the old and the new identified ones. The platform will improve the overall Data Quality level, allowing to create dashboard and calculate KQI that will ease the monitoring process.

A software selection activity is thus required, to look for the best tool already available on the market, since the internal implementation option was excluded since the beginning, due to the effort required both in terms of money and time. One of the drivers which guide the choice is the integration with other Data Governance/Data Quality tools already present in the Bank and the other systems and database architecture. More important, is the theme of usability and user-friendly tool. Given that the model adopted is a decentralized one, the tool needs to be used by many people belonging to different structure, thus a non-intuitive software will damage even the most perfect model that can be created. If users perceive the software as an obstacle to their activity and not as a supporter,

the entire project would be a failure, since they will avoid its usage. Therefore, the tool must be extremely simple, helping the user and ease its work; in the meanwhile, it has to be easily maintainable, without requiring long period of maintenance from expensive specialists.

6.4 Data Quality Framework

6.4.1 Data Quality Monitoring

The Data Quality part is probably the most relevant one in the project we are facing, since Circular 285 from Banca d' Italia requires clearness and transparency in report production, goal reachable only with a great quality of data. Needless to say, the way Data Quality is approached must be in line with Data Governance framework, otherwise non coherence will lead to a fragile model.

The Data Quality logical model implementation is mainly guided by the normative requirements imposed by Circulation 285, implemented in the directive presented in paragraph 6.1.3. According to normative provisions, it has been drawn a classification of the "dimensions" of Data Quality, also called "Data Quality General Principles", to which the entire set of data flows and outputs within the company should comply with. This taxonomy, applied in the same way in each organization function to give a unique common framework for the entire company, has the aim to reach the target quality level homogeneously. Hence, for each principle has been established a unique definition and the relative controls to evaluate the level of compliance to it. *Table 5* shows the taxonomy of Data Quality dimension, along with the controls provided to assure that they would be reached:

Regulatory Requirement	*Quality Dimension (Definition)*	*Control Objective*
Accuracy	**Accuracy**: data are produced without any material error and omission, they don't include a material error, they are properly verified. Data are timely and consistently recorded over time	**Accuracy**: verify that data don't contain material error

Appropriateness	**Appropriateness**: data are consistent with data from different time periods used for the same purpose and they are pertinent, suitable and consistent with the purposes for which it will be used	**Coherence**: related information contained in different sets must be coherent **Consistency**: related information contained in the same set must be consistent **Trend**: variations from past values must be "in line"
Completeness	**Completeness**: data contain sufficient and complete information for the purpose for which they are used for	**Coverage**: all the necessary fields are present **Existence**: NULL is not an acceptable value
Integrity	**Integrity**: data are kept intact as they were recorded, unmodified without permission	**Integrity**: verify that data have not been manipulated through illegitimate alterations
Timeliness	**Timeliness**: able to generate aggregate and up-to-date risk data in a timely manner	**Timeliness**: verify that up-to-date data are promptly available to the user when he need them
Accountability	**Audit Trail**: the procedures for extraction, management, aggregation and use of data are documented, with audit trail	**Auditability**: trace of the changes made to data allowing the possibility to analyze all old versions and owners of amendments

***Table 5*: Data Quality Dimension/General Principles**

In addition, two classification for the controls are provided, one related to the position in which it is located with respect to the entire data flow and one considering the nature of the controls.

Considering the first categorization, a clarification is needed. If we analyze the lineage of a set of data, from its birth to the last moment in which it is used,

we see that it moves from a database to another, passes through system, undergoes different elaboration or aggregation and each of those steps is a possible source of errors. Clearly, being this process a chain, an error in a single step will cause the creation of a final information which is useless and not trustable. Thus, there must be inserted controls along all the phases of data lifecycle from extraction to the last elaboration: this led to a three-level categorization of controls. A first level of controls should be provided during the *extraction* phase, the one in which the data is created or acquired from alimenting systems, either internal or external. Then, we find the *feeding* phase, which is characterized by two types of controls: reconciliation and trend ones. Lastly, the controls placed just before the *release* of the final output, adding a last layer with the aim to eliminate any eventually errors survived to the first two filters. The entire process should assure the production of valuable data, without the risk of incurring into reputation damages caused by wrong reporting toward Regulatory Authorities.

On the other hand, considering the nature of the controls, it has already been highlighted how they can be subdivided in business and technical ones. Business controls verify the contents of data, evaluating the logical meaning and thus requiring a deep knowledge of the thematic to be implemented and to give a judgment to the results. They are, clearly, defined by Business members, which communicate to the IT side the necessity to develop them in the systems. Instead, technical controls evaluate the format of data, the process and eventually anomalies in data flows, without focusing on the meaning of such data. Those are under the Technical Data Steward ownership.

Every control is accompanied by threshold values necessary to control the validity and coherence of the entered value. Boundaries are defined either by Business or IT actors, depending on the type of control we are talking about, in accordance with Data Governance predetermined objectives. The, controls listed and implemented in the related Dictionary of Controls, based on the threshold values just presented, return a feedback which could be either 'OK', 'Warning' or 'Alert' ('KO'). The first answer, 'Ok', indicates that the controls has not found any issue or inconsistency on the values controlled, thus giving a reassurance about the quality of data checked. The 'Warning' feedback laying in the middle between a successful and an unsuccessful control. It means that further analysis is necessary in order to understand if an anomaly has occurred. For example, a 'Warning' can be raised if the rules defined on business terms have not been completely satisfied, but the deviation from the acceptable range is small. On the contrary, the 'Alert' result indicates that a control has gone KO and a remediation action needs to be taken by correcting the values entered. To standardized activities and setting guidelines to be follow by every structure, the remediation in case of 'KO' feedbacks procedures are contained in the Policy document, with the

leading rule stating that issues must be solved not just temporarily, but it must be found the causes and eradicate them.

To have an overview about the whole level of Data Quality and continually monitoring the level of development, the directive requires to define specific Key Quality Indicators (KQIs) able to summarize the overall capability to produce high quality data, where with high quality we mean the ability of such data to comply with the dimensions listed before in *Table 5*. In particular, the outcomes of the controls are traced, and the indicators are used to give a synthetic information to the Data Owner about the overall quality in their structure. Some "Overall KQIs" are defined at central level, but, according to the decentralized model, greater freedom is left to Data Owners which are empowered to define "Specific KQIs" in order to give voice to the specific peculiarities characterizing each function. Nevertheless, all the indicators must be aligned with Data Governance framework and the objectives it poses, thus some guidelines are given to Data Owner in order to support them in the implementation of the directive. *Table 6* presents some possible KQIs that can for sure give some useful suggestions about the data quality status in the Bank.

Dimension	*KQI*	*Formula*	*Description*
Conformity	Identify the property of a data of being error free	$\frac{\sum_{i=1}^{N} error(v_i)}{N}$	*Numerator*: sum of the number of records with errors. *Denominator*: total number of records considered
Timeliness	Identify the update level of data, how much data are up to date	$max\left(1 - \frac{currency\ (v_i)}{volatility\ (v_i)}; 0\right.$	*Currency*: indicates the degree of updating of data (time frame between creation and availability of data) *Volatility*: indicates the average time of use of data

***Table 6*: list of possible KQIs for Data Quality Monitoring**

6.4.2 Data Quality Gap Analysis

After having introduced the Data Quality General Principles and the Data Quality framework, we show some results of the *Data Quality Gap Analysis* carried out on the basis of the information received by the company actors through the interview activity and the collection of evidence within the template referred to in the paragraph 6.3.1. During the collection of the list of outputs produced, detailed information are also collected, including a self-assessment by the Business Data Stewards about the quality of the reports they have responsibility on. This self-assessment is based on a tripartite-scale judgment (low, medium, high) regarding the coverage level of each report with respect to each of the general principles. For each output, the Business Data Stewards are asked to judge the degree to which it meets each of the six Data Quality dimensions. The first analysis that can be made is the one relating to the overall quality level of the reports produced, function by function. The overall level of quality of an output is obtained as the average of the assessments attributed to each of the individual General Principles identified in paragraph 6.4.1 for that specific output. *Table 7* shows, for each function, the number of total outputs produced and their subdivision according to the overall quality level. The results are graphically represented in *Figure 24*, where we see, for example, that the 68% of the repots produces within the Asset Management function have a high quality, 22% medium and 10% low.

Function/Area	Data Owner	Business Data Steward (Total)	Output (Total)	Output Data Quality		
				HIGH	MEDIUM	LOW
Asset Management	Marco Sesti	11	31	21	7	3
CRO	Giulio Rossi	3	12	8	2	2
CFO	Anna Rossetti	7	20	15	5	0
Transactions	Claudie Dubois	4	6	4	2	0
Other Functions	[…]	16	62	27	26	9
Total	21	41	131	75	42	14

***Table 7*: list of reports function by function, classified by Data Quality level**

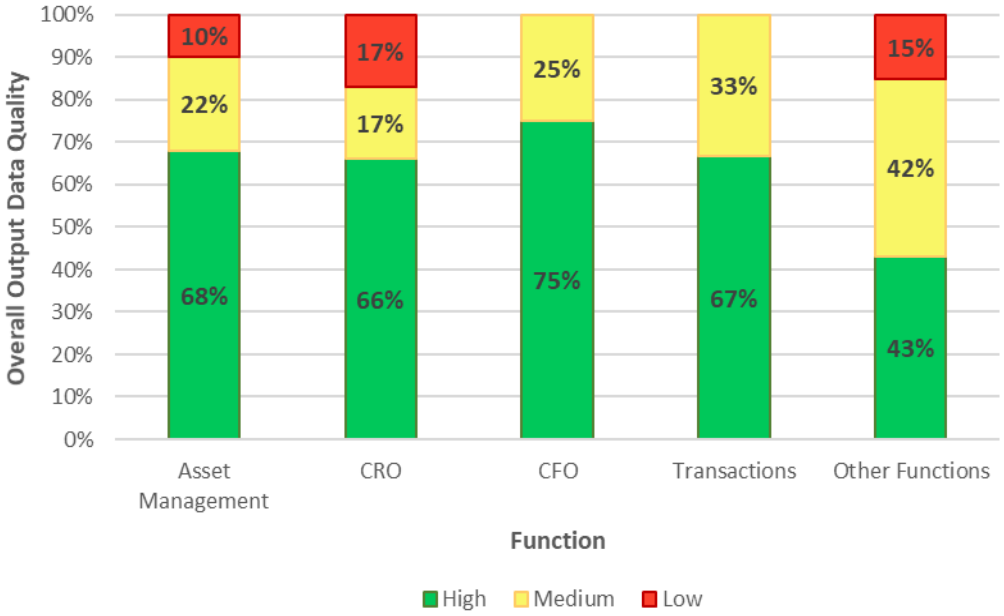


***Figure 24*: overall output Data Quality level identified in each function**

A similar analysis can be performed with regard to the input received by each function. *Figure 25* shows, for the same functions of the previous graph, the Data Quality level of the input concurring to the creation of the reports.

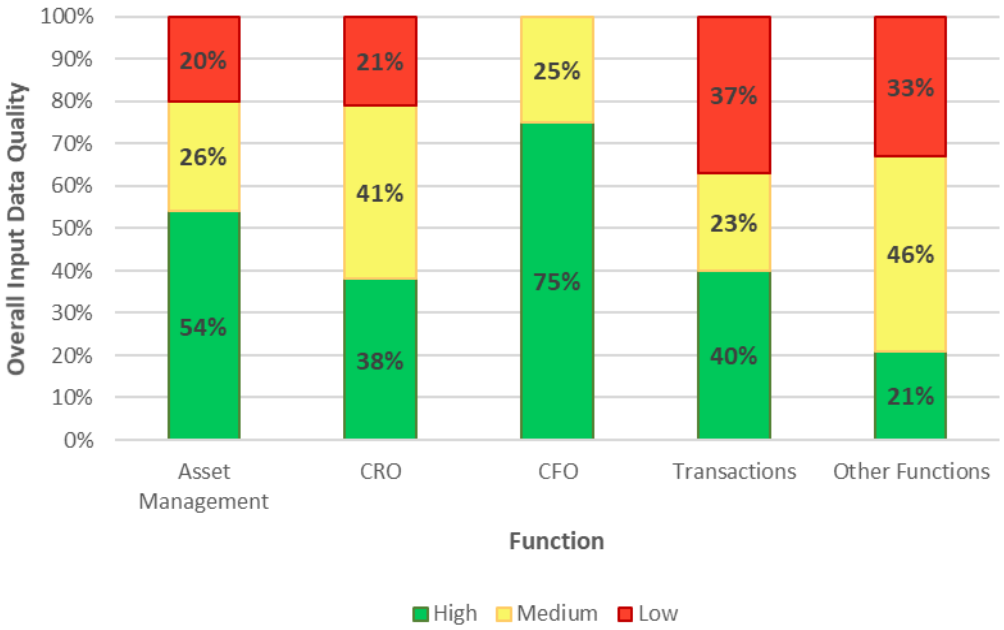


***Figure 25*: overall Input Data Quality level identified in each function**

Another interesting analysis is that which studies the level of coverage of General Principles by each of the reports produced within each function. Take for example

the Transactions structure, *Figure 26*, and evaluate for all the outputs produced, what is the satisfaction over the six dimensions of Data Quality: this allows us, for example, to evaluate those structures that resort to an excessive amount of manual controls or lacking controls and should be addressed with priority.

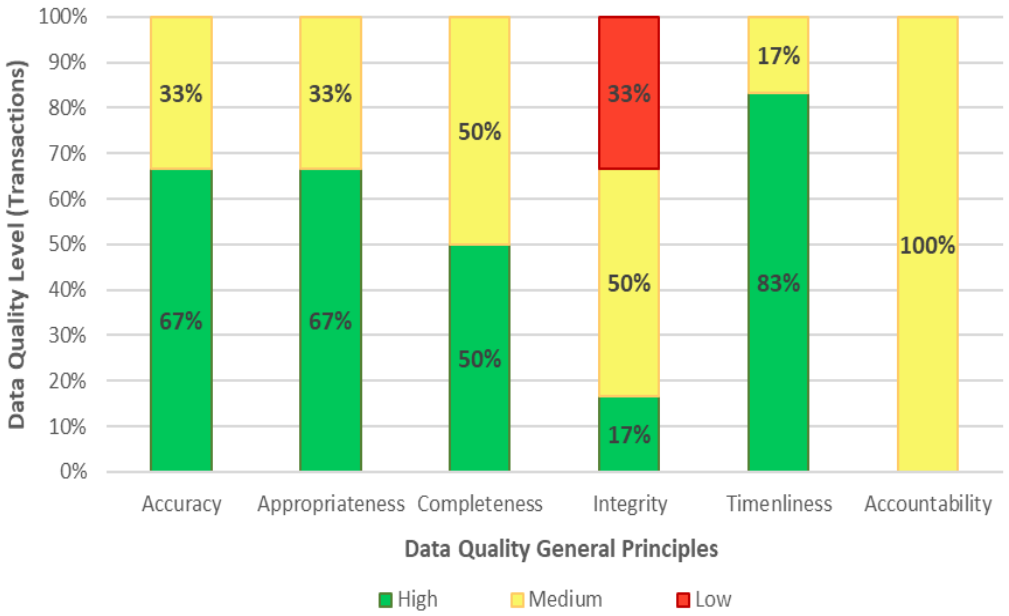


***Figure 26*:level of compliance with General Principles by Transaction function reports**

In addition to these 'more obvious' analyzes, other much more subtle and profound ones can be carried out. In fact, we remind you that a lot of other information have been collected through the template, information on which some analyzes have been carried out and many others are in progress or will be carried out in the near future. The purpose of these analyzes does not stop at the single identification of the new controls to be inserted to guarantee a good quality of the reporting production. Indeed, by entering into close contact with the organizational reality, one realizes that sometimes the quality gaps are not so much due to too few controls, but rather to deeper motivations, rooted in the business processes or in the systems and practices used to manage flows and the production of the outputs. It happens that, for example, greater coordination and homogenization in processes among structures that work closely and exchange flows continuously, can lead to an exponential growth in the quality of the output and a decrease in the effort required for the production. Or, maybe, that an improvement in data acquisition from third parties and outsourcers could improve the entire quality of the data flow downstream.

6.5 Data Directory

It has been stressed how there are Relevant Outputs on which greater attention must be paid, given their importance in terms of reputation, decision making and regulatory compliance. We have previously seen how 5 Relevant Outputs have been identified at Group level and 2 more have been added after a long skimming process. This long process is justified by a cost-benefit analysis that sees on the one hand the increase in implementation and maintenance costs and on the other a series of benefits such as decrease reputational impact, increase trust, awareness of information asset, reducing reworking and repair costs. Both such costs and benefits derive from the Policy requirements to filling in a document called Data Directory for each of the Relevant Outputs identified, that must be update at least yearly. The *Data Directory* is the main repository that centralizes technical and business information relating to data concurring to Relevant Output production and the relating Data Quality controls. The Data Directory is composed of linked sections where to collect information with different focus time by time.

Thus, it is a Data Quality documentation that gathers up-to-date and complete information regarding data in scope split into different specific sections in order to define a data, link to IT system and track the elaboration and aggregation flows. The aim is to describe Systems Information, Data Flows and Data Quality Controls referred to data (i.e. Data Instances) as soon as they are received, managed and processed within the Functional Domain producing the Relevant Output.

The Data Directory is composed of eight main sections:

1. ***Data Elements***: the Data Elements section provides information about the Business meaning of the technical information registered into the Data Instances section. They are the section contained in the Business Glossary;
2. ***System Information***: the System Information section provides information about the application systems where data are stored, including Office. The section is filled by Business Technical Stewards, coordinated by the Business Data Stewards;
3. ***Data Flows***: the Data Flows section provides information about the flows of data among the systems. It is important in order to understand where data come from, their destination and which systems process them. The section is filled by the Business Technical Steward coordinated by the Business Data Steward;

4. ***Data Instances***: the Data Instances section provides information about the overall set of data which is received, managed, elaborated or aggregated within the Functional Domain producing the Relevant Output;
5. ***Data Quality Controls***: the Data Quality Controls section provides information about the Data Quality controls (both manual, automatic or semi-automatic) in force such as, at least, control name, control rule, underlying Data Quality Principles, data under controls and ownership;
6. ***Risk & Impact Assessment***: the Risk & Impact Analysis section provides information over the inherent risk, mitigating effect of current controls in place and residual risk of the underlying data;
7. ***Data Quality Controls Results***: the Data Quality Controls Results section provides information about controls results (binary form OK-KO or other result types) related to the last run of the process, thus to the final certified/delivered data;
8. ***Issue Register***: the Issue Register section provides information about the Issues that have to be solved and their respective Remediations.

Figure 27 below shows a flow diagram of the filling in process:

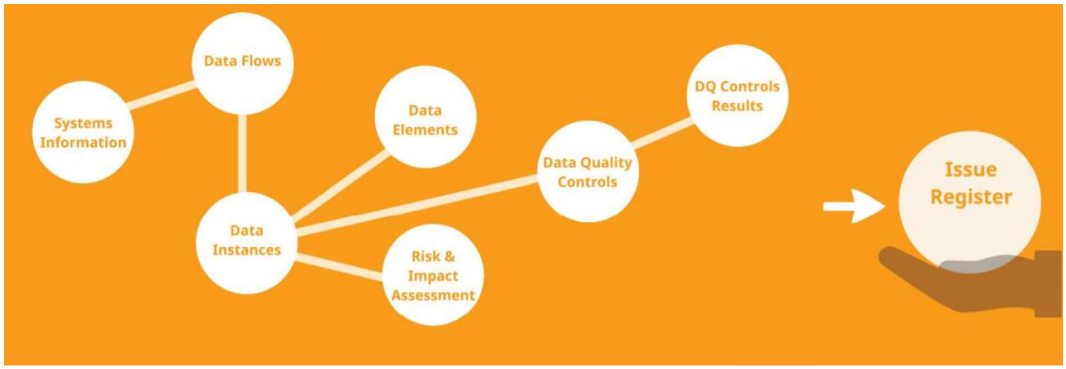


***Figure 27*: flow diagram for the filling in of Data Directory**

Finally, the Data Directory filling in will lead to the calculation of the main Key Quality Indicators (as seen in *Table 6*), that will be reported to the Board for monitoring purposes.

Chapter 7
A REAL CASE STUDY ON A BANK GROUP: TO BE CONTINUED

This last chapter will close the project, presenting the results obtained and the next steps, since the project is still ongoing. Before that, however, it will be inserted a section presenting the main activity I carried out during my internship.

7.1 Goals achieved

The leading goal of the project is to provide the Group with solid foundations for building a holistic Data Governance framework that can allow it to take the greatest possible advantage of this asset. The Real Estate company analyzed is one of the many subsidiaries of the Group in which the Policy has been extended. Clearly, within our work team, we have tried to maintain consistency, as far as possible, between the way we treat address the various subsidiaries, in order to propose similar deliverables, adopt similar procedures and methodologies and seek as much economies of scale as possible, even by leveraging on what colleagues did. This obviously means that the customer perceives our expanded team as unique and cohesive and totally dedicated to obtaining the predetermined results for the benefit of the Bank itself. However, it should be stressed that the different subsidiaries have characteristics that are sometimes very different, also due to the different business carried out, and it is essential for the success of the project to take them into account, so as to customize the activities based on the peculiarities of the specific Legal Entity taken into consideration.

Analyzing the current status, there is a greater awareness regarding the issues of Data Governance / Quality at a pervasive level, on the entire company perimeter. The awareness that a better management of data can lead to net benefits in terms of reputation and profit, begins to take root within all the corporate structures, so much so that the requests and proposals for improvements from various members of the organization are increasing. This was certainly favored by the adoption of a decentralized model that delegated the tasks to the peripheral

structures with the need to involve a huge number of actors: the consequence is the huge number of interviews carried out, with very demanding activities for the Business Data Stewards, and the final result of the expansion of knowledge about these topics companywide. We therefore proceed to an examination of the results obtained through the project activities carried out so far.

First of all, from the results of the Benchmarking it is decided to opt for a decentralized model capable of attributing responsibilities and duties to peripheral structures, as adopted for the most part in our nation and throughout Europe. A clear, scalable and flexible logical Data Governance model which assigns the roles based on those already defined by the organization charts is then provided. This allows the responsibilities on activities, data flows, reports and controls to be assigned without doubts to specific figures, so that when a Data Governance / Quality problem arises it can be addressed by the direct owner. This eliminates the risks of overlap and non-uniqueness of ownership. The people to refer to in particular problematic cases are also indicated with precision. For example, in the event that problems which can have a great impact on organizational life come to light, an escalation process is activated towards the appropriate bodies which are empowered to make a decision. In the final analysis, the System of Governance was defined.

The setting of a Data Governance Policy which clearly defines roles, rules, procedures, activities and everything needed to create a structured and well-managed framework was also important.

Moving to the local reality of the Real Estate company, a great benefit deliver to them is that of identifying the new Relevant Outputs. The long selection process and the different skimming phases allowed to identify, among all the reports and in particular among the Eligible ones, the 2 new Relevant Outputs to be added to the 5 inherited by the Group, bringing the number of local Relevant Output to 7. The activity also led to the collection of a large amount of information relating to company data flows, which allowed on the one hand to have an overview of the initial level of maturity in the company and on the other to obtain a set of essential data on to do some important analysis. These analyzes first concern the level of Data Quality recorded with regards to data flows within the company and in relations with the various outsourcer. Some of those evaluation have been reported in paragraph 6.4.2. This allows us to start tackling the priority problems by understanding how to fill the Gaps in terms of Data Quality. Secondly, the data collected allow more in-depth analyzes, through which it is possible to propose corrective actions in terms of the procedures adopted and systems used.

Linked to the theme of Gap Analysis is the definition of the Data Quality framework. An important part of Data Governance is Data Quality, and a result

achieved is that of creating solid foundations and providing the necessary tools on which to base the structuring, analysis and monitoring of the quality of the data produced from here on. In particular, in accordance with regulatory provisions, in particular Circular 285, the 6 General Principles which a report must satisfy in order to be considered of high quality have been defined. In addition to the principles, the structuring of the control levels has been defined along all data processing activities, from extraction to final aggregation.

Another important benefit is the consolidation of some Data Governance / Data Quality tools that was already present but just in a primordial state. In addiction, other tools have been introduced from scratch. First in list order is the Business Glossary, which allows you to align the Business and IT members about the Business terms used within the company. A collection of some Business Data Elements (those related to the 5 Relevant Outputs at Group level) and their porting to the Data Glossary has been already made, obtaining a first draft. The same holds true for the Dictionary of Controls, which has the purpose of containing a list of all the controls performed on the data flows and the related outline data. The mapping of the controls began through the interviews with the Business Data Owners, in which the existing controls and those that they want to implement in the future are collected. Their porting to the Dictionary will begin shortly. Although the two porting activities are in their initial state and therefore the two tools are nothing more than very simple incomplete drafts, our concern is to adequately inform all the structures of their presence, explaining their principles and operation, also considering the necessary work of the Business Data Stewards for the supply of the necessary information to be included in them.

A further deliverable is the support for the choice of the Data Quality tool. The tool will represent the new heart of the company Data Quality that will manage the controls (the Dictionary of controls will be embedded in it) and the quality of the data globally, going to rationalize the controls and eliminate manual ones as much as possible.

Part of the Data Quality framework consists of the calculation of particular KQIs indicators, defined by the individual Data Owners, to provide summary information on the controls and their outcomes and remediation. The procedure to define such indicators are also given to the company.

Ultimately the compilation of the Data Directory. The Data Directory is the main repository that centralizes technical and business information relating to data concurring to Relevant Output production and the relating Data Quality controls. The Data Directory has been compiled for the 5 Relevant Outputs identified at Group level and at the end of this year it will also be replicated for the 2 new promoted.

Therefore, both from the aspect of Data Governance and that of Data Quality, it is well evident that solid and substantial guidelines have been provided, absorbed by the corporate actors and already largely implemented with the complete satisfaction of the whole company.

The benefits brought are undoubtedly many but there is still much to be done, both to conclude the ongoing project and to be able to start others of continuation, aimed at putting into practice the results coming from the Gap Analysis. next paragraph will show the next steps to be performed in the very next future.

7.2 Next steps

As we have seen, the project is only halfway done. With our work we have already delivered the various concrete benefits presented in the previous paragraph, yet most of the remaining activities require our support. Although with the work already done we have increased the level of awareness and knowledge regarding the issues of Data Governance and Data Quality, some of the remining activities are very demanding and our help in carrying out is absolutely necessary, at least in laying the foundations. Furthermore, most of the activities fall on the Business Data Stewards hands, both for their more operational role, closer to the production and processing of data, and because of the structure and peculiarities of the decentralized model. Since the Business Data Stewards are already overburdened by their activities, and as they have already highlighted at the beginning of the project, our support also for the ongoing business is the basic element for completing the project successfully.

A first activity for which our support is required is the collection of Business Data Elements for the population of the Business Glossary. As explained during the analysis of the activities, a template and guidelines are sent to the Business Data Stewards, the figures with the responsibilities for identifying the Business Data Elements. However, several of them are experiencing problems with the compilation, so meetings with those who have asked us for support will be organized to complete the activity. The Business Elements must then be validated / confirmed by the responsible Data Owner and it will finally be possible to insert them within the Data Glossary, with all the related detailed information.

Another activity planned is a first porting of the controls on the new Data Quality tool. The Dictionary of Controls will in fact be embedded within the Data Quality platform in order to centrally manage the entire set of controls and facilitate analyzes. The activity takes a long time, since after the identification of

all the already existing Technical and Business controls, the Business Data Steward is responsible for identifying the new checks aimed at providing an overall level of Data quality able to satisfy the 6 Data Quality General Principal listed in *Table 6*. Only at the end of this they will be implemented in the digital platform. To facilitate things, as mentioned, an Excel template has been sent to all Business Data Stewards to identify the present and future controls, inserting all the information that is required in the tool in order to facilitate the transfer of them to the platform.

In addition, the Gap Analysis activities are just at an initial phase. At the end of paragraph 6.4, some results have already been presented, obtained by a first elaboration of the information collected through the template. Those results are mainly concerned with outputs and inputs Data Quality, eventually focusing on each single General Principle. This is certainly a starting point for identifying the controls necessary to improve data flows and to implement them within the Data Quality tool. However, not everything can be solved with the simple addition of a pair of controls: problems often lie deeper, in procedures, in implemented processes. Through a cross-function analysis of the information contained in the files, we are trying to understand if it is possible to propose strong improvements in terms of quality of some of the reports produced, especially for the Eligible ones. After having identified the structure which produce significant report with low Data Quality level, further investigations with the related Business Data Stewards will be made to understand the methodologies adopted for the preparation of the reports with the aim to look for any faults in the whole data lineage.

What has been done so far has provided the organization with the basis for being able to treat the data in accordance with the regulatory provisions and extract from it the greatest possible value. In the future, the company will therefore have to carry out the same activities as seen, namely the mapping of any new outputs and the probable reading of new Relevant Outputs, the update of the Business Glossary and the Dictionary of Controls, the maintenance of the Data Quality tool, the identification of Data Quality Gaps, as well as filling in the Data Directory for all Relevant outputs, etc., on their own, making leverage on what has been learned in these months. And, above all, they must update the Policy and the activities to be carried out in conjunction with normative amendments or procedural changes which could be inconsistent with what has been applied so far.

7.3 My contribution to the project

The project that have been shown is extremely vast. The number of subsidiaries in scope are several and activities to be performed are numerous. Moreover, as any project of such dimension, it is characterized by a high level of complexity, and the experience required to implement the whole framework is unavoidably high. The design of the model, the draw up of the policy, the interviews with the representatives, the analysis of the results, the drawing of the tools, the software selection and all the other activities require much experience and a deep knowledge of Data Governance topics. The experience must not only be thought in terms of hard skills and competences on Data Management, but also in term of soft skills: in a project like the one presented the ability of the advisors to deals with different clients' personalities is at the core of the success. For those reasons, the team members involved in the project ranges from interns to high seniority managers. The youngest resources were more dedicated to operational activities, while the lasts were focused on more competences demanding tasks.

Being an intern, my role has been to help in the day to day activities, and, fortunately, I have had the possibility to take part in a set of diverse activities. After a brief period needed to be introduced in the team and in Data Governance topics, I have been literally thrown in the reality of the project. At the beginning, I have been involved into two interconnected tasks, the interviews with the company members and the back office PMO activities to organize and keeping a neatly track of the outcomes coming from the meetings. Along with a colleague much experienced that me we have held about a hundred and a half meetings, either face to face or via conf. call. The people met, ranges from operational members to top management, like CFO or COO, and are representative of a wide number of different functions. It has been really interesting since gave me the possibility to learn how to deal with people belonging to opposite sides of the hierarchy. And it is clear, that the way you approach a young operative member with few months of experience or the historic CFO in charge for 25 years must be different. As different are their time availability. One major issue of this phase was the coordination of the meetings given the huge number of actors to be involved. The number of emails, messages, calls and update to be managed has become enormous, and, in some cases, the little attention some Bank' s actor dedicated to the project and to us complicated the organization of the activities. Those less available people, however, are not to blame, since, as said, Data Governance activities add to the ordinary one, requiring much effort to organizational actors. Moreover, some periods are more demanding than other: think for example at the amount of work required to CFO during the closing of Annual Financial Statements. The PMO activity thus, has proceeded in

accordance with availability of people trying to respect as much as possible their requests and commitments, keeping in mind that we also have stringent deadlines to respect. During the meeting my support has consisted in the collection of all the information coming from the interview, both the one required to filling in the Excel template and the one that may give us some insights about how to bridge Data Quality gaps. In more than one occasion, it has been left to me the possibility to hold the meeting in first person, explaining to the contact person what Data Governance is, the objectives of the project, the activities to be carried and their involvement.

Along with the meetings and the job to organized them, a necessary activity was to keep trace of the interview held in order to proceed in an orderly way and avoid forgetting some representatives, far than impossible in a project involving a such high number of people. Thus, some Excel files have been built to satisfy the need, allowing us to effectively complete all the necessary meetings.

After few months, parallelly to the activities just exposed, we start the aggregation and study of the feedbacks obtained during the interviews. In particular, I have been dedicated to the analysis of the outputs collected and the related information, like, for example, the inputs required to elaborate them. We start the process by carefully porting the contents of the templates filled out by Business Data Stewards and subordinates to an Excel file whose structure was thought to imitate the one of a database, creating keys and connection among keys of different tables. This porting activities, brought to light a set of inconsistencies among information and the lack of others. To solve such problems, it has been necessary to contact once more time the actors interested, to make clarification or integrating the missing information. Once the database has been completely populated and structured data need to be check for consistency. Thus, we implement some controls to make us sure to work on accurate and complete data. After the cleaning operation the analysis start to take place. The first analyses that have been done are related to the number of reports produced, the number of inputs, and something about the quality of them. Being performed on Excel, functions, data validation and pivot table have been used daily. As deep stressed, much other analysis must and will be done in the next future.

Talking about, Business Glossary, Dictionary of Controls and software selection activities, being more complex activities, they have been carried out by colleagues with a higher level of seniority and experience.

During all the project activities, it has been asked to us to weekly, or at maximum every two weeks, update the company project responsible (Data Governance Leader), in relations with work progresses. Those updates consist in PowerPoint slides in which the progress of the project was highlighted in terms of people involved and made aware of the Data Governance program, situations

about the output mapped, issues raised, consideration to be analyzed deeply and so on. This is another task over which I have operational responsibilities and, apart from the content, also freedom to operate.

Concluding, although the tasks carried out have not required particularly analytical skills, the activities faced have been (and still are) challenging, requiring mental flexibility, daily adaptability, and most of all, a large dose of multitasking capability.

Conclusion

The aim of the paper has been to try to fill the existing literature gap regarding the practical application of Data Governance concepts. In fact, if at the theoretical level the topic has been quite discussed, as far as the practical application of the concepts are concerned, there is still much to be done. To this end, also through the firsthand participation in a project for the implementation of a Data Governance framework for an Italian Banking Group, an attempt has been made to concretize the theoretical framework in order to meet the regulatory requirements, in particular Circular 285 of Banca D' Italia.

The first part of the document focused on the development of the theoretical aspects starting from a literature analysis. What is immediately evident is how Data Governance, as a relatively new topic, does not currently has a well-defined terminology, precise standards or proven solutions. However, I believe that this non-standardization of nomenclature and practices is not only due to the novelty of the issue, but also, and in large part, to the different intrinsic characteristics of the various organizations. In fact, there is no model applicable to all companies in a standard way, but rather it is necessary to understand what are the specific peculiarities of the entity being studied, the decision-making processes in force, the corporate culture and customize the solution in order to meet and satisfy at best his needs. This evidence also emerged from the Case Study carried out within the Banking Group.

The analysis of the literature also led to the underlining of the deep link between Financial Services and the IT world. In fact, the two areas are perfectly combined within FinTech, in such a way that they feed each other by supplying the fuel to improve one another. Surely, the future of financial services is directed towards a widespread use of technology in every service provided to the customers. This is another cue that emerged from the project, as evidenced by the number of IT systems already used by the company, but above all by the need to integrate others, as per project deliverables. But not only, in support of the strict relationship between Financial Services and Technology, there is the fact that organizational Business and IT members have necessarily had to work in symbiosis to achieve the objectives set by the project.

The project, as mentioned, has been used in order to make up for the lack of Case Studies to be utilized as a basis for a practical reflection about the topic. In addition, my first-hand participation in the performance of the activities allowed

me to observe directly how theoretical concepts are brought into the company reality. The implementation, in particular, was substantiated in the construction of a company wide Data Governance model, providing a clear definition of roles and responsibilities, a procedural framework for the processing of the data flows and the outputs produced, the tools supporting the model and a Data Quality monitoring system aimed at a continuous improvement of the quality of data handled.

The objective of the paper, in my opinion, has therefore been achieved, since, although some project activities are still in progress, the theoretical concepts have been widely applied in the organizational context, going to meet the regulatory requirements imposed by Regulatory Authorities and the specific needs of the Banking Group. Specifically, referring to the objectives set in the introductory phase of the paper, the model is extremely pervasive within the Group, thanks to the distributed model adopted which provides for the involvement of a large number of individuals. Regarding the organizational members engaged, as highlighted above, it is necessary to underline the fundamental participation in the activities by both Business and IT users who, although having significantly different objectives, must collaborate in order to achieve a solid and holistic government of data. Furthermore, the model, leveraging on national and international best practices, was created specifically for the company under analysis and adapted to the particular needs that emerged. This also allows to reach another important goal, namely that of scalability and flexibility of the model. Nowadays world characterized by hyper competition and no company can afford to be static anymore and the adaptation to external and internal changes must be maximum. The structure and procedures left to the company are such as to adapt to the constant increase in data and the complexity of the same, without the need to continually adjust the model when changes occur in the contour. But above all, through the many activities carried out and the meetings held with them, a first change of mentality has been brought to all the organizational members regarding the importance of the government and the quality of the data: it represents the real starting point for transform the organizational attitude from reactive to proactive. In addition, we note that the model provided is based on a profound Benchmarking analysis and on theoretical basis, which provide a foundation for legitimizing the model itself when a request from the Supervisory Authorities has advanced.

Finally, moving to the benefits brought, what is evident from the practical application of a project such as the one carried out, is that the advantages resulting from a well-structured governance of the data do not only concern regulatory compliance, but also a whole other set of factors. First of all, a clear definition of roles and responsibilities allows a more fluid management of all data issues so as

not to have overlaps or voids of responsibility on flows and tasks. Then, a rationalization of the reporting process that involves two orders of benefits. On the one hand, the most relevant reports are identified and subjected to an in-depth treatment, in order to obtain more reliable reports directed towards the Market and the Regulator, which, in turn, leads to strengthening stakeholder confidence and customer loyalty. On the other hand, this process leads to the identifications of the gaps in terms of quality controls along the production chain to obtain clean, clear and reliable data. This allows analysts to waste less time cleaning and checking data and to focus on the extrapolation of as much useful information as possible from data, supporting the creation of value for the customer. Finally, decision-making. Based on more reliable and high-quality data, the decision-making process is extremely improved in terms of quality and timing, becoming more solid and structured. In addition, organizations, having the security of the underlying source, can take data-driven business initiatives with the necessary confidence and outperform competitors.

However, in order to be fully achieved, the objectives set out so far require a profound change in the underlying corporate culture and in the mentality of all the people involved. If each member does not embrace the idea that the quality of the data must be pushed to the top, the whole organization will be negatively impacted, since a fault in a point of the production chain (whether we are talking about physical goods or services) damages the entire system leading to the production of poor quality outcomes and no added value for the customer. Only when the issues of Data Governance are assimilated and promoted by the entire organizational staff, the corporate culture will be positively impacted and the full potential of the Data Governance procedures adopted will emerge.

In conclusion, it should be emphasized that Data Governance is a continuous process, not linked to a short-term project such as the one shown, but rather in continuous evolution and pervasive of all the dynamics of the organization during day by day activities.

References

Accenture, 2018. *Maximizing Revenue Growth in Retail Banking*.

Accenture, 2019. *2019 Accenture Global Financial Services Consumer Study*.

Accenture, 2019. *Accenture Technology Vision 2019: Are you ready for what's next?*.

Aiken, P. & Billings, J., 2013. *Monetizing Data Management: Finding the Value in Your Organization's Most Important Asset.* Technics Publications LLC.

Alhassan, I., Sammon, D. & Daly, M., 2018. Data governance activities: a comparison between scientific and practice-oriented literature. *Mary Journal of Enterprise Information Management,* Volume Vol.31, pp. 300-316.

Anthony, R. N., 1965. *Planning and Control Systems: A Framework for Analysis.* Boston: Division of Research, Graduate School of Business Administration, Harvard University.

Armstrong, P., 2018. *Financial Technology: ESMA's Approach.* 4th Luxembourg FinTech Conference.

Arner, D. W., Buckley, R. & Barberis, J. N., 2016. The Evolution of Fintech: A New Post-Crisis Paradigm?. *SSRN Electronic Journal.*

Basel Committe on Banking Supervision, 2013. *P.E.R.D.A.R.* Bank for International Settlements .

Beyer, M. a. L. D., 2012. *The Importance of "Big Data: A Definition.* Stamford, CT: Gartner Research.

Beyer, T., Lapkin, G. & De Simoni, G., 2104. *Gartner Clarifies the Definition of Metadata.* Gartner.

Blair, M. M., 2001. *Unseen Wealth.* Brookings Institution.

Buckingham, R. A. R. A. H. F. F. L. C. J. T., 1986. *Information Systems Education: Recommendations and Implementation.* New York, NY: Cambridge University Press.

Candiotto, R., 2013. *Il Sistema Informativo Aziendale.* Torino: G. Giappichelli Editore.

Castanedo, F., 2018 . *Understanding Data Governance,* Sebastopol, CA: O'Reilly Media, Inc..

Cheong, Kuan, L. & Chang, V., 2007. The Need for Data Governance: A Case Study. *ACIS 2007 Proceedings.*

Cheong, L. & Chang, V., 2007. *The need for data governance: a case study".* Proceedings of the 18th Australasian Conference on Information Systems.

Chisti, S. & Barberis, J., 2016. *The Fintech Book.* United Kingdom: Wiley.

CMMI Institute, 2014. *Data Management Maturity (DMD) Model, Ver. 1.0.* CMMI Institute.

D'Aveni, R., 1995. *Hypercompetition : managing the dynamics of strategic maneuvering.* New York, N.Y. 10020: A Division of Simon & Schuster Inc..

David J. Teece, G. P. a. A. S., 1997. Dynamic capabilities and strategic management. *Strategic Management Journal,* Volume Vol. 18, No. 7, pp. 509-533.

De Mauro, A., Grimaldi, M. & Greco, M., 2016. *A formal definition of Big Data based on its essential features.* ResearchGate.

De Nito, E., 2008. *Organizzazione, Conoscenza e Progetti.* Economia - Ricerche. Milano: Edizioni Franco Angeli.

Deloitte Consulting S.r.l., 2019. *Tech Explosion e Digital Disruptionnei Financial Services: i Board italiani sono pronti?.*

Dijcks, J., 2013. *Oracle: Big data for the enterprise. Oracle White Paper,* Redwood Shores, CA: Oracle: Oracel Corporation.

Ernest & Young, 2019. *Global FinTech Adoption Index 2019*.

Fiocca, R., 2013. *Dalla Qualità Del Servizio alla Qualità della Relazione.* Bologna: Micro & Macro Marketing, Il Mulino.

Fisher, T., 2009. *The Data Asset - How Smart Companies Govern Their Data for Business Success.* John Wiley & Sons.

Fleckenstein, M. & Fellows, L., 2018. *Modern Data Strategy.* Springer International Publishing AG.

Francesconi, A., 2011. *I sistemi informativi nell'organizzazione d'impresa.* Giuffrè Editore, Milano.

Francis, L. & Prevosto, V. R., 2010. *Data and Disaster: The Role of Data in the Financial Crisis.* Casualty Actuarial Society Publications.

Fu, X. et al., 2011. Data governance in predictive toxicology: a review. *Journal of Cheminformatics,* Volume Vol. 3 No. 1, pp 1-16.

Gomber, P., Kauffman, R. J., Parker, C. & Weber, B. W., 2018. Special Issue: Financial Information Systems and the Fintech Revolution. *Journal of Management Information Systems,* Volume 35, no. 1, pp. 12-18.

Hagmann, J., 2012. Information governance – beyond the buzz. *Records Management Journal,* Volume 23 No 3.

Harrison, S. & Sullivan, H., 2006. *Edison in the Boardroom: How New Leading Companies Realize Value from Their Intellectual Property.* John Wiley & Sons.

Henderson, D. et al., 2017. *DAMA - DMBOK: Data Management Body of Knowledge.* 2nd ed. Basking Ridge, NJ 07920 USA: Technics Publications.

Holt, V., Ramage, M., Kear, K. & Heap, N., 2015. *The usage of best practices and procedures in the database community.* Information Systems.

ItMedia Consulting , 2018. *L'economia dei dati: tendenze di mercato e prospettive di policy,* Roma.

Kamioka, T., Luo, X. & Tapanainen, T., 2016. *An empirical investigation of data governance: the role of accountabilities.* Chiayi, 20th Pacific Asia Conference on Information Systems.

Keynes, J. M., 1920. *The economic consequences of the peace.*

Khatri, V. & Brown, C., 2010. Designing data governance. *Communications of the ACM,* Volume Vol. 53, No. 1, pp. 148-152.

Khatri, V. & Brown, C. V., 2010. Designing data governance. *Communications of the ACM,* Vol. 53(1).

KPMG, 2019. *The pulse of Fintech 2018*.

Lafley, A. & Martin, R. L., 2013. *Playing to Win: How Strategy Really Works.* Harvard.

Laney, D., March 21, 2012. *Introducing Infonomics: Valuing Information as a Corporate Asset.* Gartner Research.

Laudon, K. C. J. P. L. M., 2012. *Management Information System: Managing the Digital Firm.* Toronto: Prentice Hall.

Lin, N., 2014. *Applied Business Analytics: Integrating Business Process, Big Data, and Advanced Analytics.* Upper Saddle River, NJ: Pearson FT Press.

Loshin, D., 2008. *Master Data Management.* Morgan Kaufmann.

Marakas, G. M. & O'Brien, J. A., 2013. *Introduction to Information Systems.* 16th ed. New York, NY, 10020: McGrawn-Hill/Irwin.

Martinez, M., 2004. *Organizzazione, informazioni e tecnologie.* Bologna: Il Mulino.

Mayer-Schönberger, V. & Cukier, K., 2013. *Big Data: A Revolution That Will Transform How We Live,Work and Think.* London: John Murray.

McAfee, A. & Brynjolfsson, E., 2012. *Big Data: The Management Revolution.* Harward Business Review.

Moody, D. & Walsh, P., 1999. Measuring the Value of Information: An Asset Valuation Approach. *ECIS.*

Mosley, M., Brackett, M., Henderson, D. & Early, S., 2009. *DAMA Guide to the Data Management Body of Knowledge.* 1st ed. Basking Ridge, NJ 07920 USA: Technics Publications.

Otto, B., 2011. Data Governance. *Business & Information Systems Engineering,* Volume 3 No. 4.

Plotkin, D., 2013. *Data stewardship: An actionable guide to effective data management and data governance.* Waltham, MA, USA: Morgan Kaufmann.

PwC, 2016. *Data GovernanceSurvey Results: A European Comparison of Data Management Capabiities in Banks*.

Riggs, C., 2015. *Wells Fargo: 20 Years of internet banking.*

Ronda-Pupo, G. A. & Guerras-Martin, Á. L., 2012. *Dynamics of the evolution of the strategy concept 1962–2008: a co-word analysis.* John Wiley & Sons.

Rugiadini A, C. G. P. A. A., 1973. *I Sistemi Informativi automatizzati nelle imprese di grandi dimensioni.* Milano: Giuffrè.

Sange, M. P., 2006. *The Fifth Discipline: the art and practice of the Learning Organization.* New York City: Bantam Doubleday Dell Publishing Group Inc.

Schroeck, M. .. et al., 2012. *Analytics: The realworld use of big data.* New York, NY: IBM Institute for Business Value, Said Business School.

Shim, J. P. et al., 2002. Past, Present, and Future of Decision Support Technology. *Decision Support Systems,* 33(2), p. 111–126.

Tallon, P., Ramirez, R. & Short, J., 2013. The information artifact in IT governance: toward a theory of information governance. *Journal of Management Information Systems,* Volume Vol. 30, No. 3, pp. 141 -178.

Vercellis, C., 2009. *Business Intelligence: Data Mining and Optimization for Decision Making.* John Wiley & Sons Ltd.

Wang, R., 1998. A product perspective on total data quality management.. *Communications of the ACM,* 41(2), pp. 58–65,.

Weber, K., Otto, B. & Österle, H., 2009. One size does not fit all – a contingency approach to data governance. *Journal of Data and Information Quality,* Volume Vol. 1 No. 1, pp. pp. 1-27.

Williamson, J., 2015. *Getting a Big Data Job For Dummies.* New York City, NY: John Wiley & Sons.