



Università
Ca'Foscari
Venezia

Master's Degree
in International Management

Final Thesis

The evolution of Data Governance:
a tool for an improved and enhanced
decision-making process

Supervisor

Ch. Prof. Giovanni Vaia

Assistant supervisor

Ch. Prof. Maria Silvia Avi

Graduand

Alvise Dal Maso

Matriculation number

850071

Academic Year

2018 / 2019

Index

Introduction	5
---------------------------	---

Chapter I

Data governance: evolution of a process	9
1.1 Definition and field of investigation	9
1.2 Corporate data management: evolution of <i>data governance</i>	13
1.3 Principles and purposes of <i>data governance</i>	18
1.4 Applications and governance domains	22
1.4.1 commercial sectors	24
1.4.2 public sectors.....	27
1.5 Architecture and <i>data governance</i> factors.....	28
1.5.1 Data quality management	30
1.5.2 Data Life Cycle	34
1.5.3 Metadata architecture and data integration	39
1.5.4 Security and privacy data governance	40

Chapter II

Process of <i>data governance</i>: implementation and peculiarities	43
2.1 Building a data model: premise	43
2.2 Implement a data management program.....	46
2.2.1 Basic components for the platform: initial project.....	51
2.2.2 Design and analysis of the process.....	53
2.2.3 Synchronization and startup	55
2.2.4 Monitoring and reinforcement of the <i>data governance</i>	56
2.3 Interactive and contingent approach	57
2.4 <i>Data governance</i> e big data.....	62
2.5 <i>Data governance taxonomy: Cloud vs Non-Cloud</i>	66
2.6 Benefits, risks and criticalities of <i>data governance</i>	72
2.6.1 Data protection and security.....	73

Chapter III

Data governance: key role for corporate strategic choices	77
3.1 <i>Data governance</i> as a strategic business asset	77
3.1.1 Impact of <i>data governance</i> in the company's strategic choices.....	81
3.1.2 Value-Based data governance.....	84
3.2 Best practices.....	87
3.3 Practical case: Data Governance in Healthcare	91
3.4 Future perspective.....	99
Conclusions	103
Bibliography	105
Sitografy	111
Acknowledgment	113

Introduction

This dissertation has been developed with the purpose of analysing the topic of information management. Starting from information management, the focus is then shifted to data governance. The latter is a broad discipline that has evolved very rapidly in recent years, this is due to the increased use of information and of data within companies. Data governance has had a great impact on business, acting on the complexity and applicability of the technology that allows new structuring methodologies, choices, understandings, integrations of the information needed. These aspects are necessary to create a business, its customer service and to be in accordance with government regulations. To understand the importance of information in the current society, we can paraphrase the father of information theory, Claude Shannon, who states that information is essential to avoid uncertainty, given that our entire existence turns out to be a process of collection, analysis and understanding of information in order to make the most appropriate decisions.

Furthermore, data management applies to companies that modulate their work, both in terms of product and service; companies must continually execute a very articulated process of gathering and analysing information in order to trace their business path. Today's market is increasingly competitive and dynamic, and all kind of sectors are investing to create new business processes and new technical capabilities. These new technical skills will help companies improving their selection, analysis and understanding process of information for their different areas of interest (customers, products, partners, inventories, services, prices). Moreover, companies are becoming aware that the flow of information to be analysed is constantly growing, which makes any type of business constantly changing, therefore, the ability to manage information in an accurate and appropriate way will lead to obtaining a substantial competitive advantage.

Since the beginning of the digital age, the flow of information and consequently the amount of data, has accelerated exponentially. On a positive side, companies have access to billions of information that can prompt a growth, on a negative side, a large amount of information can destroy a company if the data obtained is not organised accurately. This is caused by information which can contain structured data, as well as semi-structured data, or completely unstructured data; it is therefore the companies'

responsibility to try and collect and organise data so that this becomes a strategic asset. The basic principle of data management resides in digitalisation and globalisation of data which have rendered business boundaries, for any business, null, raising the level of competition disproportionately. In the current digital age, a company can become competitive only when it has access to more relevant, more accurate, timely and complete data, regarding its business, its performance metrics, its customers, its partners, its product and its own market. This can only be achieved through the pursuit and implementation of data governance strategies, which help sorting data in a way that will make the decision-making process of the company increasingly simpler and effective.

This paper is outlined with the purpose of analysing ways to implement data governance strategies, aiming to facilitate the decision-making process. Initially the evolution of this discipline over the years is described, supported by a definition that also includes the mother discipline of *Data Governance* or *Enterprise Information Management*. Afterwards, the evolutionary analysis is applied to the corporate level; the reasons this discipline has been implemented by companies is described as well as the main motivations for its uses. The main aims of this discipline and the reasons that drive companies to develop data governance processes are then classified. Mostly, this occurs to protect individual and collective rights, to ensure that the trade-off influenced by data management becomes transparent, responsible and inclusive; finding the best practices, learning from their mistakes, and finally improving existing governance. The paper will then analyse the different domains attributable to the data governance together with its variety of applications; to expand, it will analyse the use of different domains in the direct sector and in the public sector. Furthermore, this thesis, will outline the different architectures and factors that compose the data governance processes. Additionally, data itself is analysed, accurately describing the data quality management, the life cycle of data within a company, and finally the aims necessary to protect the data.

The second half of this paper focuses on the process of developing a governance model; firstly by outlining the premises, then by indicating the necessary requirements, and finally by defining the operational components of the process. Consequently, the components, the design and analysis of the process, and the ways of monitoring the

governance strategy are determined. A short description is then presented of the governance processes linked to big data, with the related frameworks to be used. The main benefits, risks and criticality deriving from the implementation of these strategies are cited and expanded upon why they are functional and useful. Once more, the main methods for guaranteeing an effective governance process are specified, thus how they create a secure model, which includes data, both critical and not, in a safe and effective manner.

In conclusion, the final section of the paper specifies how the use of governance strategies related to data can become a strategic asset for companies that implement it. To thoroughly delineate the use of governance strategies, the metrics used by companies are explained, hence the best practices that have been successful for the businesses that apply them. Furthermore, a practical example is presented to show how the implementation of data governance strategies in the health sector of the American middle east works. This is achieved by focusing on the structures offering ophthalmic services, and noting how the inclusion of data governance processes has been useful to this specific structure and to companies who implement these strategies. Finally, data governance's future prospects are identified and its possible applications are outlined.

Chapter 1 – Data governance: evolution of a process

1.1 Definition and field of investigation.

Data Governance in general is only a part of a larger discipline, which is precisely the *Enterprise Information Management* (EIM).

But to analyse the term *Data governance*, we have to introduce the key concepts that we have to understand before the analysis.

These key concepts are:

- *Data Information management*
- *Enterprise information management*
- *Data information architecture*

Data (information) Management

Therefore, the first concept (DM) we can say it is the business function that plans and implements policies, practices and projects that obtain control, protect, deliver and boost the value of the *data information*. Furthermore, this is way to increase the performance in function of *data management*.

To better understand the overall definition, I have to explain some other terms. These are:

- *Business function*
- *Program-Data/information*
- *Discipline*

Enterprise Information Management

The EIM is the process by which the companies increase their value by managing the information resources to support their operations and to improve their performance. As previously mentioned, the enterprise information management handles the policies' principles, plans, frameworks, technologies organizations, people and processes in a company by leveraging in the investment in data and content.

Data Architecture

This is the third term that it is often used when we talk about the *data governance*.

It's the most technical definition between the key concepts mentioned above.

We can say that DA is the *picture* of the information management context, which includes the elements and the interactions.

Expanding the definition, the DA can include:

1. Models of how the different components used to manage the information suit each other's.
2. A list of *standards* that explain the permitted formats presentations and the use of data.
3. organization managing data description.

Data governance and data management

It is necessary to point out that *data governance* could be confused with the term *data management*. To make a distinction, the DAMA (Data Management Association) defines the data management: “the development, execution and supervision of plans, policies, programs and practices that control, protect, deliver and enhance the value of data and information assets”.

On the other hand the DGI (Data Governance Institute) defines the data governance as : “a system of decision rights and accountabilities for information-related processes executed according to agreed-upon models which describe who can take what actions with what information, when and under what circumstances, using what methods”

We can therefore assume that *data governance* and *data management* are complementary but the first one cannot replace the second, and the other way around. The importance of the *data governance* is the help that it can bring within an organization to find its mission, increase the clarity, achieve confidence in managing the different kind of data collected, clarify the organization accountabilities, pursue scope and focus that bring the organization to success.

Basically, we can assume that an effective *data governance* framework could bring lots of benefits to the organization; examples of these benefits can be: a better decision-making process, a reduction to the operational inefficiencies, ensuring the

needs of data to the stakeholders. Moreover, among the benefits that *Data Governance* brings, we have to mention the training of the organization's staff and management who will adopt common approaches, build replicable processes, have the same standards, and try to increase the efficiency reducing the organization's costs by perfecting the coordination between different units, and clarifying the transparency of all processes.

Organizations, in order to create effective data governance programs, need a strategy framework, which is a way to better understand the needs of the companies, but, mostly, the need for information. This framework, as previously mentioned, can help the organization towards their mission, to obtain clarity, to build up the confidence in using data, to increase the accountabilities and sustain their focus. To accomplish the above goals, companies must develop a scheme, a model, which will define the different responsibilities within the organization, in order to clarify each individual what their role is, and what information or data they have to manage. The top data governance experts assume that when the organization does not develop *data governance*, chaos reigns, and employees are faced with data considered "dirty", redundant and inconsistent; in addition they will not be able to integrate the data, the performance will not be the best, accountability and availability will be poor. So, there may be the feeling of loss of control by the organization.

Furthermore, there are barriers that organizations must overcome in order to develop an effective strategy in data governance. These barriers can arise from technological, organizational, legal& financial policy problems. Moreover, the poor predisposition to show and communicate the value of the data governance is considered a barrier, and the most common cause of failure in this kind of programs within organizations, is the lack of communication between the employees and the stakeholders.

So, companies, to successfully develop data governance frameworks must create a chart for the data governance program that specifies the company's mission and vision.

Data governance and Information governance

For the purpose to be clearer it is necessary to make another distinction, between the *data governance* and the *information governance*.

In recent years, companies have started to understand the importance of data to enhance their personal performance, to increase the overall value of the company, and to improve competitiveness, but most importantly to decrease costs.

Recent Big Data studies confirmed that the collection and development of information creates a significant and competitive advantage (M.schroeck, et al 2012). Moreover, the IAIDQ, which is the International association for information and data quality, (E. Pierce, et al 2012), assumes that in the largest part of companies studied, approximately 70%, data has become a strategic resource.

It is therefore clear that the ability to obtain value from data is strongly related with the performance of the company.

As I previously mentioned, *Data Governance* is the ability for an organization to obtain and use in the right way the data collected.

On the other hand we have the *information governance* which is the collection of processes that establish normative standards that can help the organization in the decision making activities, which means that the discipline consists of interactions, and it manages the different data asset of a company through all over its lifecycle.

It is universally agreed that an excellent data governance structure within companies has become necessary if said companies want to be competitive and recognized in the actual market.

Nowadays, digital transformation is deeply impacting all kind of businesses, so their data, information and their ability to use it is becoming fundamental to reach, retain and interact with clients.

To accomplish that, companies will have to adapt to new kind of technologies in order to obtain a positive economic value and a critical boost in the recent years of digital transformation. To do that companies must find where data is located; the composition of different data, so to understand the driver of this data; they need to know who can influence this data and finally how it is generated.

To summarize a data governance definition, this is not only a storage of data, it is not a database warehouse or a simple administration of both; and it is not a collection, or transformation, or reading of data.

The final goal of a data governance program within a company is to find and explain the data strategies, with all their policies, standards, architecture procedures and metrics.

(AA.VV enterprise information management); the other main goal is to resolve the various problems related to data; and finally to elevate the value of data asset.

Most of businesses today are overwhelmed with information and data, and the volume increases constantly.

The main objectives of the *data governance* are two; the DM planning and the DM supervision & control.

- 1) *Data Management planning*: in this process, the main activities are identifying the primal data needs of the company; creating an appropriate data strategy; developing the internal organization; formulating policies, standards and procedures; developing a data architecture; trying to calculate the value of the data process, in order to create of budget.
- 2) *Data management supervision & control*: this process manages the main activities of control of the process. These include the coordination of the activities; the solution to the main problems; the monitoring of policies, standards and structures previously determined; and finally, it communicates the value obtained by the Data asset.

1.2 Corporate Data Management: evolution of *data governance*

To better introduce the topic, it is helpful to look at the history of the evolution of data management.

The first certified test on the use of data in a strategic manner dates back to 1966, in the aeronautical engineering sector.

This data was used in procedural documents, for research and development of activities within the companies.

Subsequently, between 1979 and 1980, the use of data and data management per se has increased exponentially, mainly in the engineering and scientific sectors. The principal use was to help interface with complex projects. To cite one of these projects, Jayroe (1973) describes step by step the whole data management process in NASA's space mission.

Another striking example is the one cited by Engel & Shaw (1976) "problems of data management in a basic line study of the outer continental shelf" where the main problems deriving from data control are described and specifies the implementation of processes that help data control.

During this study, the different members of the research teams and their roles are defined, including the project managers, those who collect the data, and those who process it.

We can therefore affirm that initially data management was used to solve the problems in the acquisition of the data, in its processing, and in its storage; but this was done for short periods, rather than the long term.

Data management was initially used to solve immediate needs and was driven by the ability of scientists to do what was needed in their projects.

For 30 years, up until 2000, data management was used in this limited way, not in all fields, but in highly complex projects and for initial data collection and processing studies. However, it was revealing the importance of knowing that data and processing it in the right way, would bring a real competitive advantage to companies. At this moment we can start referring to "research data lifecycle", a term until now unknown in the literature of the sector.

Since 2000 there has been an enormous development of data management plans, mainly derived from public policy, these are eResearch and economic policies.

eResearch, was introduced at the beginning of the 2000s when there was a huge development of internet traffic and consequently data production. As Hey and Trefethen (2003) and Emmott (2006) wrote, this transformation was profoundly changing research too, which very quickly became digital.

One of the first and most complete reports is the one written by Lord and McDonald (2003) which deals with the potential impact of this transition from analog to digital. This report complains about the complete absence of institutional and government funds for data storage, and it was assumed that the lack of investments in this area would have easily led to the abandonment or loss of the use of these data. To try to overcome this potential problem, Lord and McDonald (2003) deployed data management strategies, which they tried to share, to preserve data management policies, mainly aimed at institutions and governments, instead of the individual researchers.

Now shifting to the economic factors of this change towards digital research, in 2004 a group of researchers from the organization for cooperation and economic development (OECD), published a "declaration on access to research data from public funding", which recognized the positive impact of open access to data, committing to work to ensure that agreements are created to have public research funds.

This declaration meant that from 2004 a group of experts was formed in order to draw up recommendations approved by the OECD. This statement made it clear that the nations participating in the OECD had few returns on public funds for research caused mainly by the absence of data re-use. The report therefore underlines the economic importance of data sharing but does not consider the impact that can have for researcher's tasks.

To better understand what the future evolutions of this discipline might be, we need to analyse its origins more deeply, then we will talk about corporate and information technology governance.

Corporate governance is: "the set of processes, customs, policies, laws, and institutions affecting the way a corporation is directed, administered or controlled".¹

Among the practices that are part of corporate governance, we find IT governance, which is defined as: "the leadership and organizational structures and processes that ensure that the organization's IT sustains and extends the organization's strategies and objectives."² We can therefore say that IT governance is a highly developed discipline; many organizations, among which we find the ISO (International Organization for Standardization) have clearly defined processes and standards that the companies that want to submit to these certifications must implement.

Many of these IT governance practices have subsequently evolved based on common architectures and structures, which have resulted in applications such as ERP (Enterprise Resource Planning) or CRM (Customer Relationship Management).

¹R. A. G. Monks and N. Minow, Corporate Governance, Chichester, England: John Wiley and Sons Ltd, 2008.

²Board Briefing on IT Governance, 2nd ed., IT Governance Institute, Rolling Meadows, IL, 2003.

However, IT Governance practices are mainly developed in applications for the organization.³ In most companies the owners, processes, and policies are clearly specified in order to better manage the organization.

Despite having this type of framework, companies lack a governance structure that manages the most valid data: the one concerning customers, employees and the product. It is therefore understood that data is an indispensable asset for companies today even if they are not included in the budget. Representing, as mentioned above, customers, employees and suppliers, data also represents the main activities, transactions and results. Data if used and managed its best, remains the finest reusable asset within a company. It can help managers make short-term decisions, they can provide information regarding the profitability of their product or customer portfolio; they can also provide useful information to the various business units improve the planting of their work. It is therefore clear that the data relating to a company, in order to be effective, should be shared within the company itself, between the various sectors and processes; to do so, and to get the most out of the data, this must be managed as a real asset. To obtain maximum value from data, in an economic perspective, companies must develop and establish common standards, common policies and processes that deal with data management, so to be able to build the right corporate structure and the related technological support.

The motivations that drive a company to develop data governance plans are 3:

- Ensure data meets the needs of the business.
- Protect, manage and develop data as a valued enterprise asset.
- Lower the costs of managing data.

Nowadays an increasingly large number of companies rely on data management programs and invest more funds in such initiatives.

A study carried out by the TDWI, The Data Warehousing Institute, shows just how the number of companies adopting data governance programs is growing. To mention a few numbers, 8% of the 750 companies interviewed have already developed data governance initiatives; 17% are developing one, and 33% are planning to do so.⁴

³ G. Reese, *Cloud Application Architectures: Building Applications and Infrastructure in the Cloud (Theory in Practice)*, Sebastopol, CA: O'Reilly Media, Inc., 2009.

⁴Panian, Z. (2010). Some practical experiences in data governance, World Academy of Science, Engineering and Technology.

But what are the main motivations that drive companies to develop and adopt programs of this type?

The main ones, as Panian Z. (2010) mentions, are:

- Growing revenues
- Lowering costs
- Ensuring compliance

Growing revenues

The main objective for any company is to increase profits; one of the most effective ways a company has, is to increase cross-sell / upsell ratios and increase customer loyalty. This is possible if the company increases the knowledge of its customers

Lowering costs

Another fundamental aspect that characterizes a company's success is the reduction of costs; for this purpose, companies must increase and guarantee operational efficiency. the main way to do this is to automate or standardize internal business processes.⁵

Ensuring compliance

Data governance is an optimal way to achieve compliance⁶. Ensuring compliance succeeds in establishing the standards, policies and processes that are compulsorily required by government policies, it helps to standardize compliance-related processes, thus leading to cost optimization. Moreover, it provides a fundamental help in the responsibility and verifiability of data.

⁵J. Jeston, Business Process Management: Practical Guidelines to Successful Implementations, 2nd ed., Burlington, MA: Butterworth- Heinemann, 2008.

⁶ J. Bowerman, (2007, 02). Effectively Auditing and Reporting on Database Activity for Compliance". DM Review. Available: www.dmreview.com.

1.3 Principles and purposes of *data governance*

With the objective to further develop the topic of *Data Governance*, it is necessary to understand the main principles and goals that data management practices can bring to a company.

Beginning with the principles, they must not be considered as rules, they must not limit the actions undertaken by the employees, but can be used as guidelines and instructions that help the various business units. These principles should be applied at the base of the various governance models for the different units, eventually providing an overview of the data governance framework. Moreover, these principles provide the possibility of designing governance solutions suitable for different sectors and businesses, creating customized models.

There are four principles, all four are contained in a macro principle called *promote human flourishing*. These principles are:

- *protect individual and collective rights and interests*
- *ensure that trade-offs affected by data management and data use are made transparently, accountably and inclusively*
- *seek out good practices and learn from success and failure*
- *enhance existing democratic governance*⁷

Analysing the guiding principle, the promotion of the flourishing human should be the basis of all governance systems. The main purpose of this principle is to ensure that, in the development phase of the system, “human” is at the center of the aforementioned system. However, the meaning of human flourishing is really complex, and we can say that there is not a single definition that adapts to each context, but this can vary according to the applied field.

Protect individual and collective rights and interests.

Often, the incorrect use and management of data, afflicts the rights, benefits, interests of both the individual and the collective.

The damages caused by bad collection, sharing and processing of data, can be both material and immaterial. Data can help individuals assert the rights that they previously

⁷AA.VV. (2017). *Data management and use: governance in the 21st century*, The Royal Society, British Academy.

failed to enforce, or that they struggled to enforce effectively. On the other hand, if they are not handled with caution, data processes can become very risky for goods and benefits of the communities. In fact, effective governance models pay much attention to how human rights can be damaged by data management. If personal information is incorrectly disclosed, people's privacy will be violated.

Analysing the tangible damages these can refer to health, personal finances or discriminatory treatments. Furthermore, there can be intangible damages which may refer to the exclusion of services, facilities or fears that personal data may be used in the wrong way.

To make them effective, data governance models must protect individuals and communities from all possible damages; in case of failure and damage verification, efficient systems should be able to guarantee a compensation to those who have suffered the damage. An effective model, in conclusion, requires high monitoring of the processes and results of the system in which damage can occur.⁸

Ensure that trade-offs affected by data management and data use are made transparently, accountably and inclusively

Data management requires a balance between a series of factors, among which we find benefits, risks and interests. If the governance system is to be transparent it should therefore allow anyone interested to be able to effectively take part in it. But transparency alone is not enough, and neither is responsibility. Responsibility is a more complex concept, but it can be supposed by transparency. In addition, data management systems must ensure multi-stakeholder participation; this implies that who is interested in participating is clearly established how they will be enticed to participate.

Seek out good practices and learn from success and failure

In order to have a constant improvement of practices and standards, an effective governance model should show its commitment to adopt correct practices and to continually learn. To promote and disseminate even more of the best practices,

⁸AA.VV. (2017). *Data management and use: governance in the 21st century*, The Royal Society, British Academy.

information should be shared actively. Although, good practices may not always be received positively, and can lead to failure. They should still be accepted and not avoided in order to learn from failure and use it to improve in the future. The best practices must therefore be developed, maintained and adapted constantly. Furthermore, since there is a continuous change practices must be regularly reported to avoid said practices to fail in the future.

Enhance existing democratic governance

In order to have an efficient and effective data management system, processes, decisions and controls must receive appropriate support proportionality in the management of this data must be guaranteed. Proportionality is the balance between interests and objectives, a set of means and goals and finally a real commitment to be consistent.⁹

Main data governance functions

Analysing the data governance's functions, we can identify three main macro categories:

- *Anticipate, monitor and evaluate*
- *Build practices and set standards*
- *Clarify, enforce and remedy*

Anticipate, monitor and evaluate

These three actions are necessary for an efficient governance model in order to be able to understand if the current methodologies are right or if adjustments are needed to operate better in the future. In fact, a good governance model cannot operate in a reactive manner as it must try to foresee problematic situations and avoid them. One of the most important activities therefore appears to be the scanning of the horizon, as it can act as an early warning for potentially harmful situations. this can therefore lead to shifting the focus to riskier activities making them safe.

⁹AA.VV. (2017). *Data management and use: governance in the 21st century*, The Royal Society, British Academy.

Build practices and set standards

Standards help to spread common practices in the different sectors, and ensure that everyone carries out their task, thus reducing the risk of failure to a minimum. Therefore, for a good governance model, standard designing is fundamental for the coordination of activities between the different sectors.

Recently, the standards and mechanisms for certifying processes or products are an increasingly widespread method of governance but obviously these mechanisms also have limitations and there is no one best way.

Clarify, enforce and remedy

Another fundamental purpose of an efficient governance model is enforcement; this can be practical or technical. Enforcement, clarification and certain standards are not effective for the organization if it is not able to backup these. This function is probably one of the least developed in terms of data governance, but it is also one of the most important.

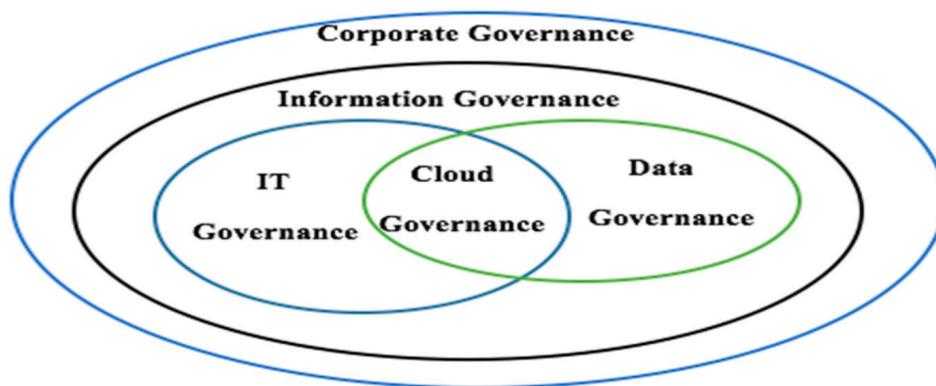
However, data governance has other success factors than those mentioned above. One of these is an organizational factor and it is one of the fundamental components of data governance. To achieve its goals, an efficient data governance structure must have ownership and standards and policies must be established. There must also be a clear definition of roles and responsibilities; an involvement from business and IT, so that both the executive and the employees are involved in this kind of practices. Furthermore, there must be a sponsorship by the executive; winning data governance programs are recognized by senior partners as valid and prolific and then sponsored within the company; for this reason, many companies establish recommendations, councils and boards specifically for data governance functions. Finally, we have the integration competency center (ICC), which is a business function able to increase agility and reduce the implementation of business costs. The second kind of success factor is technology. Technology helps to achieve data governance goals by automating data integration processes. There are ten main factors

that allow the automation of processes through technology and these are: access, discover, cleanse, integrated, deliver, develop, manage, audit, monitor, and report.¹⁰

1.4 Applications and governance domains

Considering the constant growth of governance within organizations, the different types of governance in the different corporate environments can become confusing. To cite the most important typology of governance we find: Corporate governance, IT governance, Information governance, Cloud computing cover and finally Data governance. The figure below will help to understand the link and connection between the different governance of data.

Fig.1 relationships between domains



Source: Hameed, K., Al Ruithe, M., Benkhelifa, E., (2018). Data governance taxonomy: cloud vs non cloud, Sustainability, 10, 95.

The macro-group, the one that contains all the other disciplines, is called *Corporate Governance*, and it makes the business operate adequately. Moreover, it keeps corporate the companies accountable for their work.

To quote the OECD, the Organization for Economic Cooperation and Development, corporate governance is defined as: “a set of relationships between a company’s management, its board, its shareholders, and other stakeholders, corporate governance also provides the structure through which the objectives of the company are

¹⁰ AA.VV. (2017). *Data management and use: governance in the 21st century*, The Royal Society, British Academy.

set, and the means of attaining the objectives and monitoring performance are determined”.¹¹

IT Governance discipline has assumed considerable importance within companies, a bit like the just mentioned *Corporate Governance*. *IT governance* turns out to be the way companies establish authority, responsibilities and communication through the use and design of policies, standards and control and measurement mechanisms. The definition of IT governance is therefore: “procedures and policies established in order to assure that the IT system of an organization sustains its goals and strategies”.¹²

Firstly, we analyse information governance. This discipline has not yet been developed and adopted by many companies as there is not a profound difference with IT governance. Information governance can be understood as a part belonging to *Corporate Governance*, whose primary purpose is to improve and accelerate the flow of information in the company, in order to develop the decision-making process, saving costs and risks. To provide a definition of information governance, Gartner says: “The specification of decision rights and an accountability framework to ensure appropriate behaviour in the valuation, creation, storage, use, archiving and deletion of information”.¹³

Focusing next on *Data Governance*, this is the driver that unites all the other kinds of governance and it is at the center of all the different domains. The stronger link that *Data Governance* has is with *Information Governance*, since information governance cannot work without data governance, thus becoming a prerequisite.

To understand the final hierarchy, we can use the following image that shows us how *Data Governance* is fundamental for the functioning of all other domains. From a managerial perspective *Corporate Governance* is more important than all the others;

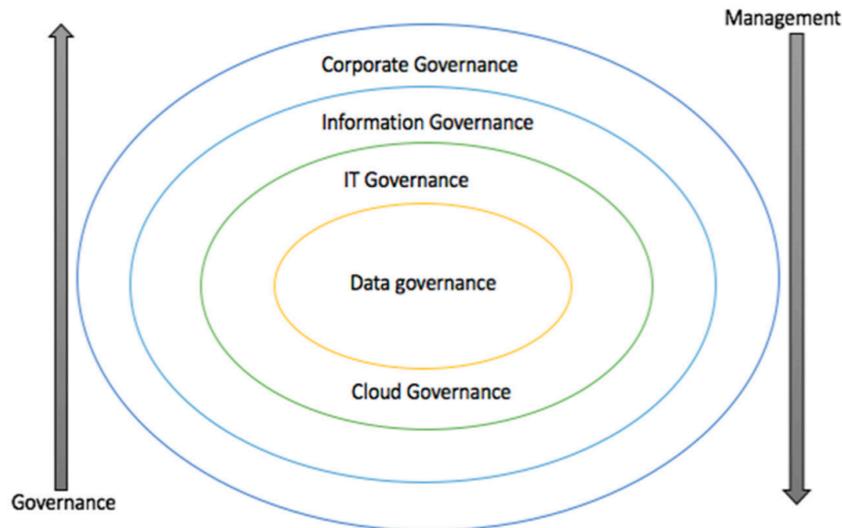
¹¹ Government, A. The National Cloud Computing Strategy. *Natl. Broadband Netw.* 2013, 2013, 36.

¹²Herbst, N.R.; Kounev, S.; Reussner, R. Elasticity in Cloud Computing: What It Is, and What It Is Not. In Proceedings of the 10th International Conference on Autonomic Computing, San Jose, CA, USA, 26–28 June 2013.

¹³ Gartner, Information Governance. 2016. Available online: www.gartner.com.

from a governance point of view it is the other way around: without *Data Governance*, the other domains fail to function effectively.¹⁴

Fig.2 Hierarchy between domains



Source: Hameed, K., Al Ruithe, M., Benkhelifa, E., (2018). Data governance taxonomy: cloud vs non cloud, Sustainability, 10, 95.

1.4.1 Commercial sector

For a practical analysis of how data management and data governance are used, it is necessary to look at its commercial applications, both in the private and public sectors.

Starting from the private sphere, the leading sectors that started to use *Data Management* programs were finance, banking and insurance. They immediately found the need to manage millions of data regarding its customers and future prospects for its own sector. To mention some Morgan Stanley, Google, eBay or Twitter that have built their businesses on data management.¹⁵

¹⁴ Hameed, K., Al Ruithe, M., Benkhelifa, E., (2018). Data governance taxonomy: cloud vs non cloud, Sustainability, 10, 95.

¹⁵ Groenfeldt, T. (2012). Morgan Stanley Takes On Big Data With Hadoop. Forbes. Retrieved from www.forbes.com.

Among private sectors that use such programs we can cite for example the telecommunications industry, the healthcare system, the pharmaceutical industry, the transport and logistics industry and retail services.¹⁶

It is interesting, however, to see how even small and medium-sized companies try to adopt *data management* and *governance* programs in order to improve their businesses. Major companies, belonging to the aforementioned sectors, have a better predisposition to deal with *data management*, because they have a large number of resources. This is why large companies are able to set up special teams whose sole purpose is to interface with the collection and processing of data to operate a better business' system.

In addition, large companies have pre-set frameworks available to them which, based on their organizational structure, culture, politics and policies, are adapted to improve the company itself.

The business drivers that characterize large companies are more or less the same as small-medium enterprises; they, too, need to make decisions every day, to increase revenue, optimize work and even try to reduce costs. A good data management and governance system, specifically designed for those who do not have the huge amount of resources of large companies, can therefore help SMEs to pursue their own objectives, while ensuring security, privacy, quality, usability and credibility in data management¹⁷

Why are we talking about SMEs? There are many definitions referable to SMEs, but I will use the one provided by the European commission which states that small and medium-sized enterprises are all those that work with less than 50 employees, have a turnover or an annual balance sheet of ten million euros or less.¹⁸

Furthermore, it is necessary to take small and medium-sized companies into consideration as they are the ones that most strongly characterize the overall market. At a European level, SMEs are 98.7% of the entire market and employ 49.2% of workers.¹⁹

¹⁶ Berson, A., Dubov, L. (2011). *Master data management and data governance*, McGraw Hill Education, New York.

¹⁷ Nwabude, C., Begg, C., McRobbie, G. (2014). *Data governance in small business*, International Conference on Business, Management and Governance, IPEDR, 82.

¹⁸ E. U. Commission, "Commission recommendation of 6 may 2003 concerning the definition of micro, small and medium-sized enterprises," Official Journal of the European Union, L, vol. 124.

¹⁹ G. Dimitri, M. Patrice, G. Edward, C. Cecilia, F. Michael, B. Gabriela, et al., "Annual Reports on European SMEs 2012/2013," 2013.

Unlike large companies, SMEs have less resources, less time to use, less technology and less opportunities to consult experts for research and development. Moreover, they have an internal hierarchical structure totally different from the big businesses.

Given that in recent years there has been a substantial increase in data and information related to businesses, companies have started to rely more and more on them. As mentioned above, however, with the increase of data and its use, the possibility of risk can rise as well as a lack of security and privacy related to the collection of data.²⁰

The main goals of data governance is to match data and its use, with the needs of the company, solving the problems related to it, trying to reduce data management costs and succeed in packaging the data in a way that becomes a valid asset for the company, providing it a substantial competitive advantage.

A data governance framework suitable for SME needs to have specific characteristics; these are:

- *Clear implementation*: implementation needs to be clear and easy, if it is not it will not be used by the company. The efforts made by the company, given the scarce resources, should be minimal.
- *Inexpensive*: having limited resources, SME's management costs, maintenance and training costs should remain relatively low. The lower the cost, the greater the probability that the system will be implemented by the company.
- *Take into account the internal structure*: the system has to take into account a hierarchical structure totally different from large businesses and realize that decisions are taken by fewer people inside, instead of a board like in large companies. Decisions in SMEs are taken quickly, and usually by one or two people.
- *Have few roles*: with maximum fifty employees in the company, the decision-making process is carried out by few people, so even the data governance framework should be designed to be used by fewer individuals.

²⁰ Nwabude, C., Begg, C., McRobbie, G. (2014). *Data governance in small business*, International Conference on Business, Management and Governance, IPEDR, 82.

- *Require no IT staff*: not having the technical skills and the adequate competencies, if data governance programs won't be used by SME's, it should be put into practice by people who are not experts.²¹

1.4.2 Public sector

Following what the private sector has been doing for a long time, the public sector is realizing the usefulness of *data management* and *data governance* programs. The commercial sector and even government organizations are faced with consumers and businesses (G2C and G2B), therefore they try to optimize their actions, increasing efficiency, trying to reduce costs and improving the services offered.

While recognizing their effectiveness, governmental organizations do not seem to be sure about the use of data governance programs; this is mainly due to the belief that they are inadequately prepared to use specific programs.

However, with the advanced technologies available today, it is relatively easy to deal with this kind of data; moreover, thanks to the improvement of technology, tools to manage data are becoming more accessible and usable for everyone.

While the private sector is moving forward, the public sector remains behind.²² Big data, however, could seriously help the public sector; for example, it could improve the supply of information, improve developed policymaking processes, improve the provision of services to citizens shaped by their needs.²³ Moreover, it could drastically improve social issues that afflict cities and citizens like the problems related to public transport, healthcare inefficiencies or the production of clean energy.²⁴

Despite these potentials, the governmental organization's approach seems stalled, or rather being evaluated. As stated by Klievink, governments do not yet act on the matter since they are not sure they are ready to face this introduction which would lead to substantial changes.²⁵

²¹ W. H. Delone, "Firm Size and the Characteristics of Computer Use," MIS Quarterly, vol. 5.

²² Mullich, J. (2013). Closing the big data gap in public sector. Bloomberg Businessweek.

²³ Chen, Y.-C., & Hsieh, T.-C. (2014). Big data for digital government.

International Journal of Public Administration in the Digital Age.

²⁴ Scholl, H. J., & Scholl, M. C. (2014). Smart governance: A roadmap for research and practice. In iConference 2014 Proceedings.

²⁵ Klievink, B. et al. (2017). *Big data in public sector: uncertainties and readiness*, Information System Frontiers, 19.

1.5 Data governance Architecture and Factors

To deeper analyse into what the architecture of data governance is, it is necessary to go back to its definition.

The data governance is a process and quality control discipline focused on managing the quality, consistency, usability, security, and availability of information. This process is complex, multifaceted, and closely linked to the notions of data ownership and stewardship, data quality management, metadata management, and the definition of effective organizational structures to support data governance efficiently and in a n agile fashion.²⁶

By the definition, *data governance* becomes a pivotal point for companies that want to implement data management strategies.

Data management and *governance* provide substantial help to companies in making decisions regarding data usage, obtaining real value from these data, trying to minimize the costs and complexity of operations, managing the associated risk to data, and finally to ensure compliance with various legal restrictions.²⁷ Moreover, it provides a concrete tool to companies allowing them to design procedures, processes and rules in the use of data, applying a discreet form of control, companies make sure that the rules created are positively used.

Finally, data governance strategies help to ensure that the different types of data delivered, are strategically used by the right people, at the right time and in the right place.

The architecture of a data governance strategy when analysed closely, it is formed by principles needed to be effective to those who implement it.

These principles are:

1. *Define a data governance process*: doing this allows those who implement the strategy to monitor and reconcile the data between the focus of the data and its sources and users.

the data governance process should cover not only the initial data upload, but also the refinement, standardization, and aggregation of data.

²⁶ Berson, A., Dubov, L. (2011). *Master data management and data governance*, McGraw Hill Education, New York.

²⁷ *Ibidem*.

Data governance should define the mechanisms that create and maintain that cross-reference of information by applying different approaches.

the fundamental thing that a data governance process should have, in addition to those already mentioned, is that of having the possibility of being able to manually correct the defects deriving from the bad data assembly, so as to avoid errors.

2. *Design, select, and implement a data management and data delivery technology suite:* when data governance strategies are developed, data management, distribution and delivery technologies are fundamental for the implementation, and play a key role for the company. In fact, these technologies allow companies to fully integrate these solutions without using specific patterns
3. *Enable auditability and accountability for all data under management that is in scope for data governance strategy:* Auditability is extremely important for data governance strategies since, in addition to providing verifiable records of activities, it provides a tool for achieving the required compliance parameters. Furthermore, auditability is in close contact with the accountability since it requires the strengthening of some company roles such as the data owners and the data stewards.²⁸

Data owners are the people who can manage and exercise substantial control of the data within the company.

The role of data steward is slightly different. Data stewards not own the data, and above all, they do not have full control of it. This position is fundamental to ensure that the data is constantly adequate, and qualitatively acceptable.

To be effective those who perform data control should work closely with data architects, database administrators, designers ETLs (Extract, transform, Load), business

²⁸ Berson, A., Dubov, L. (2011). Master data management and data governance, McGraw Hill Education, New York.

intelligence and reporting application architects and business data owners in order to use specific data policies and data quality metrics.²⁹

1.5.1 Data quality management

Nowadays, an even growing number of decisions within companies are data based; to make the best decisions, therefore, data must be qualitatively very high. The level of efficiency, effectiveness and quality of data is called *Data Quality*.

There are multiple data quality definitions. Berson. A defines this discipline as: " one of the key components of any successful data strategy and data governance initiative, and is one of the core enabling requirements for master data management. Conversely, master data management is a powerful technique that helps enterprises improve the quality of master data".³⁰

Data Quality can also be defined as the measure of accuracy, completeness, consistency timeliness, interpretability and believability, and this parameters make data appropriate or of high standard.³¹ Furthermore, in the current scenario, the degree of excellence and completeness shown by the data can be defined as *Data Quality*.

Fig. 3 Representation of *Data Quality*



Source: Patel, J. (2015). Data quality management, tools and techniques, IJCS, 7, 1

²⁹ Berson, A., Dubov, L. (2011). Master data management and data governance, McGraw Hill Education, New York.

³⁰ *Ibidem*.

³¹ Patel, J. (2015). Data quality management, tools and techniques, IJCS, 7, 1.

Data quality, however, is not meant to solely improve the quality of the data, but it concerns the entire organization, the way in which it manages data itself, and how it can improve it based on continuous learning. The desire and ability to improve the quality and usability of data must be the main goal for those who implement data governance strategies.

The main activities for those who implement this kind of strategies, paying particular attention to data quality are:

- Identify stakeholders, establish decision rights, and clarify accountabilities
- Set direction for data quality
- Monitor data quality
- Report status for quality-focuses initiatives.³²

On the practical level, companies have numerous tools at their disposal that they can use to raise the quality of available data. These tools are:

1. *Profiling tools*: these tools improve the accuracy and correctness of data at the source. They can determine the cardinality of certain attributes, value ranges of attributes in the data sets, and missing data values, in addition to other things.
2. *Data-cleansing tools*: these tools use various machine learning techniques, deterministic and probabilistic, to correct problems found in data by profiling tools. In general, these tools compare the data used with a set of business rules and domains stored in the metadata repository or in external repositories.
3. *Data-parsing and standardization tools*: these types of tools, subdivide a single record, or single datum, in several parts so that it can be used in different moments. The standardization tools, on the other hand, convert the data attributes into a so-called “canonical format” or a “canonical data” model used by all the components of the data acquisition process.

³² AA.VV. (2011). Enterprise information management: best practices in data governance, Enterprise Architecture, May.

4. *Data Extract, Transform, and load (ETL) tools*: these are not strictly defined as data quality tools. ETL tools are designed to extract data from known structures of systems based on prepared and validated data mapping sources, transforming the inputs deriving from the data extracted into predefined targets, loading the processed data into specific data environments. Furthermore, they are useful for being able to maintain integrity constraints.
5. *Hybrid packages*: these tools are the most versatile since they contain both ETL components, parsers, standardization, profiling and data cleansing components.³³

There are different strategies and techniques to improve and increase data quality; these can be *Data Driven* or *Process Driven*.

- *Data Driven*: this type of strategy allows you to change the company's value directly by working on the data. The techniques related to this strategy are acquisition of data, standardization or normalization, error localization and correction, record linkage, data and schema integration, source trustworthiness and cost optimization.³⁴
- *Process Driven*: this strategy is process-driven, which restructures the processes and the various procedures or modifies the data in order to improve and increase its quality. The main techniques used in the implementation of this strategy are process control, where data is controlled and managed throughout the manufacturing process, and the redesign of processes, where invalid procedures are eliminated and replaced by qualitatively better processes.

Over a long period of time, between the two strategies and the different types of techniques, *Process Driven* is better suited for the long term; moreover, the *Data Driven* is very expensive, but it is more efficient and effective in the short term economically speaking.³⁵

³³ Berson, A., Dubov, L. (2011). Master data management and data governance, McGraw Hill Education, New York.

³⁴ C. Batini, C. Cappiello, C. Francalanci, A. Maurino, "Methodologies for data quality assessment and improvement," ACM Computing Surveys (CSUR), vol. 41, p. 16, 2009.

³⁵ *Ibidem*.

Analysing the specific data, we can have different classifications of it, in different areas. Data can be *implicit* or *explicit*, and within this division data can be of different three kinds. A second type of classification sees data considered as a *product*, again in this area data can be in three different classes.

Among the *implicit* or *explicit* data, we find:

- *Structured data*: Generalization or aggregation of items described by elementary attributes defined within a domain.
- *Unstructured data*: A generic sequence of symbols, typically coded in natural language.
- *Semi structured data*: data that have a structure with some degree of flexibility.

36

On the other hand, between *Product Data* we have:

- *Raw Data Items*: smaller data units are used to create information and to compose data items.
- *Component Data Items*: data is constructed from raw data items and stored temporarily until final product is manufactured.
- *Information Product*: Data, which is the consequence of performing manufacturing activity on data.³⁷

Finally, there is a third data classification method, which is less used, but is effective especially in the analysis phase. This type of classification sees the data being divided into *elementary* and *aggregates*. Among the *Elementary Data* we find information such as age, sex, height, weight and so on. Among the *Aggregated Data* we find for example the average salary, the average revenue, average income.³⁸

In conclusion, it can be affirmed that data can be classified in different ways, in order to find it easier to use for our decision-making process.

³⁶ Sidi, F. et al. (2014). Data quality: a survey of data quality dimensions, Journal of Quality Management, 4.

³⁷ *Ibidem*.

³⁸ *Ibidem*.

1.5.2 Data Lifecycle

Not knowing how and how much a data can be useful, companies have different ways to create usage patterns for data, which are necessary to ensure that the data is suited for the company. This adaptation helps to increase efficiency and effectiveness, and to minimize usage costs. We can therefore say that the data have a cycle of life within the company, and the discipline that manages this life cycle is called *Data Lifecycle Management (DLM)* which is the process of managing business data throughout its lifecycle from conception until disposal across different storage media, within the constraints of the business process.³⁹

The life cycle is understood as the time that elapses from the moment in which the data is created until it is cancelled or stored by the company.

The lifecycle data management is, however, composed by different phases, depending on the way they are used. A widely used model is the one called POSMAD, which includes six different phases which are: Plan, Obtain, Store and Share, Maintain, Apply and Dispose.⁴⁰ The other model, mainly used in the industrial sector, has an extra phase, so seven in total, which are: Data Capture, Data Maintenance, Data Synthesis, Data Usage, Data Publication, Archive and Purgings.⁴¹

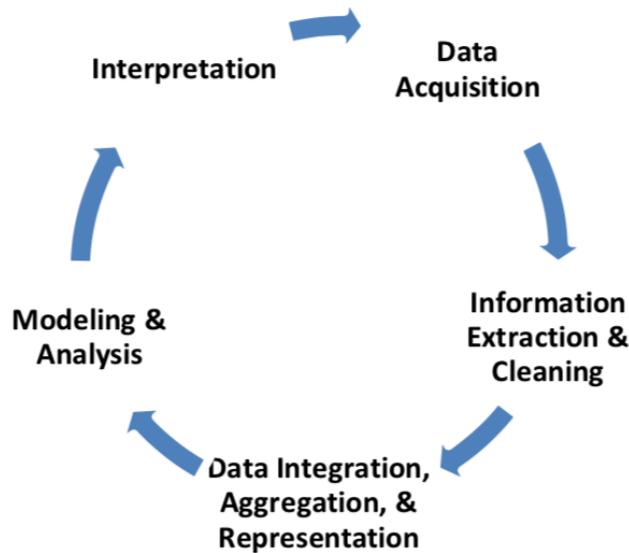
It is, however, possible to synthesize the two points above in a single model, to analyse it. A lifecycle is characterized by five phases, which are: Data Acquisition, Information Extracting and Cleansing, Modelling and Analysis and Interpretation.

³⁹ Lundberg, J. (2003). Understanding data lifecycle management, Veritas Architect Network.

⁴⁰ D. McGilvray, Executing Data Quality Projects, Amsterdam: Morgan Kaufmann Publishers, 2008.

⁴¹ M. Chisholm, "7 phases of a data life cycle," Information Management.

Fig. 4 Data Lifecycle phases



Source: Kourik, J., Qang, J. (2017). the intersection of big data and the data life cycle: impact on data management, *International Journal of Knowledge Engineering*, 3, 2.

- *Data Acquisition:* in this phase there are many challenges concerning for example the volume intended as quantity, as we often gather more data than we can actually process; speed challenges when data must be processed as quickly as possible while maintaining a very high quality; moreover, the variety of sources will increase, so data comes from different parts and it is necessary to manage them adequately.
- *Information Extraction & Cleaning:* once the data has been acquired and processed one must be able to extract the right information from the acquired data; this therefore appears to be the most delicate phase since it is necessary to know how to shape the data to make it useful, so to have the right information.
- *Data Integration, Aggregation and Representation:* if the data is not complete, it an intermediate phase is inserted. This corrects, integrates and aggregates data in such a way as to make the incomplete data, complete. At this stage, it is necessary to be very careful since the manually corrected data are then difficult to correct automatically, through the computation systems.
- *Modelling and Analysis:* this is one of the final stages through which the data is modelled and subsequently analysed to find a final result.

- *Interpretation*: this is the last step of the cycle, where the results found in the analysis phase are interpreted and evaluated. While the previous phases, relied heavily on the work done by the "machines", like the computational systems, this phase relies mainly on human evaluation. This phase is crucial because, based on how the data is interpreted, future business decisions are made.⁴²

In recent years, a different model has been developed called Smart Data model. The development of this new type of model was due to the fact that according to a research carried out by Capgemini⁴³, between 2014 and 2015, only 27% of the IT managers interviewed were satisfied with their data governance projects. This new type of cycle was therefore created to transform Big Data into Smart Data, forming the Smart Data Lifecycle. This type of cycle has similarities with those mentioned above but being composed of twice the phases of the previous ones, is much more accurate.

The different phases of this new type of cycle are fourteen: Planning, Management, Collection, Integration, Filtering, Enrichment, Analysis, Visualization, Access, Storage, Destruction, Archiving, Quality and Security.

- *Planning*: this step is part of the management process and covers the lifecycle across the board. It is monitored by a special team, which regulates the flow of capital and resources necessary for implementation. In this phase it is also necessary to foresee all possible problems and try to correct them before they happen. Moreover, the team provides a detailed description of the necessary data, on how it will be used and how it will be managed.
- *Management*: this phase is part of the planning process; here all the operational resources that directly influence the data are planned. This phase covers the whole lifecycle of data, from the beginning to the end, and it is also a way to make the different phases communicate effectively. This phase has a dedicated team.

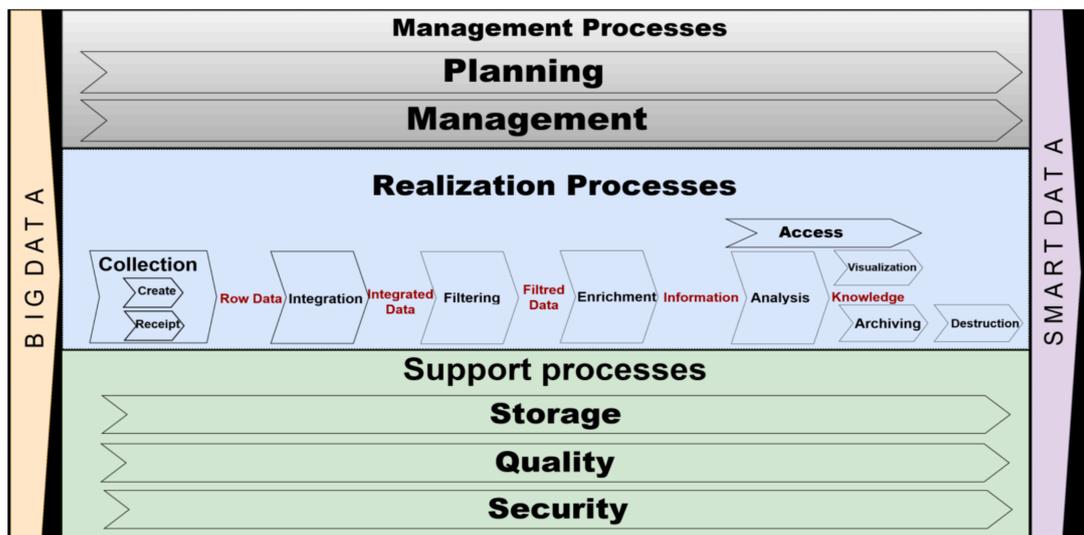
⁴² Kourik, J., Qang, J. (2017). the intersection of big data and the data life cycle: impact on data management, International Journal of Knowledge Engineering, 3, 2.

⁴³ Arass, M., Souisse, N. (2018). Data lifecycle: from big data to smart data, Conference Paper.

- *Collection*: this is usually the initial phase for each data lifecycle. All raw data is collected and classified from different sources, and it begins to be organized in the best way. During this phase it is necessary to perform data quality processes that optimize data to improve its future use; it is better done at this stage because if done later, it could be expensive and ineffective. Data, depending on the company, can be taken both from the outside or created internally to the company.
- *Integration*: during this phase, a model is provided that is suitable for data coming from independent, distributed and heterogeneous sources, so as to facilitate users in managing this data. Data from multiple sources is put together to create a data set that is easy to analyse.
- *Filtering*: During this phase, the enormous amount of data is carefully filtered and reduced. Data is not automatically discarded but is filtered because unfiltered data can be useless and would not add value to the company. This phase is necessary for the subsequent ones, since it decreases the time needed for computation, and above all reduces storage space. Scarce attention at this step can have a negative impact on the company, especially if the decisions taken are based on the data found.
- *Enrichment*: during this phase data is structurally or hierarchically modified; this allows to add information to increase its quality.
- *Analysis*: this is one of the steps that all data lifecycles share. During this phase the data is exploited and analysed to reach conclusions and interpretations relating to the decision-making process.
- *Access*: in this phase the access within the big data application provider is focused on the communication or interaction with the data consumer.
- *Visualization*: it consists in exposing the data found in a clean and intelligent way, so the decision-makers could quickly understand the results found.
- *Storage*: this phase is common to all data lifecycles and must be transversal in the project. As the name implies, this phase foresees the storage of all the data, so its traces remain for future use and to be consulted repeatedly. This phase must be managed with reliability, availability and accessibility.

- *Destruction*: at this stage the data is deleted once it has been used successfully; however, this phase is not common to all data lifecycles.
- *Archiving*: in this phase the data is stored for the long term, so that it can be reused. The crucial part of this phase is the process in which data is stored as it must be done as effectively as possible. Three fundamental operations are therefore necessary: encryption techniques, long-distance storage and data retrieval mechanism.
- *Security*: this phase includes three basic parameters which are: data Integrity, access control and privacy. However, these parameters must be checked throughout the lifecycle. This allows the data to be reliable and secure.
- *Quality*: quality is pursued during all the phases of the process and is defined thanks to the imposition of well-defined parameters, the level of precision and the implementation of controls that measure the satisfaction of the data quality.⁴⁴

Fig. 5 Smart Data Lifecycle representation



Source: Arass, M., Souisse, N. (2018). Data lifecycle: from big data to smart data, Conference Paper.

⁴⁴ Arass, M., Souisse, N. (2018). Data lifecycle: from big data to smart data, Conference Paper.

1.5.3 Metadata architecture and data integration

Metadata describes what data is for, moreover, they provide a mechanism for making a clear and exhaustive description of data representation. In a data governance strategy, the use of metadata is necessary because it make the data interpretable by users; standardizing the metadata ensures the ability to effectively use the data and to ensure that the information is traceable.

There are different types of metadata that play different roles in discovering, retrieving, collecting, managing and analysing data.⁴⁵

Singh et al classify metadata as physical, domain-independent, domain-specific and user metadata.⁴⁶ Neera Bhansali defines metadata based on types of business, technical, process and operation.⁴⁷

This type of use is done because companies must focus on reducing costs and pursuing operational excellence, and this implies using an integrative data architecture.

The main activities involved in these programs are:

- Identify stakeholders, establish decision rights, clarify accountabilities
- Ensure consistent data model and definitions
- Support architectural policies and standards
- Support metadata programs, master data management and enterprise data management
- Bring cross-functional attention to integration challenges.⁴⁸

⁴⁵ Zhang, N., Yuan, Q. (2016). An overview of data governance, Economics Paper, December.

⁴⁶ Singh, G., Bharathi, S., Chervenak, A., Deelman, E., Kesselman, C., Manohar, M., Patil, S., & Pearlman, L. A metadata catalog service for data intensive applications. In Proceedings of the ACM/IEEE SC2003 Conference on High Performance Networking and Computing.

⁴⁷ Neera Bhansali (2013). Data Governance: Creating Value from Information Assets. CRC Press.

⁴⁸ AA.VV. (2011). Enterprise information management: best practices in data governance, Enterprise Architecture, May.

1.5.4 Security and privacy data governance

With the regular and growing acquirement of data, but above all with the use and the reliance that companies have on it, it is also mandatory to find a way to protect this data.

Protecting data does not mean limiting its circulation or ability to share, but it will ensure that this is kept secure, and it is released without any risk.

It is important to clarify the difference between privacy and confidentiality. Privacy usually refers to data relating to a person, while confidentiality is attributed to data relating to a company.

The main objective of data security is therefore to protect the data relating to a company and its consumers or suppliers. There are different mechanisms to ensure the privacy and confidentiality of the data. The following listed methods make sure of that:

- *Anonymization*: personal and identifiable information are removed or masked from data sets.
- *K-anonymity*: process in which data points are suppressed or generalized so people contained in data sets cannot be distinguished from at least k-1 individuals in the same dataset.
- *Data perturbation*: process by which “noise” is added to data in order to be distorted and less identifiable.
- *Differential privacy*: this method incorporates built-in mechanisms that ensure that the output of queries to a differentially private database yield the same conclusions irrespective of whether a particular individual’s data is present in the database.⁴⁹

Analysing now security, a company in order to better protect its data, must work to develop and implement *Data Security Plans*.

To be effective and efficient, a data security plan should consist of three main elements, or phases: *Data Use and Handling Procedures*, *Vulnerability Assessment and Risk Management*, and *Data Security Incidents and Breaches*.⁵⁰

⁴⁹ AA.VV. (2018). Data management, privacy and security in connected systems, Interact Lighting.

⁵⁰ Hendey, L. et al. (2018). NNIP’s resource guide to data governance and security, NNIP, September.

Data Use and Handling Procedures

Procedures must be established to ensure that confidential or protected data is transferred in a secure manner and above all by authorized persons. Depending on the data, it must be encrypted, or password protected.

Furthermore, it may be useful to limit the number of entries that can access the data to reduce the risk of errors or flaws in the system. that is why it is useful and effective to physically control data access.

Moreover, to ensure security, even the storage of data must be secure, so for physical devices, such as flash drives or CD's, it is preferred if they are locked in protected places. while for the data it is good to insure it with password or encrypted codes.⁵¹

Vulnerability Assessment and Risk Management

An effective strategy to increase data security is that of conducting vulnerability assessments and developing a plan to address risk.

Therefore, the data security audits are useful tools that the company can use to test its predisposition to data security. The company can in fact incur in various threats. These types of risks can be of two main categories, technical and non-technical. Technical risks can be for example no existing security architecture, unpatched client-side software and applications, "phishing" and targeted attacks, websites, poor configuration management, mobile devices, cloud computing, removable media, botnets, and zero-day attacks. Listing non-technical threats, they are insiders, poor passwords, physical security, insufficient backup and recovery, improper destruction, social media and social engineering.⁵²

Data Security Incidents and Breaches

In this phase, procedures are set up by the organization to solve or better manage the problematic situations or accidents. Liability insurance is one of the considerations that the company can make to remedy to possible accidents.

⁵¹ Hendey, L. et al. (2018). NNIP's resource guide to data governance and security, NNIP, September.

⁵² *Ibidem*.

The procedures that manage these unpleasant situations should be easily implemented by the employees, they must have clear instructions on what to do. Accidents should be reported internally, and actions should be taken, which will ensure the right training and the right revision of the procedures.⁵³

⁵³ Hendey, L. et al. (2018). NNIP's resource guide to data governance and security, NNIP, September.

CHAPTER 2 – Data Governance process: implementation and peculiarities

2.1 Building a data model: premise

In this chapter I will try to analyse closely how to build and implement an efficient, effective and useful data model for the company. Building a good data model is a fundamental part of the structure in which the company data resides and evolves. This process allows the company that implements data to preserve it, ensuring data integrity, finding the basic requirements for its business and finally communicating the primary purpose of company data. There must be an alignment in business strategies, in business operations, and in supporting technologies.

Developing a data model has constantly changed over the years, becoming, nowadays, a discipline that determines the common languages between companies for data management. This discipline determines and manages data modelers, business analysts, database administrators, application and database developers, data governors and data stewards. Data modelling is used in different sectors, and can be of canonical or conceptual type, determining the different relationships within companies, either logical or physical, determining the different systems or applications.

A unique data model provides business-level and structural development for companies that implement it, provides standards for data integrity, information on how data management can support business processes, and determine what data can be selected or used, and the different paths that the company should undertake.

Data modelling also provides multiple tools to those who design and manage the model, data developers and data administrators, which is able to evaluate a multiplicity of data within numerous databases.⁵⁴ Data model can therefore be fundamental for the development of a company that wants to grow and be competitive.

The basic requirements that a data model must have are:

- Optimization for entity resolution, relationship resolution, and fast searches of master entity records.

⁵⁴ Berson, A., Dubov, L. (2011). Master data management and data governance, McGraw Hill Education, New York

- Flexibility to support multiple data domains with a variety of master entities and relationships
- Ability to support consuming applications such as data warehouses, data marts, operational data stores, and so on.⁵⁵

Data models define the different actions, mechanisms, relationships and interactions that the company must develop and implement.

In order to be effective, models must be made of different components. These components are:

1. A structure that allows the company to manage design and reporting architecture, the composition of the board and to control corporate actions interdependently.
2. Try to assign the different responsibilities according to the oversights of the board, the committee and the management, also defining the authority to hire or dismiss those responsible.
3. The disposition of talent and culture, which shapes the character and actions of those who work in the company, establishing remuneration, especially incentives, promotions, operating principles, evaluations and performance management and lastly training.
4. Infrastructures, which include the management of the possible risks of bankruptcy deriving from the procedures, reports, assessments, and capacity of the management.

An effective data model allows the execution of different responsibilities within organisations, in all of its existing levels. This is obtained through to the ability of these models to clarify company hierarchies and relationships, specifying the decisions and the risks that may occur during the work of the company; also by clarifying the roles, responsibilities and limits of the authority of the various employees. Therefore, the developed model must possess certain peculiarities which determine its effectiveness.

These peculiarities are: *Board oversight and responsibilities, committee authorities and*

⁵⁵ Berson, A., Dubov, L. (2011). Master data management and data governance, McGraw Hill Education, New York.

*responsibilities, organisational design and reporting structure, management accountability and authority, performance management and incentives.*⁵⁶

- *Board oversight and responsibilities:* the model should allow the council to specify the different peculiarities and abilities required to avoid oversights, thus trying to compose the board in a more appropriate way; entice management to improve the flow of information they deliver to the board, and that it uses to make decisions; put employees in charge on the policies that characterise the government, so that they align themselves with them; try to understand the different activities necessary for the various units of the company, so as to increase their efficiency.
- *Committee authorities and responsibilities:* efficient models allow to better define the quantity, quality, composition of the board members, specifying their responsibilities and clarifying the mechanisms that exist between the different members and the different interactions that occurs. This implies the presence of responsibility charts, the different items the board must pay attention to, and the definition of the reporting methods to be used.
- *Organisational design and reporting structure:* a good data model also manages to specify the different hierarchies that exist in the decision-making process, the management of the risk and the activities that can be problematic, the description of the financial and regulatory work, and the mechanisms for responding to crisis situations.
- *Management accountability and authority:* authority and responsibility at all company levels must be appropriately specified and aligned with the objectives. Objectives must be distinguished to distribute the various responsibilities in a popular way; clarifying authority in the decision-making process for each business unit, to avoid risky and problematic situations; define the different directions for each employee so that the authority and consequent limits are clear.
- *Performance management and incentives:* obviously, the entire remuneration and incentive part must also be proportionate to the company's work and must reflect the effort that the company makes to implement data models. The model

⁵⁶ Deloitte (2013). Developing an effective governance operating model, in www.deloitte.com

must therefore be useful to the board to establish the performance objectives, align incentives to find the balance between resources used and expected objectives, carefully assessing the existing risk, and finally specifying the different roles with the respective assessment methods.⁵⁷

The benefits deriving from the implementation of these models are numerous, and overall, they overcome the disadvantages, going to improve the work and the value of the company. These benefits are:

1. Improve clarity by implementing the governance principles described in the model in governance practices. This improves the definition of roles, information flow responsibilities, and guidelines to follow.
2. Increase visibility to better understand the path to follow in the decision-making and risk management process.
3. Improve coordination by addressing the complexity inherent in the governance of multiple businesses into a single organization.
4. Increase effectiveness, specifying the information needed by the board in order to proceed with the decision-making process.⁵⁸

2.2 Implementation of a *Data Management* program

There are two main thrusts that lead a company to develop and implement data governance programs; these thrusts derive either from business strategies or from IT strategies.

When the programs of data governance are driven by business motivations, *business strategy driven*, the main source that lead to the development of data management programs comes directly from the executive (CEO, CFO, CIO etc). This occurs when the aforementioned executives want to avoid the operational inefficiencies that could occur within the company. Moreover, this methodology is used when the company communicates inefficiently with its customers, with its suppliers, and has little control over its product. The implementation methodologies can be the development of a new central operating system or develop a new conceptual master data model.

⁵⁷ Deloitte (2013). Developing an effective governance operating model, in www.deloitte.com

⁵⁸ *Ibidem*.

The second scenario that can occur when data governance programs are developed, are those driven by *information technology strategies*. This methodology turns out to be the most particular and significant, since those who deal with information technology often do not have the same knowledge relating to the client or to the executive product. In this scenario a so-called “golden model” is much easier to develop. A golden model implies perfect data collection, or as close as possible to perfection. This method, however, has limitations mainly related to the relations with users, which can lead the system to fail.⁵⁹ However, the choice of the two different scenarios is different for each sector, industry and company.

Benefits that nevertheless occur from the above scenarios can be of two main categories: *Quantitative Benefits* and *Qualitative Benefits*.

- *Quantitative Benefits*: they can be calculated based on cost savings, increased revenues, or the elimination of the possibilities where the revenues are wasted. These benefits are also called “hard” benefits.
- *Qualitative Benefits*: these “softer” benefits concern corporate aspects such as risk, company competitiveness and problems regarding potential compliance that could be harmful to the company but are difficult to quantify.⁶⁰

To estimate the impact of implementing a data governance program within a company, and to understand its benefits or disadvantages, there are two main approaches that can be used: *traditional bottom-up approach* and *less traditional economic value approach*.

The former quantifies the benefits deriving from the implementation of data management programs using an in-depth analysis of current processes, evaluating inefficiencies and developing a ROI model that identifies cost savings and revenue growth.

The latter involves a very accurate estimate of the impact that the data management program and other information governance initiatives have on the business.

⁵⁹ Berson, A., Dubov, L. (2011). *Master data management and data governance*, McGraw Hill Education, New York.

⁶⁰ *Ibidem*.

Furthermore, there is a third type of approach which is called “interview-style” approach, which is used to understand the level of security on data that a company possesses, or the level of risk to which it is exposed.

When programs of this type are developed and implemented, attention must be focused on three main areas: *People*, *Process*, and *Technology*.

- *People*: data should be central to the entire company. It creates interconnections that involve the entire company. In fact, "abandoning" data only to those involved in information technology becomes counterproductive for the entire business, not supporting the demand and the various opportunities that appear.⁶¹
- *Process*: *Data Management* programs are never complete in themselves, they must continually be adapted to the changes that characterize the environment in which the company operates. We can state that the process of data governance must be an evolutionary process, composed of small and measurable steps, which lead to long term objectives.

As mentioned in the previous chapter this process must be composed of several phases, five in all, which are: *Data investigation and explorer*, *design*, *employment*, *maintenance and redundancy*.⁶²

- *Technology*: technology plays a key role in determining which data governance program to implement in the organization, significantly reducing risk and increasing data value over the long term. Technology can help standardize data in order to improve quality, manage the associated risk and develop a more "secure" program within the company.⁶³

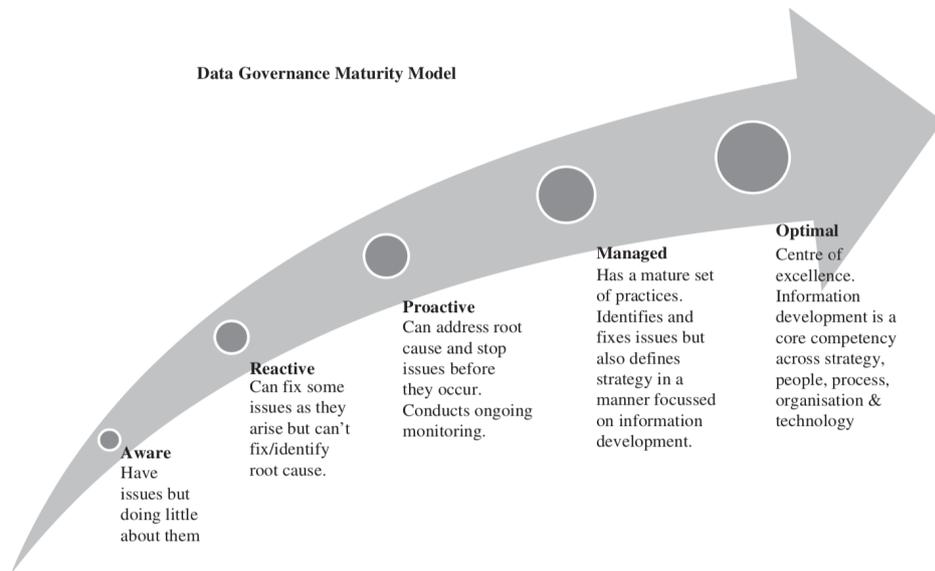
When developing data governance programs, it is also necessary to evaluate what the maturity of the current management program is. For this to be possible, it is useful to use Data Governance maturity method which identifies the capacity of governance programs following different steps. This process must be developed in the long term. The figure below can help understand the approach described, in a graphical manner.

⁶¹ Gregory, A. (2010). Data governance – protecting and unleashing the value of your customer data assets, *Journal of Direct, Data and Digital Marketing Practice*, 12, 3.

⁶² *Ibidem*.

⁶³ *Ibidem*.

Fig.6 Data Governance maturity model.



Source: Gregory, A. (2010). Data governance – protecting and unleashing the value of your customer data assets, *Journal of Direct, Data and Digital Marketing Practice*, 12, 3.

To now expand on the graphic above:

1. *Aware:* since the data is managed in "macro-containers" you have to follow some rules and policies in order to use them in the most correct way. At this first level there is no supervision by the executive, or if there is, it is very limited; the commercial and financial part must not be synchronised. The main features in this phase are: inconsistent and duplicate data; the company is unable to adapt to market changes; localized data management; the projects are mainly guided by the information technology department, and the technology is mainly used for specific problems or special needs.⁶⁴
2. *Reactive:* reactive companies are committed to obtaining regulatory requirements. Data, however, is not managed by the organization itself, and data sharing is rare. However, at this stage, the company pursues data quality but the actions that are undertaken are deriving from problematic situations that the company tries to cope with. An example may be the failure of the unit that deals with customer relationship management that leads the company to focus on the

⁶⁴ Gregory, A. (2010). Data governance – protecting and unleashing the value of your customer data assets, *Journal of Direct, Data and Digital Marketing Practice*, 12, 3.

data possessed in order to resolve the situation. The main stages that characterize this phase are: cross-functional collaboration; high maintenance costs for multiple databases and applications; unexpected success is given to "heroes" the little technology used results in the storage of data and relative to CRM.⁶⁵

3. *Proactive*: In this phase the data is monitored and analysed much more accurately. At this level the data is used as a strategic asset by the entire organization and the company perceives data as the true added value of the business. The main features of this phase are: the employment of specialized and experienced people who know how to use the data in order to obtain a value; IT and business staff work closely together; data is perceived as a corporate asset.⁶⁶
4. *Managed*: at this point the company believes that the main value of data possess comes from the sharing of the data; the prevailing technologies in this phase are CRM programs, Enterprise Resource Planning (ERP), and technologies related to data storage. The predominant characteristics that we find at this level: to perceive the level to which one finds oneself and to evaluate how to improve any defects, the technologies used are related to the automation of data security and compliance management technologies.⁶⁷
5. *Optimal*: the only goal left at this point is to reinforce the best practices and to code the different data throughout the organization. The main attention in this phase is reserved to the integration of the processes more than the use of data in itself. The main features encountered in this phase are: automated and repeatable business processes, business requirements are managed by information technology projects, business operations are optimized, data governance strategies are unified, and there is also a use of data deriving from the outside that are integrated with those already held by the company. The main technologies are found in the automation of business processes, and in the master data management.⁶⁸

⁶⁵ Gregory, A. (2010). Data governance – protecting and unleashing the value of your customer data assets, *Journal of Direct, Data and Digital Marketing Practice*, 12, 3.

⁶⁶ *Ibidem*.

⁶⁷ *Ibidem*.

⁶⁸ *Ibidem*.

2.2.1 Basic components for the platform: initial project

Initially, before starting the development of a data governance project, it is good to identify and classify the different types of data, necessary for our project. Data is classified into two macro domains: *transactional* and *non-transactional data*.

Among the *transaction data* we can find customer call records or the tracking of a person's banking transactions. Usually this type of data is fundamental for compiling regulatory reports; furthermore, this type of data plays a key role in every business intelligence strategy.

On the other hand, we find the *non-transactional data*, called *pure reference data*. among these data we find the personal information relating to the various users (name, address, age etc), suppliers and employees. These data are fundamental for business as they provide useful details to the company regarding the development of its product or the provision of services.⁶⁹

Another useful distinction is made between *operational* and *non-operational data*. the *operational* ones are data that the company uses as support for its daily activities, those carried out every day. *Non-operational data*, on the other hand, are collected, stored and used with less frequency, only when business intelligence programs are developed.⁷⁰

This distinction is made to divide data governance programs into *operational* and *analytical* programs; there is also a third hybrid category that includes a combination of data, both *analytical* and *operational*, which is called *Enterprise Data Management*. *Operational* governance programs therefore include applications such as enterprise resource planning (ERP), customer relationship management, and supply chain management. Among the *analytical* ones we find applications such as integration of customer data or financial performance management. Finally, enterprise data management is used to maintain and publish all the master data in the organization.⁷¹

Subsequently, after having distinguished the different types of data, we move on to the definition part of the basic components for the development of the project. To make sure that the activities carried out in the company and the decisions taken by the

⁶⁹ Purohit, P. (2014). Master data management, Conference Paper, March.

⁷⁰ *Ibidem*.

⁷¹ *Ibidem*.

executive are reliable, timely, complete and precise, IT should facilitate access, sharing and integration of data.

The platforms that manage the data are tools that companies can use to obtain competitive advantages. To be effective these platforms must have some basic characteristics, which are relevant to the business, especially for data governance strategies⁷². These features are:

- *Data Integration Services*: this feature allows the company to access all those fragmented data without counting its origin, who has already used it, or what type of application created it. It is a service that starts to model the data in such a way that they become an asset for the company.
- *Data collection and detection services*: data integration allows the platform to identify related data and to group it in a central index. The aggregation also allows the platform to add ability to find the main features of interest for its business.
- *Data quality services*: this type of feature ensures that the platform finds and uses data that is consistent with the quality standards of the user company. It can therefore involve specific tags, data classification, data filtering and enrichment of data with additional metadata, to increase the quality and value for the company.
- *Data discovery and search services*: one of the fundamental and most particular features is how the data is channeled through this service. In fact, some data remain in their place of origin, while others are aggregated and centralized through this platform
- *Data utilization and curation services*: this feature allows the company to continuously evaluate the usefulness of the data, what type of data they are, where they are located, what needs to be done with the data once they are no longer usable, who can have access to the aforementioned data and who is the responsible.
- *Data exchange service*: this is the interface of the industrial standards that allows the application to interact with data from a centralized location. they are

⁷² Baker, S., Sjoberg, P. (2018). Intelligent data governance, John Wiley & Sons, Hoboken.

also the entry points that developers use to integrate other systems and services into the platform in question.⁷³

- *Data management service*: with this service the data can be checked, secured, protected, and made immediately available. The policies that manage the data, the degree to which they are left in "movement" can be managed entirely by this type of service. The policies governing data and data policy applications are therefore defined. This is what makes this service so valuable to the company.⁷⁴

2.2.2 Design and analysis of the process

To define a data governance model, it is necessary to define and document three different categories concerning the company: components, sub-components and activities. The purpose is to define decisions and actions useful to the company in order to reach its goals. The process of documenting the governance model can create as much value as other documents used by the company. However, if a company does not have a documented model, those who carry out the decision-making process can focus on things that are incorrect, such as balancing competing goals, defining responsibilities, allocating resources and choosing decisions, all activities that are necessary to the implementation of a governance activity.⁷⁵

When a company is in the decision-making phase of its own governance model it should take into account its current position occupied in the market, understand and define the future position that it would like to achieve, defining, during the process, the various steps to be followed to obtain the desired results. The company should consider objectives such as: *Compliance issues, Cultural shift, Governance and Management decision rights, Process and System issues, Regulatory relations, Human resources.*

- *Compliance issues*: these are all the activities relating to compliance. It is important to align the compliance of the imposed standards with risk management, responding to situations of need in an integrated and globally coordinated manner.

⁷³ Baker, S., Sjoberg, P. (2018). *Intelligent data governance*, John Wiley & Sons, Hoboken.

⁷⁴ *Ibidem.*

⁷⁵ Deloitte (2013). *Developing an effective governance operating model*, in www.deloitte.com.

- *Cultural shift*: the company must implement a good cultural level internally, by improving the organization. To resolve conflicts, if there are any, among the local customers; to learn the norms and the needs of the different business units; and to avoid tensions between the centralization of risk policies and the decision-making process.⁷⁶
- *Governance and management decisions rights*: there must be a manager who takes care of decisions regarding strategy, budget, various company funds, liquidity plans, recruitment and dismissal of personnel, performance management, compensation and risk management. It is needed to clarify the roles and responsibilities of management.
- *Process and system issues*: it is necessary to try to increase business processes and systems for “risk reporting” in order to support risk management. This is achieved by continually updating processes and systems in order to generate data and statistics required by the authorities and finally trying to make the controls efficient to improve performance and limit costs.
- *Regulatory relations*: these establish a business role that responds to local regulatory authorities when necessary and meet regulatory requirements related to the various units.
- *Human resources*: this take into account the internal human capital of the company; therefore, the roles and activities of each employee must be defined, and the skills required by the administration board, especially for risk management, must also be evaluated.⁷⁷

Moving to the architecture itself, we need to analyze the different models that can be used. The first one is the so-called Single Central Repository Architecture (SCRA); in this framework the data is stored in a central repository, which will be loaded by the governance services and applications. The advantage of this type of architecture is to be able to guarantee data consistency, while the disadvantages are the high implementation costs, since they require huge efforts to convert data from multiple

⁷⁶ Deloitte (2013). Developing an effective governance operating model, in www.deloitte.com.

⁷⁷ *Ibidem*.

different systems. Fortunately, however, once this architecture has been implemented, and the huge upfront costs are incurred, the maintenance costs are contained.⁷⁸

The second type of architecture is the Central Hub and Spoke Architecture (CHSA). As the name implies, this type of architecture has a central repository, the central hub, which has the ability to retain and provide individuals with data extension. The main advantages of the development of this architecture are extreme flexibility, which is very important if development costs are taken into account, which remain more contained.

The main disadvantages of this architecture are that it does not address problems related to time and latency. However, data conversion efforts remain necessary.⁷⁹

Finally, there is the Data Service Federation (DSF) which is an architecture commonly used in a virtual way. This virtual model aggregates different data from different sources into a single form but keeping the definitions for all the sources used. The advantage of this model is it is inexpensive, since the data must not be physically copied, furthermore, since there is nothing concrete, there is no need for physical space to store data. The only issue that can be found is that the improvement of the data is not brought back to the application of origin.⁸⁰

2.2.3 Synchronization and Startup

Once the data governance process with the respective architecture has been designed and analysed, the synchronization phase of the various parts is evaluated. Data synchronization is one of the fundamental steps for implementing data governance strategies, as they are able to maintain data consistency. This step is necessary no matter what the architecture chosen for its governance project is.

There are three main types of synchronization: *Trigger based*, *Message-based Data synchronization and integration framework* (MDSIF) and *Conflict resolution*.

- *Trigger based*: in this approach what starts of the synchronization is the development or insertion of an additional source in the database record. This approach ensures that all data in different positions are always synchronized. However, this type of approach is convenient to implement in small realities,

⁷⁸ Purohit, P. (2014). Master data management, Conference Paper, March.

⁷⁹ *Ibidem*.

⁸⁰ *Ibidem*

since in larger ones, where there are many tables to analyse, the synchronization of all the data becomes very complicated.⁸¹

- *Message-based Data synchronization and integration framework*: this type of approach is based on the advantage of not having network constraints, but of being able to rely on messages obtained from the middleware. It allows not to overload the internal network, avoiding bottlenecks in general performance. The only disadvantage in this software is its large cost both in terms of initial expenses, maintenance and also infrastructure support.⁸²
- *Conflict resolution*: since multiple data can be used, deriving from multiple databases, it may happen that these data conflict with each other. With this approach we try to assign a confidence level for each datum based on the reliability of the source that last modified that datum. The one with the higher level of reliability will have a higher preference, while those found not to be reliable enough will have a lower level, and consequently will be rejected. This approach is most effective because of its simplicity, flexibility and easy ways to be adopted, even if it has to be constantly maintained by the staff.⁸³

2.2.4 Enforcing and monitoring *Data Governance*

Nowadays, the amount of data companies are faced with has increased exponentially, dashboards, reports and warnings resulting from these data have also increased. These dashboards, reports and warnings, often mask the important details for the companies, providing data that in substance is not reliable. It is therefore necessary to try to avoid these overloads of dashboards, reports and warnings as they allow the use of data that has already been governed or managed effectively, thus allowing the use of quality descriptors that attest to its reliability and accuracy. Furthermore, to ensure its quality, a company should:

- *Know its audience*: even if the data is qualitatively high, dashboards, reports and alerts are only useful if they are for their users.

⁸¹Purohit, P. (2014). Master data management, Conference Paper, March.

⁸² *Ibidem*.

⁸³ *Ibidem*.

- *Set priorities and classify alerts based on relevance:* try to increase the controls on the reports in such a way as to ensure that the right information reaches those who really need it to make decisions.
- *If everything is quiet or the metrics aren't changing, sometimes is likely wrong:* there is a need for the reports to be constantly monitored so as to guarantee their accuracy.
- *Timestamps can denote relevancy:* as not to use data that is temporally "old", to ensure that when creating reports, dashboards and warnings they also have a clearly visible date and time so as to inform the user on the accuracy of the data.
- *Document everything:* finally, it is very important to describe how the report, dashboards and warnings have been created, defined and set up in order to make them understand their main purpose.⁸⁴

2.3 Interactive and contingent approach

Given the continuous increase in the quantity of information, companies are obliged to continuously adapt their business models to the various changes. This adaptation is necessary whenever business processes are to be designed for different continents and customers. These two factors therefore influence the architecture of corporate procedures and consequently the strategies of data governance. Therefore, the quality of the data, pursued through the right programs of data governance, serves the companies to adapt their strategies to the different characteristics of the markets.

There are different approaches that companies can use to better adapt their strategies to the different needs of the markets: a contingent approach and an interactive one. The first helps companies to understand how to adapt the data governance strategy to the various organizational contingencies. There are two factors that help to understand how organizational contingency affects the company's governance model: organizational positioning and coordination of decision-making authority.⁸⁵

⁸⁴ Baker, S., Sjoberg, P. (2018). *Intelligent data governance*, John Wiley & Sons, Hoboken.

⁸⁵ Wende, K., Otto, B. (2014). *A contingency approach to data governance*, Conference Paper, November.

Organizational placement of decision-making authority

analysing the organizational positioning of a data governance strategy, this approach can be characterized by two main factors: a centralized or decentralized approach.

- *Centralized*: all decision-making roles are centrally positioned in the strategy; for example, we can mention the chief steward or the data quality board. The decisions made with this type of approach are valid for all the business units, univocally.
- *Decentralized*: approaches with these characteristics instead give the decision-making authority in the hands of external administrators, corporate and technical. This type of approach, however, is often obsolete as it is difficult to implement and above all very expensive.⁸⁶

Coordination of decision-making authority

the fundamental characteristics of this second contingent parameter are *hierarchical* and *cooperative* data governance models.

- *Hierarchical*: as the name suggests, this feature requires a decision-making process that goes from top to bottom, following several hierarchical levels within the company; the decisions therefore start from the top and slowly reach down in the different business units.
- *Cooperative*: this model, contrarily, applies coordination mechanisms that can be formal or informal in the decision-making process. Therefore, working groups, task forces, and committees are set up to help the board in choosing operational solutions. No one here will make decisions independently, but these will be discussed with other members. Moreover, the formal and informal mechanisms can be integrated with each other to find the right perfect match for your company.⁸⁷

⁸⁶ Wende, K., Otto, B. (2014). A contingency approach to data governance, Conference Paper, November.

⁸⁷ *Ibidem*.

Focusing next on the second approach, the *interactive* one, before applying it, it is necessary to analyse the degree of maturity in terms of data governance within company's boundaries. There are six different levels of data governance development: *none, initial, managed, standardized, advanced* and *optimized*.

In the first level, the *none*, no type of data governance is developed in the company.

In the second level, the *initial*, there are some authorities in terms of data governance, but these do not influence company processes; IT and business collaboration is relatively inexistent.

In the third level, the *managed*, there is an initial definition of ownership and administration of government programs, while data standardization programs are at an early stage.

In the fourth level, the *standardized*, consistent data standardization programs are included in the company, and there is also a central, easily accessible, repository of data policies; governance programs are constantly measured and monitored.

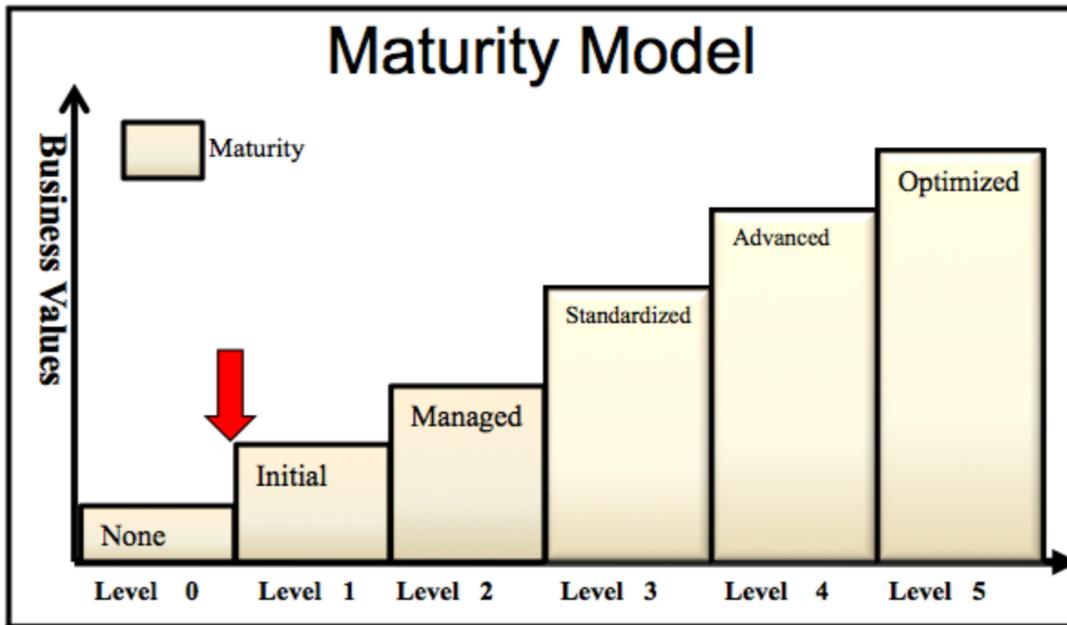
In the *advanced* stage, data governance programs are seen as institutional and perceived as a corporate asset; while objectives are set up to ensure the quality of the data to be used.

In the last level, the *optimized*, data governance programs become the main business and decisions are made based solely on the data collected.⁸⁸

The figure below helps us to understand the value, in terms of business, that each level brings to the company, and the level of maturity of the data governance programs developed.

⁸⁸ AA.VV. (2011). Enterprise information management: best practices in data governance, Enterprise Architecture, May.

Fig.7 Maturity model



Source: AA.VV. (2011). *Enterprise information management: best practices in data governance*, Enterprise Architecture, May.

The purpose of companies that develop such programs is to reach the last maturity level, the *optimized* one. There are, therefore, three phases that help companies increase their maturity for data governance programs: *explore phase*, *expand phase* and *transform phase*.

- *Explore phase*: in this phase solid foundations of data governance are built and leaders are established for these programs. These improvements should start at a local level and then rise to globalization.

The main objectives of this phase are, first of all, to understand and protect the needs of data governance. Subsequently it is obligatory to design and plan data governance documents for the implementation; it is necessary to create schemes that ensure the confidentiality, quality and integrity of data. The scheme to be defined must be integrated with the corporate objectives and activities.

Furthermore, a council or committee must be established to discuss data

governance programs. It is very important, however, not to pretend that the implementation of these programs necessarily leads to success.⁸⁹

- *Expand phase:* in this phase the coverage of the governance programs must be increased to all the business units; it then passes from a local coverage, as in the first phase, to a more extensive one. Enterprise data governance programs are developed during this phase, so the entire user base exponentially increases. One of the objectives of this phase is to increase the scale of the governance program, so as to have a greater impact.

In this phase a central repository for the data is established which enhances the collaboration and the exchange of information between the various divisions. We must continue to pursue the benefits and evaluate the increase in value for the business.⁹⁰

- *Transform phase:* at this stage the term “think-globally” is put into practice. New models of data governance must be established as the focus has expanded globally and is no longer local. The main objectives of this phase are to optimize the costs of the programs implemented. Moreover, the sustainability of programs and operations is also becoming more relevant, so the operations to be performed in order to make the program efficient, cost effective, sustainable and optimized must be identified. The main disadvantages that occur at this stage are the loss of agility and flexibility.⁹¹

⁸⁹ AA.VV. (2011). Enterprise information management: best practices in data governance, Enterprise Architecture, May.

⁹⁰ *Ibidem.*

⁹¹ *Ibidem.*

2.4 Data governance and big data

Big data management can be defined as the composition of two different terms: big data and data governance. The definition of data governance we have extensively analysed in the previous chapter, as well as the definition of big data.

To analyse the composition of both we can say that big data management is a set of disciplines, tools and platforms that are used to manage big data. Traditional data in fact differ a lot from big data because big data are "data that exceeds the processing capacity of conventional database systems. The data is too big, moves too fast, or does not fit the structures of your database architectures"⁹²

Big data is covered by four main phases: Big Data Generation, Big Data Acquisition, Big Data Storage, and Big Data Analytics. In the first phase the data is generated in large scale from different sources; then in the acquisition phase there are sub-phases which are data collection, transmission and pre-processing. In the third phase the data is stored, making it available and ready for those who need it. Finally, in the fourth phase data is analysed accurately to make forecasts, market trends, product and services.

As previously seen, the data governance discipline establishes all sets of policies and practices so that companies can manage their data correctly. Data governance framework is "a set of processes that ensures that important data assets are formally managed throughout the enterprise". This allows you to deliver the right set of information to the right people so they can make the best decisions.⁹³

The concept of big data governance has become increasingly important in the business environment, as it is a factor of innovation for companies. The more companies "understand" data and big data, the more they are able to implement data governance strategies, and the more they are able to capitalize on these adoptions on an economic level. Implementing strategies of this type can increase the challenges and risks to which the company is subjected but managing it carefully it can bring benefit to the company. According to the survey carried out by Russom P, concerning the number of companies that currently implement big data governance strategies, they are numerous. Out of a sample of 491 companies, 26% already know how to implement

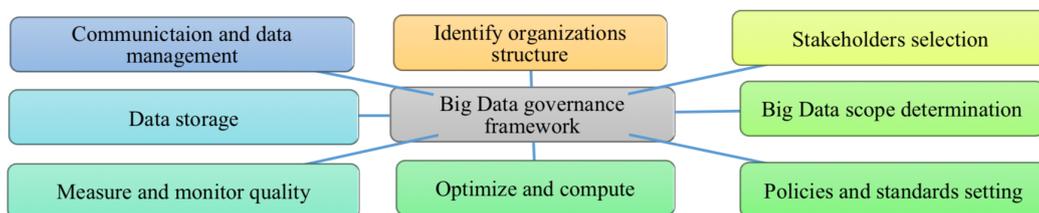
⁹² Dumbill, E., (2012), *What is big data? An introduction to the big data landscape*, in www.oreilly.com.

⁹³ Sarsfield, S. (2009), *The data governance imperative*, IT Governance Publishing.

such strategies, with a particular focus on big data; 31% say that this type of strategy is not yet being implemented, but they are gearing up to do so; 38% have never thought about implementing this type of strategy at all and finally 5% do not know what to do about it.⁹⁴

When wishing to implement strategies containing big data governance, it is necessary to develop a framework. This framework needs to be made by eight components, which are identified in the figure below.

Fig.8 Big data governance framework



Source: Al Badi, A., Tarhini, A., Khan, A. (2018). Exploring big data governance frameworks, Procedia Computer Science, 141

- *Identify organizations structure*: in this phase those who design the framework should carefully analyse the company structure; this is necessary because the framework, therefore the entire big data governance, should be in line with the objectives, the vision and the mission of the company.
- *Stakeholders selection*: the framework should take into account the parties interested in the data managed in the big data governance, in order to design a suitable program for: data scientists, data analysts, business stewards leads, data stewards, steering committee.
- *Big data scope determination*: this phase is necessary to understand if the big data taken into consideration is suitable for the corporate destination for which it is intended. It is therefore necessary to define within the framework the destination of each data so that there are no misleading situations.
- *Policies and standard settings*: the various rules, policies and practices that manage the various data must also be defined; this optimises the use of data,

⁹⁴ Russom, P. (2013). Managing big data, TDWI Research, Fourth Quarter.

reduces the possibility of risk and also tries to contain costs related to the program.

- *Optimise and compute*: this phase of the framework definition consists of data acquisition and transformation. Given that the company can benefit from data analysis.
- *Measure and monitor quality*: priority should be given to measuring and monitoring the quality of big data as it is the cornerstone of the whole big data governance program. The employees, at this stage, should take into account all the changes made with respect to the original process. Moreover, the data, given that it must be of a certain qualitative standard, must be filtered so as to have it as clean as possible.
- *Data storage*: the data must be stored in a safe place, but at the same time must be accessible and available at any time necessary.
- *Communication and data management*: finally, communications must be prepared, in report form, that the program must process for its customers.⁹⁵

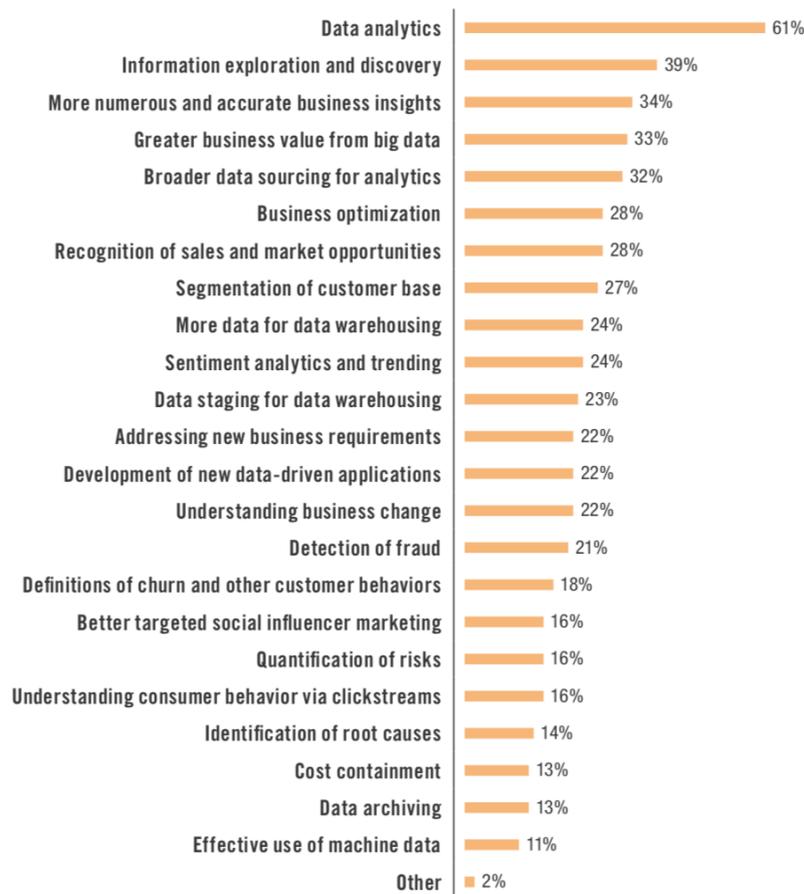
The development of big data government programs has multiple benefits but also barriers. Referring to the survey carried out by Russom P. on the sample of 491 companies, it was found that the greatest benefits are obtained by business analysts, so that the development of these programs work in close contact with the aforementioned category.

The table below helps to understand how companies perceive the implementation of big data governance programs as positive.⁹⁶

⁹⁵ Al Badi, A., Tarhini, A., Khan, A. (2018). Exploring big data governance frameworks, *Procedia Computer Science*, 141.

⁹⁶ Russom, P. (2013). *Managing big data*, TDWI Research, Fourth Quarter.

If your organization were to successfully manage and leveraging big data, which business and technology tasks would improve? Fig 9. (Answers based on 461 respondents)



Source: Russom, P. (2013). Managing big data, TDWI Research, Fourth Quarter.

There are also numerous disadvantages in the implementation of these programs. When a company has not got expertise in big data management, the main issues are found within the staff or the skills possessed; furthermore, there can be major limitations in managerial infrastructures, and immaturity in the sources used by the company. This type of problematic is overcome with an adequate system of recruitment and training of the company’s personnel. Moreover, there may be problems in data administration, when a company is not an expert in big data management, problems concerning the complexity of the data used and in the organizational architecture of the company. Once the team's experience has improved these problems are easily overcome.⁹⁷

⁹⁷ Russom, P. (2013). Managing big data, TDWI Research, Fourth Quarter.

2.5 Data governance taxonomy: *Cloud vs Non-Cloud*

Noticing a progressive increase in data in all their sectors, companies, in order to get the most from this data, need to be able to create *taxonomies*. Taxonomy is the creation of a data classification, represented by a formal structure of classes or types, organized on the basis of defined parameters. This is used to organize the knowledge of the data so that it is easily traceable. A business taxonomy allows data to be searchable and organizes data so that this is normalized through the organization.

A fundamental aspect of corporate taxonomy is to specify the difference between structured and unstructured data. Structured data is usually stored in a database. Numerical data uses the metadata or attributes as a classification. Unstructured data on the other hand is characterized by notes, texts, and other features that are not part of the metadata.⁹⁸

The development of a corporate taxonomy can lead to numerous benefits for the company. For example, it allows to save costs on paper purchases and shipments, since the data is stored in the same place, and users use that data in that place. Having a single place to look for data, this is easier to find so less time is spent looking for documents, doing work tasks faster. Although, it is necessary to be careful because if the data are lost, they are difficult to recover since being the only place where they have been stored there are no other possibilities to find that single datum.

To construct a business taxonomy the typology of approach used, Dimensional based approach, must be defined. This methodology allows to divide the categories of the taxonomy created in multiple areas. This supports the organization in being increasingly flexible and in terms of *Data Governance* this helps to manage the data more effectively.

To understand the aspects of data governance and cloud services, we use analytical theory to study the different dimensions of data governance. This is possible by following three steps. Firstly, the state of data governance of the company is understood; secondly, the size or characteristics of data governance and cloud computing is analysed; thirdly, the dimensions for data governance and cloud

⁹⁸ Wibbenmeyer, K. (2015). How to successfully implement a corporate taxonomy, *Global Journal of Computer Science and Technology*, 15, 5.

computing are developed, based on the results obtained from the previous analyses carried out in the previous steps.⁹⁹

To carry out a detailed analysis on how to create a taxonomy, it is key to observe a taxonomy for traditional data governance systems, and then for cloud data governance systems.

Traditional Data Governance

Analysing traditional data governance, it can be divided into three main categories: people and organizational bodies, policy and technology.

- *People and organizational bodies*: data governance influences the decisions and actions of stakeholders within the company. In the traditional data governance, people and organizational bodies play a fundamental role in the development of data governance systems, especially as people have multiple tasks, such as authority, data stewardship, business rules, collaborations, accountability and culture attitude.¹⁰⁰
- *Policy and Process*: a fundamental aspect for the development of traditional governance systems; Policies are a set of measurable rules for a set of governance functions that ensure benefits to business processes. Policies and processes in the development of traditional systems include principles, policies, standards, and processes.¹⁰¹
- *Technology*: this aspect is crucial for the development of traditional systems; through technology automation can be assured and reinforcement and better control of governance policies.¹⁰²

⁹⁹ Hameed, K., Al Ruithe, M., Benkhelifa, E., (2018). Data governance taxonomy: cloud versus non cloud, Sustainability, 10, 95.

¹⁰⁰ *Ibidem.*

¹⁰¹ *Ibidem.*

¹⁰² *Ibidem.*

Cloud Data Governance

The development of these cloud systems has not yet been adopted by a large number of companies mainly due to security issues related to legal issues. Undoubtedly, however, in the coming years companies will increasingly move towards systems of this type. Governance in the cloud requires an understanding, moderation and regulation of the relationships that exist between the different stakeholders in terms of roles and responsibilities. Data governance therefore is used to classify and assign responsibilities between the different actors.

To create a cloud taxonomy the different categories that classify data within the system must be defined. In this case we have many more sub-categories of traditional systems. These are: Data Governance Structure, Policy and Process, Cloud Deployments Model, Service Delivery Model, Cloud Actors, Organizational, Technological, Service Level Agreement, Monitor Matrix, Legal Context.

- *Data governance structure*: this category is very important as it ensures that the roles and responsibilities within the company at any given level are clearly defined. Doing a good job at this stage can help the whole organization to operate effectively and efficiently. the main roles defined at this level are: *Executive sponsorship, data management committee, compliance committee, data stewardship team, cloud manager, cloud provider member, IT member, legal member.* ¹⁰³
- *Policy and Process*: making sure to establish the right policies and the right processes, will lead the entire organization to operate in an accurate, precise, available manner and also guarantees security to the data managed. Data governance functions can help in the corporate decision-making process so it must be ensured that the established policies and processes are correctly designed.¹⁰⁴
- *Cloud deployments model*: fundamental aspect in the development of cloud governance; the implementation models can be private, public, community

¹⁰³ Hameed, K., Al Ruithe, M., Benkhelifa, E., (2018). Data governance taxonomy: cloud versus non cloud, Sustainability, 10, 95.

¹⁰⁴ *Ibidem.*

hybrids. In choosing the model, however, the risks and complexities deriving from each model must be taken into consideration.¹⁰⁵

- *Cloud service delivery model*: this type of service can be divided into three different models: Software as a service, platform as a service and infrastructure as a service. Depending on the model there will be a different level of control on the data held. In choosing the model, the policies and rules to be added to the model must be taken into account in order to reinforce it.¹⁰⁶
- *Cloud actors*: this category refers to those who work within the company, in particular those who participate in processes, transactions, and operate in cloud computing environment; each participant has certain roles and responsibilities that must be respected.¹⁰⁷
- *Organizational context*: this category includes all organizational aspects of the company. To implement cloud data governance, we need a change at the management level, and an increase in the participation and commitment of all staff. Moreover, the support that top management gives to the organization is necessary for implementation. For the organizational context we mean the definition of all the internal factors of the company considering the risk contribution they can cause. There are three sub-categories that can occur: *Strategical, tactical and operational*.¹⁰⁸
- *Technical Context*: technology can play a key role in the success of cloud governance strategies. By technical context we mean the evaluation of the problems that incur in unravelling a joint technology rather than another. Being superficial at this stage, therefore not adopting the right technology, can lead to failure of the entire governance strategy.¹⁰⁹
- *Service level agreement*: at this point all the agreements between the various actors involved are defined, then the consumers and the service provider. The agreement established by the type of service provided, the methods and

¹⁰⁵Hameed, K., Al Ruithe, M., Benkhelifa, E., (2018). Data governance taxonomy: cloud versus non cloud, Sustainability, 10, 95.

¹⁰⁶*Ibidem*.

¹⁰⁷*Ibidem*.

¹⁰⁸*Ibidem*.

¹⁰⁹*Ibidem*.

consequences for the failure to pay. The agreement is compiled through the insurance of the rules and the rules to be respected by both parties.¹¹⁰

- *Monitor matrix*: is the implementation of the authority of control and sharing in the decision-making process in the management of company assets. Adding aspects of control and monitoring to existing ones reinforces the data quality processes. The matrix data governance monitor for cloud computing services includes cloud control matrix, KPIs and monitoring tools.¹¹¹
- *Legal context*: determines all internal and external laws and regulations relating to data, which could limit the benefits of cloud technology.¹¹²

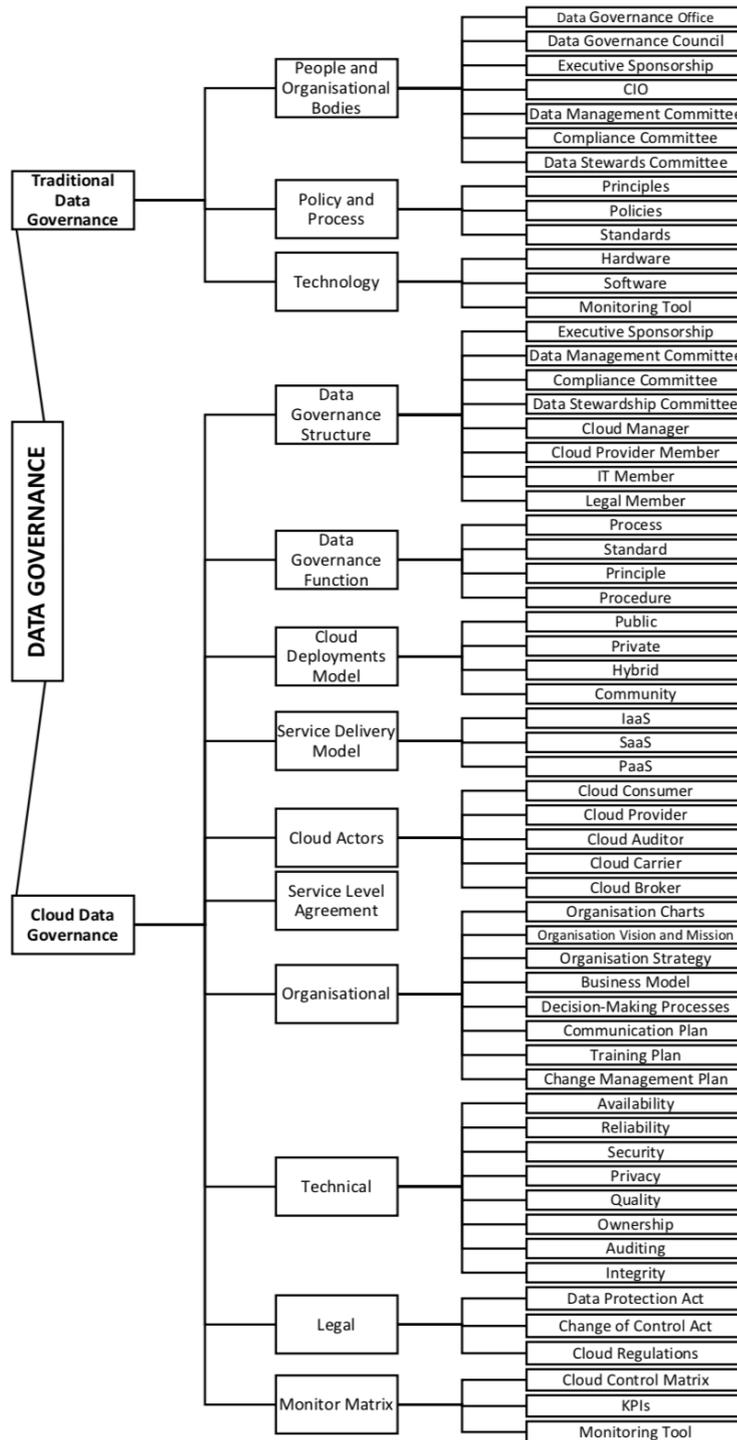
The figure below shows all the elements described above in a single framework that includes both the characteristics of the development of traditional governance programs and the more modern cloud governance programs.

¹¹⁰Hameed, K., Al Ruithe, M., Benkhelifa, E., (2018). Data governance taxonomy: cloud versus non cloud, Sustainability, 10, 95.

¹¹¹*Ibidem.*

¹¹²*Ibidem.*

Fig. 10 Data governance Taxonomy (Cloud and Non-Cloud)



Source: Hameed, K., Al Ruithe, M., Benkhelifa, E., (2018). Data governance taxonomy: cloud versus non cloud, Sustainability, 10, 95.

2.6 Benefits, risks and criticalities of *data governance*

The implementation and development of data governance strategies can bring numerous benefits to different disadvantages, or at least can create risky situations. These strategies, as extensively described above, are necessary to align the company's infrastructure with its objectives, trying to provide the material needed to make the right decisions.

Describing the critical success factors that the development of this type of strategy can bring, we must mention the major sources of oversight on the part of the executives, and how these strategies succeed in preventing it.

The first type of benefit is found in *accountability* and *strategic accountability*. The executive must in fact have the possibility to personally manage the governance processes; to be able to do so the roles and responsibilities must be explicitly clarified to all the actors involved in the system.

The second proven benefit is that found in the definition of the *standards*; this process is fundamental for a company so that it is "fit for purpose".

The third type of benefit is the *managerial blind spot*: it is necessary to align the technology, the processes and the organizational bodies specific to the data, with the objectives of the company.

Moving to *complexity*, the use of data governance strategies manages to contain the complexity encountered in data management.

Cross divisional issues refer to that series of benefits that lead to designing a structure in a way that it will include the participation of each company level.

Metric benefits are found if the right valuation methods are used within the company. If a company shares its data with partners, benefits can be met if the different parts are made responsible for their actions, so the data produced must be reliable and therefore be beneficial.

Choose the right *control points*, so as to make sure that everything is done in the most appropriate manner.

Monitor compliance, so that it is followed by employees according to the established plans.

Finally make sure that everyone has the right level of *training*, so that the quality of the data can be adequately achieved.¹¹³

The risks that can be encountered during the implementation of governance strategies derive from different categories: transactional, reputational, strategic and legal.

- *Transactional risk*: also known as operational risks, these risks are the easiest to find, they can derive from fraud, errors or incompetence in the supply of products or services. This type of risk can affect other of the business' processes such as credit, interest rate, liquidity and price compliance.¹¹⁴
- *Reputational risks*: this category risks if fed into errors, delays and leaks in the information security system, and the company's reputation but also the safety of its users can be compromised.
- *Strategic risks*: the lack of precision or incompleteness of the information can bring strategic risks to the company as it could compromise the entire corporate decision-making process. this type of risk is the one that should be avoided the most.
- *Compliance risks*: the inaccuracy of the data or the lack of precision in terms of time related to customer compliance can expose the organization to serious problems relating to lawsuits or substantial compensation, so when confidential data is managed it must be done with extreme precision.¹¹⁵

2.6.1 Data protection and security

As presented in the previous paragraphs, the protection that a company creates around its data becomes a strategic asset for every business. however, securing data is a more complex and articulated process, as there are constant threats that can plague corporate data from multiple fronts.

¹¹³ Cheong, L., Chang, V. (2007). The need for data governance: a case study, Australasian Conference on Information System, December.

¹¹⁴ Berson, A., Dubov, L. (2011). Master data management and data governance, McGraw Hill Education, New York.

¹¹⁵ *Ibidem*.

Berson A. and Dubov L. define data security "the protection of data from either accidental or unauthorized viewing, modification, destruction, duplication, or disclosure during input, processing, storage, transmission, or output operations".¹¹⁶

Therefore, the main objectives of data security are to ensure *data privacy* and *confidentiality*, ensuring that information or data can be used only by those who actually have access to it. *Data integrity* guarantees the quality of the data and allows this to be modified or influenced by those who have authorization. Finally, we have the *data availability*, the assurance that the data is accessible to those who need it when they need it.

The continuous evolution of the techniques of damage or deprivation of data by agents external to the companies, obliges them to continually update the data protection techniques; furthermore, the continuous increase in data volume requires higher protection. This is the reason new approaches to protection can ensure the security of customers' and company's information.

Traditional protection systems have specific types of requirements: authentication, authorization, confidentiality, integrity, verification and nonrepudiation, audit and accountability, availability and security management.

- *Authentication*: this type of request serves to ensure that the user of the data is recognized or identified in order to verify that he is officially authorised to use the data.
- *Authorization*: this request serves to ensure that the person using the data can actually interface with that data, or that he has access to it; to reinforce this category, technologies can play a key role.
- *Confidentiality*: is a mechanism that allows to define rules and processes that can protect certain information from those who are not authorized to access it.
- *Integrity*: is the process that ensures that a datum's information is protected during its life course, and that it is not changed, unless the change is authorized. The integrity is pursued when a message arriving at its destination is exactly the

¹¹⁶ Berson, A., Dubov, L. (2011). Master data management and data governance, McGraw Hill Education, New York.

same as when it was sent; integrity ensures that there are no malicious data changes.

- *Verification and nonrepudiation*: this refers to the legal aspects that ensure a continuous verification of the actions of those who deal with data undertake.
- *Auditing and accountability*: this category defines the processes through which data and information are reviewed by those involved in these actions, so that the authentication and authorisation mechanisms produce the results desired by the security policies created by the company.
- *Availability*: this category ensures that data is provided promptly to those who request it.
- *Security management*: the security management deals with user's administration which is the process that defines, creates, maintains and chases the authorizations of users with their respective resources; security management also deals with key management which are all the processes that establish, generate and distribute the methods that ensure solutions in terms of security.¹¹⁷

One of the new systems of protection and data, which operates on a wider range, also on company boundaries, is the so-called *Layered security framework*. Berson A. and Dubov L. define this new framework as: “security domains can be organized into a layered framework that looks at security from “outside in”: perimeter security, network security, platform (host) security, application security and user security”¹¹⁸

This model indicates the areas of the company that must be protected.

- *Perimeter security*: this deals with the threats deriving from company boundaries, therefore from the outside. As the name suggests, the security of the perimeter must deal with authorizations, authentications, and control of access to resources; the main technology used is firewalls.
- *Network security*: this deals with authenticating the users accessing the network, the resources available and the protection of information passing through the

¹¹⁷ Berson, A., Dubov, L. (2011). Master data management and data governance, McGraw Hill Education, New York.

¹¹⁸ *Ibidem*.

network. The main technologies used, the tools, are PKIs (public key infrastructure) and virtual private networks (VPN).

- *Platform (host) security*: this deals with the dangers of the attacks that the current platform can suffer from the outside and make it therefore vulnerable. There must be the aforementioned forms of authentication, authorization and access control are used, in order to limit access to intruders and potentially dangerous persons.
- *Application, data and user security*: this is the fundamental category of the whole security system; the main objective of every ill-intentioned person is to access personal information concerning users or the company and use them against it. Therefore, the security systems gave assurances that there are 3As (authentication, authorization, and administration), encryption systems, digital signatures, impartiality, integrity, privacy, and virus protection.¹¹⁹

¹¹⁹ Berson, A., Dubov, L. (2011). Master data management and data governance, McGraw Hill Education, New York.

Chapter 3 – Data governance: key role for corporate strategic choices

3.1 Data Governance as a strategic business asset

In today's society, information has become extremely important. In an appraised economy, information is the most valuable asset that a company can have, and therefore knowing how to correctly manage and govern information can provide a substantial competitive advantage to those who have the right information. As previously outlined, data management, when carried out correctly, is able to guarantee that the data is accurate, properly shared and protected. As Eric Sweden states, data is essential for creating information; information is essential to create knowledge; knowledge is finally needed to create the wisdom that helps in the decision-making process. We can therefore affirm that the quality of the data and its relative information influence the quality of the decisions made.

As aforementioned, the key aspects of data governance can be of three types: they can relate to strategic planning, control mechanisms and the metric key.

Strategic Planning

Strategic planning regards all functions that determine the need for the company data and the strategies to adopt in the use of this data. At this stage, the level of maturity reached by the company in terms of data management is analysed and certified. The company will then try to establish the future peculiarities of data management, develop and subsequently approve the policies, standards and procedures concerning the data. Consequently, the projects and services are planned and finally the value to be associated with the data as a company asset and the relative implementation costs is established.¹²⁰

Ongoing Control

The control mechanisms, include the coordination of data governance activities, managing and trying to solve the problems related to the management of the aforementioned data, monitoring and reinforcing compliance with the policies,

¹²⁰ AA.VV. (2011). Enterprise information management: best practices in data governance, Enterprise Architecture.

standards and data architecture, and finally communicating the value of goods related to data¹²¹.

Key Metrics

In this section, the data's value is defined, the cost of the whole data management process; the objectives to be achieved in the short or long term and the number of decisions made are defined. The stewards of the process are also defined, and consequently all the related staff as well.¹²²

Companies and government authorities are driven to focus on the development and implementation of strategies for an effective data management because at least 25% of the fortune 1000 companies still continue to handle the data incorrectly. Producing incorrect information subsequently leads to making wrong decisions.

Both on a governmental and on a corporate level the need for accurate and precise information continues to grow and it is precisely in this area that companies are trying to gain an advantage over their competitors. With the right data management there can be an improvement of the entire company performance, at all its levels. The largest creation of value is obtained through the improvement of the cross-organizational performance that is achieved by improving the flow of information shared between the different company levels.

To understand how companies are interfacing with this new type of asset or resource, the CDI institute carried out a study that analyses the companies belonging to the Fortune 5000 category, to understand how companies are investing in data governance. Many of these companies in terms of budget, staff and structure could be compared to the top 500 of the fortune list. The study provides several explanations on the motivations to develop this type of data governance strategies.

Given the importance that is given to data and information, companies should use this information in all respects as a very valuable resource for the company. Although, depending on the value and therefore on the risk attributable to information, the resource must be properly managed and protected; the more value the information has,

¹²¹ AA.VV. (2011). Enterprise information management: best practices in data governance, Enterprise Architecture.

¹²² *Ibidem*.

the higher is the need to invest in security, protection and its use. To be considered reliable, as explained in the previous chapters, data must have high quality standards; hence data and information must be treated as financial or human resources. The information must also be integrated into all company levels in order to facilitate connections between the different levels and units of the company. Finally, to guarantee an effective decision-making process the data resources must be managed correctly in operational terms in a company.¹²³ These circumstances, which also include the respective commercial and technological innovations, push companies to have to foresee or anticipate them so as to be able to set the greatest advantage with respect to their competitors; this creates a further motivation to improve the interaction between the different units of your business, and this improvement can only be achieved if you rely on accurate, precise, timely, integrated and available information.¹²⁴

The evaluation of information thus becomes a fundamental aspect for the company. To ensure that the information is correctly evaluated, the acquisition cost of the information must be taken into consideration, the degree of sharing that the information possesses, the type of user who has the information available, the context within which information is entered and finally the added value that the company obtains from the use of that information. Needless to say, that not all information, thus not all data has the same associated value, accordingly, they cannot all share the same risk. It is particularly simple to analyse the level of risk possessed by data or information for a specific company. The risk is calculated by multiplying the probability that an unwanted event occurs, due to the impact that this possible event would have on the company. To point out the worst-case scenario, where the risk would be higher, is when it has a high probability of occurrence combined with a high impact for society. There are circumstances where the information turns out to be void for the company, in that case there is a zero tolerance for the loss of the data, and circumstances where the data or information is public and does not reflect a risk for the company.¹²⁵

Companies should have a risk portfolio in which to classify all data. This risk portfolio should catalogue the data by analysing the value deriving from the information; its

¹²³ Sweden, E. (2008). Data governance. Managing information as an enterprise asset, Nascio Governance Series.

¹²⁴ *Ibidem.*

¹²⁵ *Ibidem.*

value for the company, its potential value for criminals; the value that follows from the change of the information and the secondary effects deriving from this change for the company and obviously the necessary measures to adequately protect the data.

Therefore, we can affirm that before being protected the data and the information must be classified to be able to evaluate the most suitable method of protection.

However, the implementation of these strategies forces companies interested in this kind of asset to "overcome" certain barriers. The major barriers that companies encounter are:

- Determining the rules and requirements; interpreting and understanding the rules concerning data sources
- Gaining agreement of all parties regarding policies.
- Developing new tools and software to enable data governance
- The cost of implementing policies
- Incompatible systems
- Competing priorities within the organization
- Getting management to understand what is necessary
- Building the project process¹²⁶

As previously stated, the quality of data directly affects the quality and effectiveness of future decisions, hence highlighting the impact that the company will have on its business. The methodologies that ensure the quality of the data are essential to ensure that the data is available promptly for those in charge of making decisions.

To ensure that data processing is regularly improving to provide companies with increasingly effective and reliable resources, new frameworks are being developed, which form *data governance 2.0*. The definition of this new approach provided by Baker, S and Sjoberg, P is: " agile approach to data governance focused on just enough controls for managing risk, which enables broader and more insightful use of data required by the evolving needs of an expanding business ecosystem". To ensure the effectiveness of data as a corporate asset, four types of possible data governance systems are identified.

¹²⁶ Sweden, E. (2008). Data governance. Managing information as an enterprise asset, Nascio Governance Series.

1. *Systems of record*: here we have the traditional databases that handle transaction processing and analytical processing; the main focus of these systems is data quality and compliance.
2. *System of engagement*: in this system there is a new management and a new coincidence of the customer, employee and partner; in this new framework, data management must focus on the personalization of data taking into account their confidentiality¹²⁷
3. *System of automation*: This is driven by the integration of the Internet of Things (IoT), data on operational technology (OT), correlation of events, analysis and artificial intelligence (AI); in this type of framework, governance must not focus on the context of data m especially on the origin and quality of them.
4. *System of design*: These are creativity and innovation systems in which stakeholders can be product designers or researchers. Data governance in this case may involve social, legal, ethical and moral issues.¹²⁸

3.1.1 Impact of Data Governance in the company's strategic choices

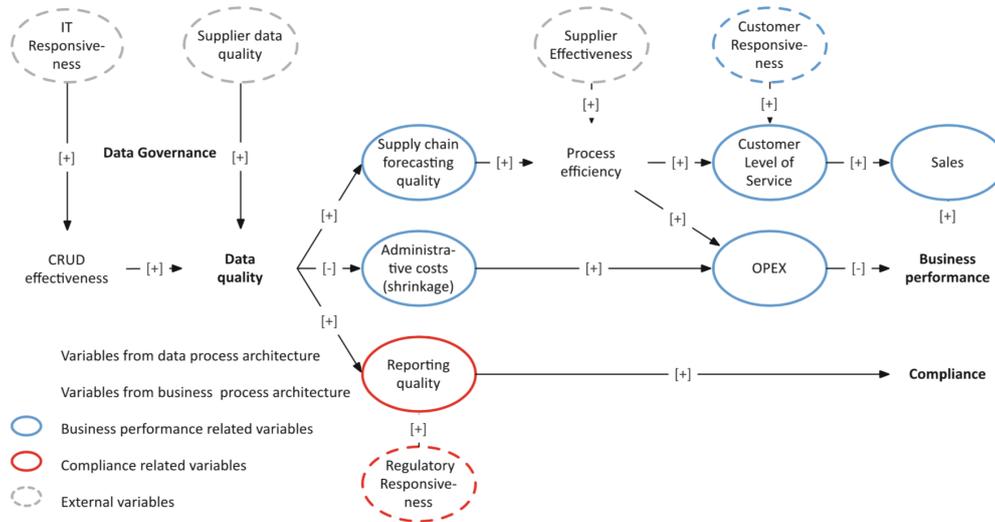
After in depth analysis in the previous paragraphs, we have established that data is considered one of the most valuable assets within companies. The value attributed to data derives from two main applications: the measurement of business performance and compliance reporting. First of all, the increase in complexity and above all in the range of action of the different businesses, requires the support of highly accurate and precise data; secondly, data within the companies is used to carry out different types of reports. However, to fully understand how data governance strategies can be useful to a company at a practical level, we can use the model provided by Martijin. N et al. Below is the graphic representation of the model, which will subsequently help us to understand the different influences of the government of the general performance data.¹²⁹

¹²⁷ Baker, S., Sjoberg, P. (2018). *Intelligent data governance*, John Wiley & Sons, Hoboken.

¹²⁸ *Ibidem*.

¹²⁹ Martjin, N. et al. (2015). Determining the effects of data governance on the performance and the compliance of enterprises in the logistics and retail sector, *International Federation for Information Processing*, 37, 7.

Fig. 11 effects of Data Governance programs



Source: Martjin, N. et al. (2015). Determining the effects of data governance on the performance and the compliance of enterprises in the logistics and retail sector, *International Federation for Information Processing*, 37, 7.

In order to better understand the different relationships with the figure - [+] -> positive reactions are indicated, while with - [-] -> negative reactions.

Starting the process of analysis from the upper left we can see how the strategies of data governance can positively influence the denominate CRUD actions, which then subsequently improve the quality of the data obtained. Obviously, the initial "cleaning" of the data, fundamental in every Data Governance strategy, helps the whole process. The consultation of these quality data also benefits the communication between the various departments of the company, which consequently leads to a better reactivity of the IT activities. The reinforcement of the standards also reduces the possibility of making mistakes, and the higher the standards, the greater the possibility of measuring their quality. Also, the sources outside the company play a key role, as improving the quality of the information flow shared with its partners in the supply chain will consequently improve the quality of the data deriving from the aforementioned suppliers. Improving data quality leads to improved forecasts made along the supply chain, so if the forecasts are not reliable, the whole process will be less effective. Improving the quality of data, however, the costs of management and administrative expenses are reduced. According to a research carried out by (GSI 2011) data of poor-quality leads to a high expenditure of administrative resources; these costs are mainly

due to the difference that forms between what is shipped by the suppliers and what is actually sold to the end customer. In addition to being efficient in the first part of the process, it also reduces operating expenses, which consist of all those costs associated with supply chain operations such as transport and transaction costs. The administrative costs have a big impact on the operating activities, so it is preferred to limit these expenses in favour of a more fluid operating process. If the grade of data is improved or in the case it remains reliable, the quality of the forecasts on which to entrust the full production is also improved; furthermore, the quality of the reports would be made accordingly, enhancing and increasing sales and the general performance of the company. In the lower part of the figure, the improvement of the data quality allows to increase the regulatory compliance.¹³⁰

The statements given above explain that the effects of data governance are positive and also tangible. The first visible effects will be of a qualitative type, but with improving knowledge and experience, quantitative measures will also be noticeable in the company. The model always depends on what type of company it is applied; many of these variables are carefully defined and constantly evaluated, such as OPEX, sales, customer service level; the others can be defined once a good operation has been found.¹³¹

To better outline how data governance and data quality can impact and improve companies that implement it, I will provide a practical example. This is a practical case that deals with an international company in the logistics and shipments business.¹³²

As modern society and our daily habits prove to us, the shipping and logistics market has a high rate of competition, therefore the company is forced to rely heavily on the quality of its suppliers' data. The quality of the data is fundamental due to the efficiency of the processes of the company in question, since on the one hand it must comply with the regulatory requirements of the various countries in which it operates, and on the other hand, the high competitiveness of the market in which it operates requires a supply chain increasingly efficient. For this reason, the company has previously

¹³⁰ Martjin, N. et al. (2015). Determining the effects of data governance on the performance and the compliance of enterprises in the logistics and retail sector, International Federation for Information Processing, 37, 7.

¹³¹ *Ibidem*.

¹³² Martijn, N., Jonker, R. (2015). The effects of data government in theory and practice, Compact, I.

encountered problems regarding customs and with a general reduction in its competitiveness, all due to poor data quality and consequent poor reliability. All this has led to a reduction in customer satisfaction that the company cannot afford if it wants to remain competitive. To try to resolve this crisis, the company has decided to implement data governance strategies, specifying within the company, the different roles and different responsibilities. The entire strategy was focused on reducing the costs deriving from the supply chain to guarantee the company's advantage over the competitors. All this proved to be absolutely effective for the company, which has estimated that the return on investments, resulting from the implementation of this strategy, is of around 3.5 million euros. The main reason is that the company is in a situation of control over the quality of the corporate data since it is able not only to guarantee the availability of the data but also to guarantee its correctness. All this has led the company to operate in a more efficient supply chain that has fewer compliance problems, which in general translates into better customer satisfaction. All this is done thanks to the implementation of data management strategies.¹³³

3.1.2 Valued-based *Data Governance*

The common tendency for all companies is to interface with an always increasing data. This is mainly due to the fact that the use of data and consequently the management and management of this data goes to align the company objectives with its own resources, which in these cases are precisely the data. In fact, data is increasingly seen as a strategic asset for companies.

Data governance provides the necessary supervision to the company and a method to set goals. Regularly seeking higher quality data, companies invest in data quality so they have insurance against the various risks that could occur if the data were not shaped appropriately for the company.

The main necessity is to have a profound knowledge of how to define the attribute value to one's data, since it constitutes a resource. It is therefore possible to define different criteria in order to specify the connection between data value, metadata, business processes and data quality tools.

¹³³ Martijn, N., Jonker, R. (2015). The effects of data government in theory and practice, Compact, I.

There are innumerable studies dealing with data quality, data assets, information data assessment; however, the main problem is that none of the aforementioned studies provides a formal model on how to actually evaluate the data, so as to select only the one best suited to the company. Moody and Walsh define seven "laws" that help define the behaviour of the people in charge of data selection and the relationship between the value of the data and the value they can bring into the company; unfortunately, even this list of rules does not provide any concrete measurement techniques. Moody defines three methods of data evaluation: Utility, Market Price and cost. After having analysed all three of them, he defines the first model, that is the one that determines the usefulness of the data, in theory the best, but impractical, since it is difficult to define the usefulness of a data a priori; therefore finds that in the cost-based model, that is the one that takes into account the collection price the most effective method.¹³⁴

The biggest challenge for companies is to define quality standards a priori for their data. The formal models of data quality allow to better understand, to better evaluate, and above all to easily predict the connection between the production of the datasets, the costs, the utility obtained, the methods of use, quality metrics , metadata, topic domains, workflows, provenance, and value would enable new approaches called value-driven in data governance strategies. In addition, machine processable models form the basis for the transfer of knowledge of data quality and value to the entire business ecosystem, enabling automation of data quality management, such as the definition of metrics, tools for selection, and the priority of tasks. These models also provide the possibility to develop new intelligent applications for data quality such as data asset value profiling or improvement, and decisions supporting the design of qualitative processes.

However, the tendency is to move from a more human control to an automated control for qualitative monitoring, for analysis and reinforcement that maximizes the value of the business and minimizes human effort. This type of methodology is applicable to those business units where human oversights must not be eliminated since they are fundamental for the company.

¹³⁴ Brennan, R. (2017). Challenges for value driven semantic data quality management, Proceedings of the 19th International Conference on Enterprise Information Systems, 1.

Therefore, the presence of formal hubs that put in connection the quality and the value of the data allow to have the presence of automated decision-making processes, or people supporting the decision-making process as they possess a deep knowledge of the data they deal with, such as the data steward. Consequently, curves can be developed that monitor the performance of the technology for the processes and the quality of the data, which together with the model of knowledge of the data help to foresee and plan the quality of the data themselves. The effectiveness of the system will consequently be measured by statistical analysis compared to historical models of data quality.¹³⁵

To ensure that the data is qualitatively traceable, and in a way that positively influences the decision-making process, among the quality standards we can mention: Completeness, Uniqueness, Timeliness, Validity, Accuracy and Consistency.

- *Completeness*: measures the percentage in which the data field contains the expected data. For example, if a data field in which the date of birth should be entered is empty or contains letters, we can assume and confirm that the data is not complete.
- *Uniqueness*: measures the content of the different data sources with one another to understand the repetitiveness of the data. Repetitiveness is based on the duplication of the data or on the fact that the data becomes obsolete.
- *Timeliness*: measures the time which a data remains reliable and valid before it is "exceeded" by "younger" or otherwise updated data. This feature plays a key role specifically in highly transactional workloads.
- *Validity*: verifies that the data are compliant with the standards established by the company, and therefore usable.
- *Accuracy*: assesses whether the data relates to or uses taxonomies in the established reference data or whether the data correctly reflects what it identifies.
- *Consistency*: evaluates how data is standardised in form and content and also to what extent the data aligns with preconceived data models that are expected to be seen or automated in the business.¹³⁶

¹³⁵ Brennan, R. (2017). Challenges for value driven semantic data quality management, Proceedings of the 19th International Conference on Enterprise Information Systems, 1.

¹³⁶ Baker, S., Sjoberg, P. (2018). Intelligent data governance, John Wiley & Sons, Hoboken.

The above list as it is, can be applied to the strategies of governance, or it can be extended to include specific metrics of governance of the data related to access, to the authorized use of data, to their degree of mobility within the company, to the degree of recovery and their safety characteristics.¹³⁷

3.2 Best practices

Not all the efforts that are employed in the implementation of data governance strategies lead to the desired results; it occurs frequently that the hoped results of effectiveness and efficiency are not reached. There are many obstacles that tend to undermine the effectiveness of these programs. These obstacles can come from multiple fronts: cultural, political and organizational. However, there are certain ways that can help companies succeed in implementing governance programs.

Take a holistic approach but start small

Since the process of data governance is an iterative process, it must necessarily start on a small scale and then increase in tension until it touches all company areas. In order for it to be effective it is convenient for it to start from people, politics and culture, and then move on to the processes of governance and data management, concluding with technology. Multiple steps are required to gradually increase the level of knowledge and maturity. all this aims to balance the company's strategic objectives and the relative tactical compromises to ensure that the company is moving in the right direction.¹³⁸

Obtain executive sponsorship

Given that data governance strategies involve significant changes both at a cultural level and at a behavioural level, as well as costly investments in technology, they require a high level of support at the executive level without which nothing would

¹³⁷ Baker, S., Sjoberg, P. (2018). Intelligent data governance, John Wiley & Sons, Hoboken.

¹³⁸ AA.VV. (2011). Enterprise information management: best practices in data governance, Enterprise Architecture, May.

change. It is therefore necessary to analyse the interested parties and align, involve, the fundamental decision makers of the company that represent the functional areas and the main lines of activity that have a strong influence and determine the majority of decision-making power. By obtaining this type of support, you will have a greater chance of success.¹³⁹

Define data stewardship during early stage

Speaking of who will personally administer the process, they are called data stewards, their main responsibility is to ensure effective and efficient control and use of data resources. The best data administrators cannot be created, but must be found. Therefore, a large amount of time must be budgeted to create the right team, thus identifying the most capable people, hence forming experts in the field for all business areas.

There are different opinions for the official recognition of this position, and the answer to these uncertainties is: it depends. It mainly depends on the phase in which the governance project is developed and also depends on the cultural and political context in which its organization is inserted. The only fundamental need that is required is that the definition of the role is included in the descriptions of the different tasks of these persons and moreover a correct allocation of time to this administrative work must be indicated.¹⁴⁰

Establish quantifiable benefits by building a business case

The main benefits deriving from the implementation of certain governance strategies show their effects over the long term. Unfortunately, however, some of the effects may or may not be immediately visible. For this reason, it is not always easy to obtain funding from third party stakeholders who cannot be shown tangible results. However, one way to deliver concrete results is to focus on the relationship of the key data elements and the related supported business processes. Calculate the cost of managing these data items through repeated and duplicate manual integration and

¹³⁹ AA.VV. (2011). Enterprise information management: best practices in data governance, Enterprise Architecture, May.

¹⁴⁰ *Ibidem*.

validation. Moreover, the risk that the company runs using these data must be quantified as they may become unavailable or incorrect, which results in missing transactions or loss of customers. It is therefore necessary to identify the opportunities that high quality data offer in terms of revenue generation and improvement, managing to create better customer service through up-sale and cross-sale. In conclusion it is necessary to be able to build a business case to articulate and highlight the tangible benefits deriving from the project or strategy, which are then necessary to obtain financing and support from third parties.¹⁴¹

Establish, collect, and report on metrics to measure the progress

As a starting point, strategic combinations must be chosen which have led to success and long-term tactical improvements. It is therefore necessary to measure the immediate returns of the fastest victories in order to consequently obtain positive feedback in order to support the commitment shown and in conclusion obtain even more support. The measures must be planned at the beginning of the project, to subsequently concentrate on the quantitative parameters to support the objectives of the project and the general strategic program. The metric indexes must analyse the company values and some metrics must include the data values, the relative management costs, both those prior to the process, and subsequent ones. Finally, the number of decisions taken, and the maturity of the governance process must be evaluated. An effective way to automate the communication of the relative progress is to create a dashboard for the main key performance indexes (KPIs).¹⁴²

Link and build in incentives to award and re-enforce participation

The worst aspect of data governance, and the most difficult, is the assurance that the project has a continuous duration, and that it is not implemented for a specific and limited period. Those who manage the project, Data owner and Data Steward, are usually assigned a high number of operational tasks and the business units vary significantly on the different allocation of resources. The positive implementation of

¹⁴¹ AA.VV. (2011). Enterprise information management: best practices in data governance, Enterprise Architecture, May.

¹⁴² *Ibidem*.

this specific type of governance program depends on the level and priority commitment of all the various corporate functions. To build an efficient system based on incentives that employees can achieve by putting more effort into the whole system must be supported by all the different company levels. However, it is fundamental to consider that the incentives must not be based solely on financial criteria, since they would not be effective. Formal and informal incentives have great potential to bring out the greatest commitment from employees, instilling enthusiasm and promoting cultural changes within the company.¹⁴³

The main benefits related to the decision-making process that the organization can obtain, include the identification of new opportunities to interact with the customer and the supply channels. The achievement of a superior performance in terms of efficiency through the different business processes. An improved reporting, business intelligence and data analysis capabilities. Finally, the optimisation of the manual effort required to manage and use the entire company's data.

In conclude the analysis of the best practices to implement, in order to have a successful program, companies must implement a holistic type of data management strategy. The right management model in fact helps the organization in multiple fundamental activities in all its businesses: from reporting, cross-selling and up-selling to decision-making and compliance.¹⁴⁴ Moreover, the chances of success of a Data Governance initiative increase even highly in a business if:

1. the company considers the Data Governance program as a program within their basic structure rather than as a project for its own sake or even outside the company.
2. focus the attention mainly on the return on investment (ROI) right away, or from the formulation of the business case to the post implementation.
3. to constantly involve the executive and the top management of the company.
4. Provide an organizational structure to address governance problems at regular intervals¹⁴⁵

¹⁴³AA.VV. (2010). Best practices for a successful MDM implementation, White Paper, September.

¹⁴⁴ *Ibidem*.

¹⁴⁵ *Ibidem*.

3.3 Practical case: Data Governance in Healthcare

Wanting to observe a practical case to understand the use of data governance strategies within its sector, we can analyse American healthcare system. Every sector, from economic to medical to public, has seen a constant increase of the amount of data received by numerous and various sources; this has been especially notable in the healthcare sector, where large amount of information can make it extremely difficult to manage "vital" patient data. By using data governance and data management strategies, patients' management can be safer and doctors can make more informed and accurate decisions. The number one mistake made using patient data is to apply them as a technological asset rather than as a company asset that can be useful to the company. Using data as a technological good, leads to a series of problems linked to data that can generate: loss of accountability, poor quality, and noncompliance with external adjustments. Furthermore, health data can have problems related to the ownership of data that appears to be fragmented, with little authority and procedures, standards, policies non-existent. Consequently, patient's data can be exposed to anyone, and potentially could lead to wrong decisions being made about people's health, as well as being able to cause loss of money and time.¹⁴⁶

Theoretically speaking, data governance in the healthcare sector has three main related modules: administrative, technical and business.

- *Administrative*: this module includes the formal governors whose role is to define the main data management, specifying goals and policies, solving problems, and assigning roles and responsibilities.
- *Technical*: this module presents the personnel linked to information technology who are responsible for the technical aspects related to data management such as data integration and data modelling standards.
- *Business*: this module represents the stewardship that are responsible for the standardisation and definition of data, and compliance with the policies, rules and processes related to the data.¹⁴⁷

¹⁴⁶ Alofayasan, S. et al. (2014). The significance of data governance in healthcare, Proceedings of the International Conference on Health Informatics.

¹⁴⁷ *Ibidem*.

Healthcare organizations are beginning to interface with data quality challenges, as clinical systems are becoming increasingly complex and above all data structures within hospital companies are increasingly relying on data. As mentioned above, the poor quality of the data has a dramatic negative impact on the efficiency and effectiveness of healthcare organizations, both at the strategic and the operational level. As reported by the Data Warehouse Institute, the problems related to data quality cost, in the American market, in terms of efficiency, six hundred billion dollars a year. Data governance solutions that can help healthcare organizations to solve data quality issues and find the prime way to make organizations efficient in all their aspects.

This deep-seated transformation of the whole organisation brings extreme changes in the strategy that manages healthcare organizations and serves primarily to align people, processes, and technologies more closely to their business strategy. Business transformation is useful for simplifying compliance, improving quality and integrating the business by managing changes in terms of data ownership and use. Business transformations within healthcare organizations are poorly implemented as they are perceived as harmful to the business itself; in fact, a research carried out by RapidBI shows how the failure rate of the business transformation programs for the health sector is very high, between 70 and 80 percentage points, in addition to the fact that the organizations that implement data governance strategies can obtain at maximum 80% success.¹⁴⁸

The benefits obtainable from the implementation of data governance strategies are many and of different types, just like the challenges are. The main challenge related to data governance strategies in the health sector is not so much linked to technologies, but, as seen in the previous chapters, to people and processes. The first challenge concerns political leadership to tabulate the one who will govern the council. Furthermore, the lack of company participation and sponsorship of the executive level appears to be another very determining challenge. Obtaining the commitment from the managers, or at least from the council, is particularly challenging. This is caused by the data management being perceived as an abstract problem, rather than as something concrete that provides a useful tool to the company.

¹⁴⁸ Alofayasan, S. et al. (2014). The significance of data governance in healthcare, Proceedings of the International Conference on Health Informatics.

As reported by a Data Warehousing Institute survey, the second substantial difficulty companies face during the implementation of governance programs is the lack of understanding on the part of the people involved, which includes unsustainable executive sponsorship and lack of business justification. The last barrier that implementers detect, concerns the lack of data ownership, resistance to change and resistance to responsibility.¹⁴⁹

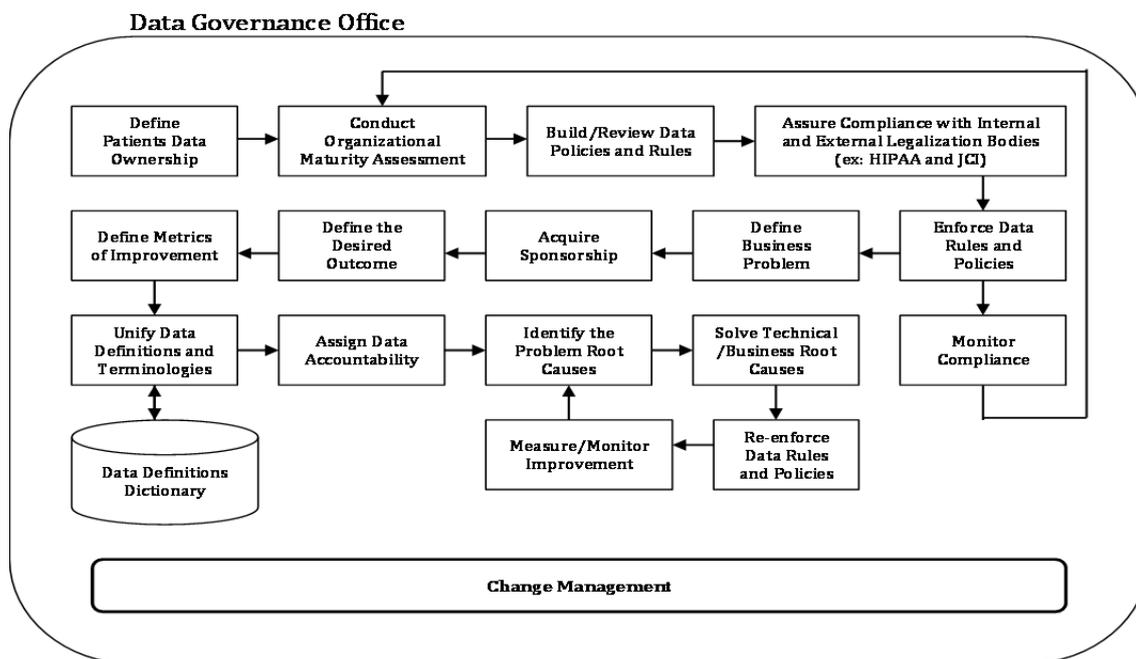
It is, therefore, necessary to create a model that uses and exploits data as a real strategic company asset, as the models already available on the market are not effective for healthcare companies. Furthermore, there is a shortage and deficiency of frameworks capable of efficiently implementing data rules and policies, monitoring and building the services offered and solving problems in a reactive manner. The needs of healthcare companies are profoundly different from those of other sectors, which is why they need a different framework than the others. Unlike in any other field, healthcare companies require compliance with health regulations and standards.

The image below can assist the understanding of how healthcare organizations should manage their data at an operational and organizational level.¹⁵⁰

¹⁴⁹ Alofayasan, S. et al. (2014). The significance of data governance in healthcare, Proceedings of the International Conference on Health Informatics.

¹⁵⁰ *Ibidem*.

Fig. 12 Data governance framework for healthcare.



Source: Aloyfayan, S. et al. (2014). The significance of data governance in healthcare, Proceedings of the International Conference on Health Informatics.

The model shown above begins with defining who owns the totality of the data and then follows with the analysis of the level of maturity of the organization in terms of data management. The main objective is to determine a clear understanding of how the company currently manages the data, in order to specify the gap between the current position and the desired level of efficiency and effectiveness. The number one advice is to examine this analysis annually in order to objectively assess the desired objectives and the current trend. From an operational perspective, those who manage data define the business problems and assign a steward to each problem. The managers and stewards will then define the different metrics of measurement of the various performances to analyse the various improvements over time. Once the problems have been investigated, stewards are assigned responsibility of the data related to each problem in order to avoid the occurrence of the same problems again. They also work to identify and resolve the technical, commercial and operational causes of each problem. The merchandise that is specified at the start in each business unit facilitates the managers to monitor the various improvements relating to the data. Any adaptation,

addition or deletion of regulations or data policies must be forwarded to the board in order to update the framework.¹⁵¹

In order to understand how the aforementioned framework could be implemented, we can focus on the study carried out in central-east America, on a sample of 2200 beds and 12,000 employees, doctors and nurses. The existing data management system is not sufficiently suited to increasing numbers. Annually, in fact, the outpatient visits are generally around 485,000 while the number of admissions, inspected through admissions and resignations, is equal to 40,000. In 2007, this scenario, which includes fifteen main hospitals and administrative systems, was integrated with the electronic medical registration (EMR), which brought the number of data managed, transferred and analysed to grow exponentially; the average daily number of data managed, collected, transferred and examined is equal to 5,832,000. The introduction of the EMR, which can be defined as a data governance system, has demonstrated substantial changes and improvements to the collection and transfer of the manual data, from how it was previously used.

The survey was supported by analysing the ambulatory care services division, consisting of eighteen different clinics working nine hours a day for five days a week. The complexity that this disagreement with the different clinics encounters is due to the fact that they have an enormous volume of patients that passes through every day. Three of these clinics were selected for the survey, based on them having greater flexibility than the others in the acceptance system, as well as having a tendency to be more inclined to the use of EMR systems. The three clinics investigated are: the clinic for employees' health, the clinic for gynaecology and the ophthalmology clinic. Within these three clinics, employees, doctors and nurses were asked to collect the data of the patients they managed manually, as well as compose statistics related to the client, again, manually. In addition, groups of people involved in collecting and analysing data in manual mode were asked to clarify, as well as define, the different formulas for key performance indicators (KPI). Below, the table summarises the different terminologies defined by the groups involved. The data that has been collected, both manually and

¹⁵¹ Alofayasan, S. et al. (2014). The significance of data governance in healthcare, Proceedings of the International Conference on Health Informatics.

electronically via EMR, belongs to a period of time spanning from January 2011 to December 2012.¹⁵²

Fig.13 Table of definitions

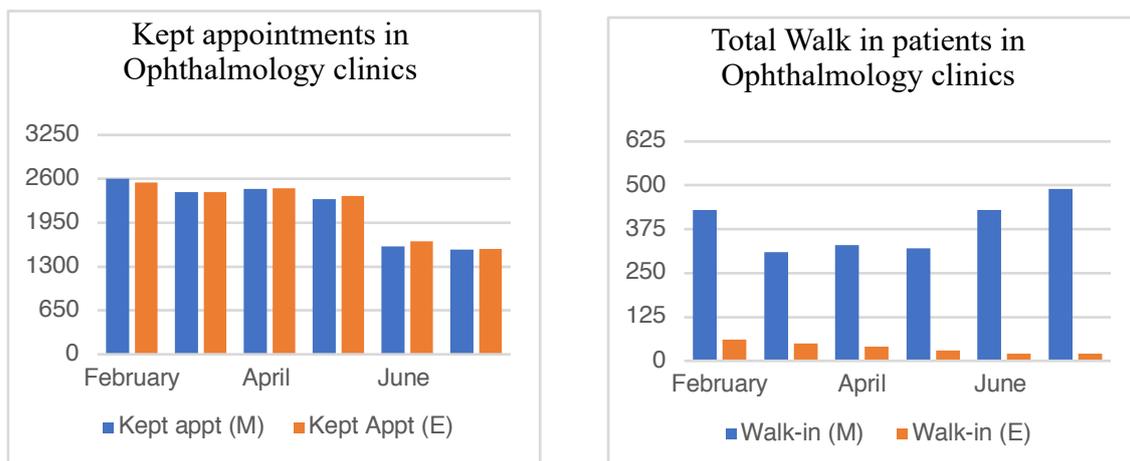
Terms	Definitions
Manual Data	Data that is collected manually by nurses in a clinic using paper and pen
Electronic Data	Data that is captured from an electronic source of data such as Data Warehouse
Nurse-Clinic	Outpatients clinic that is run by a nurse
Physician-Clinic	Outpatients clinic that is run by a physician
Booked Appt.	Patients who did register an appointment for consultation
Kept Appt.	Patients who did attend their registered appointments and had their consultation
Walk-in Patients	Patients who attended a clinic without an appointment being made
No-Show Patients	Patients who did not attend their appointments
<i>Seen</i> Patients	Patients who visited the clinic as a walk-in or with an appointment. The total <i>seen</i> patients are equal to the total walk-in patients plus total kept appointments $Seen\ Patients = (Walk-in + Kept$

Source: Alofayasan, S. et al. (2014). The significance of data governance in healthcare, Proceedings of the International Conference on Health Informatics.

¹⁵²Alofayasan, S. et al. (2014). The significance of data governance in healthcare, Proceedings of the International Conference on Health Informatics.

To simplify the analysis, I will take into account only the data deriving from the ophthalmic clinics; these offer the widest range of services. Ophthalmic clinics' patients may receive treatment ranging from small eye check-ups to major surgical operations. To attend these clinics, it is mandatory for patients to book appointments before any type of visit. Although, if patients arrive without a reservation they can be accepted under certain conditions. As a result of the study we can see, with help from the tables below, that there is an effective discrepancy in the total number of registered patients. The analysis shows that electronic data, collected through the proposed framework, provided 20% fewer registrations than manual ones. The number and patients seen is the sum of those who had an appointment and those without; therefore, the greater discrepancy is due to errors in registering the number of patients without an appointment and those with an appointment. Electronic recordings showed up to 95% less recordings than manual recordings.¹⁵³

The substantial difference that is created between the two types of data collection is the failure to record visits without an appointment in the electronic system, which is a very serious violation of hospital policies, and causes profound discrepancies in the analysis of company data.



Source: Alofayasan, S. et al. (2014). The significance of data governance in healthcare, Proceedings of the International Conference on Health Informatics.

¹⁵³ Alofayasan, S. et al. (2014). The significance of data governance in healthcare, Proceedings of the International Conference on Health Informatics.

After having closely analysed the three ophthalmic clinics, and having noticed that these did not comply with the patient registration policies both manually and electronically, since June 2013 the policies issued by the health system have been enforced, as well as the type of monitoring made to hospital companies has been tightened. Moreover, direct and indirect training initiatives were added to the clinical components, so that they understood and accepted the systemic change. At the end of 2013, data analysis was unexpected, showing a substantial improvement within the three ophthalmic clinics under consideration. Proving this, electronic data provided 21% fewer registrations than those handled and were used to provide up to 95% of patient records without a reservation in less as illustrated above.¹⁵⁴

In conclusion, this analysis assists to perceive how in the health sector, like in any type of sector, there are problems related to the quality of data. It shows how the individual business processes implemented in the clinics influence the results of data. It is sufficient to see that 95% of patients without a reservation have been placed by the EMR system. This is due to the absence of a well-defined business process and to the weak application by the hospital system, and above all, by the lack of adequate training on the electronic system for the employees of the clinics. This affects the quality of data itself, which is negatively influenced by the actions of its employees. Failure by personnel to enter electronic systems is a violation of hospital policy.

The most valuable point that the analysis of the American healthcare presents, is that it needs to find a solid governance framework designed specifically for healthcare needs, therefore one should be created for each different need. The new framework used helped to proactively use data, to resolve data problems quickly and effectively and to monitor violations relating to company policies. Furthermore, this analysis has shown how there are errors in those systems that do not implement data governance. These errors mainly derive from the lack of data-related policies and the lack of adequate responsibility for the data entered in the various management systems. This analysis

¹⁵⁴ Alofayasan, S. et al. (2014). The significance of data governance in healthcare, Proceedings of the International Conference on Health Informatics.

was useful to ensure that the board of the health organization in question decided to start a data governance project.¹⁵⁵

3.4 Future perspective

Given that data governance and data management solutions continue to grow based on the continuous supply of information, I therefore try to formulate and understand what the future trends could be for this sector.

To understand the future trends of data governance we can first evaluate the market trends and then the trends related to the technical capabilities of data governance.

Data Governance market trends

Even if data governance is a multidomain discipline, the main domains are Party / customer and product; especially in the first domain, party / customer, many initiatives will be put into practice to allow companies to be at the centre of all information relating to their customers. The main aims of future projects will therefore be much more specific and will include activities focused precisely on the management of business processes and a redesign of applications with the aim of developing more agile, comprehensive and client-centered strategies.

Many governance initiatives will be defined and implemented specifically to address the growing problem of managing different sources in such a way as to allow them to have a well-defined hierarchy giving them order.

There will be an increase in master data governance solutions in every sector of the market but will be characterized by two different forces: the first is the development of specific implementations of the sector or of the business that characterize the strongest and differentiating skills of each company, while another is to develop governance projects using the latest, most flexible and multi-domain capabilities to avoid creating too many different platforms that may be incompatible with one another, as they cannot be integrated with one another.

¹⁵⁵Alofayasan, S. et al. (2014). The significance of data governance in healthcare, Proceedings of the International Conference on Health Informatics.

In the future, most governance projects will follow a service-oriented model, but there will equally be the possibility that both real-time projects and batch projects are implemented by companies.¹⁵⁶

Data Governance Technical Capabilities trends

Data governance solutions combined with the products to be sold will continue to expand the data capacity along with many other dimensions such as:

- “organically”, by evolving core functionality within governance engine
- Through integration with new complementary technologies such as advanced data quality and matching solutions.
- Leveraging new computing paradigms such as virtualization, cloud computing and cloud related technologies such as software as a service (SaaS), platforms as a service (PaaS), Data as a service (DaaS) and so on. These computing paradigms dramatically and often positively change the cost, scalability, manageability and deployment factors of the governance project equations.¹⁵⁷

Data governance programs, moreover, will become open source solutions, thus allowing for an improved development cycle, shorter time to market, and reduced costs.

Furthermore, the definition of open source can leverage on the community that will be created by this openness, giving the possibility to get new ideas, new requirements and different testers. All this can create more robust, collaborative models, managing to develop a more efficient, more acceptable and more sophisticated technology. A direct consequence of the support of governance programs for structured and semi-structured data will be the positioning of governance as a potential platform to manage privacy policies and to help companies to strengthen and make strongly reliable the relationships that come to create with your customers.

Finally, by relying on the availability and efficiency of execution of so many rules and policies, it will be possible to extend the existing vision of data governance to support

¹⁵⁶ Berson, A., Dubov, L. (2011). Master data management and data governance, McGraw Hill Education, New York.

¹⁵⁷ *Ibidem.*

new governance information by creating data governance policy hubs that will manage other governance data hubs.¹⁵⁸

Governance processes must therefore be seen as an added value for companies in the research field; specifically, they will allow an iterative interaction with the purpose of revision and updates. Focus on data collection methodologies, their processing and analysis in research fields can be easily integrated into the process and be scaffolded through training and other associated resources and materials that from part of a governance system potentially consisting of:

- A training manual/online module
- An institutional storage options chart
- A referral map of all research support services across the institution
- A data governance self-assessment rubric¹⁵⁹

The most relevant challenges will be to safeguard the privacy of individuals and their security and try to improve data protection techniques and related unauthorized access.

the main areas of interest are:

1. *Unauthorized personal data use*: technology companies use daily methodologies that allow them to obtain huge amounts of their users' personal information. This data is often used to aim selected advertising and sold to third parties for the same purposes. Although, users are not always aware of the use that is made of their data. In the future there will therefore be the need to focus and invest in protecting users and their information.
2. *Lack of industrial standard data*: speaking of industrial and economic information, industries, even within the same sector, are often not encouraged to share their data because they believe that this is not qualitatively up to standard. One of the goals will be to set appropriate standards in motion, so that

¹⁵⁸Berson, A., Dubov, L. (2011). Master data management and data governance, McGraw Hill Education, New York.

¹⁵⁹Smale, N., Unsworth, N. et al. (2018). The history, advocacy and efficacy of data management plans, International Research Institutional.

information is available and useful to the industry and everyone will be able to benefit economically from it.¹⁶⁰

¹⁶⁰ Crawford, M. (2019). Governing the future: better data governance requires better standards, Mowat Center.

Conclusion

The aim of my dissertation was to outline and analyse the decision-making process within the use and application of data governance. From this study it became clear how necessary it is for companies to start developing data management strategies to efficiently obtain a substantial competitive advantage over other companies. Data governance is, for this reason, the most important component of the Enterprise Information Management discipline, and above all it is related to all the different business functions (product, customer, suppliers, partners, etc.). Observing companies' structures, it emerges that the level of dependence and reliance that the various companies have regarding the use of the data can vary. Looking at the definition of the term *governance*, this discipline is based on the levels of control to be applied to data, in such a way as to filter it, sort it, and manage it in one's favour. Executing data governance therefore implies applying the right level of control over data. This level of control can only be beneficial, having a positive effect on the quality of the data itself. The increase in data quality thus has a positive impact on the efficiency of business processes and consequently on the performance of the company itself.

From a practical study of the American health sector, it has emerged that whenever data governance practices are implemented, these bring an overall and welcomed benefit to the initial structure and consequently to the whole sector.

The implementation of the process of data governance, leads to an improvement in the entire performance of the business. Whilst in the preliminary phase of development, it can be difficult to communicate and specify the expected effects, data governance is born as a solution to corporate "problems".

The main motivations that drive companies to implement this type of process is the fact that data management, using the processes described in this paper, is able to raise the quality of the data flow belonging to the company, sorting it in a way that data becomes the main instrument of competitive advantage. The competitive advantage is pursued by improving the decision-making process. Therefore, with reliable, clean and tidy data, companies can make decisions related to data in a faster way, anticipating the competitors' next moves. This makes companies more flexible and dynamic, bringing them to succeed in their own sectors.

In developing this paper, I decided to focus on the importance of the different aspects pertaining to the implementation of governance strategies; above all, the focus is on the fundamental analysis and recognition companies need of their own level of maturity in data processing and managing, in order to form a good starting point for the business. It was at this point necessary establish the data relative to the company and essential to define the standards to which data must comply in order to be useful and effective for the company itself. Once these two specifications are defined, the governance program is automatically effective for the company.

The paper continued by establishing how fundamental the data for the company's decision-making process is and how to make the data a strategic asset. Accordingly, the right governance model enables companies to improve performance in different areas: reporting, cross-selling and up-selling, compliance and, as we have before mentioned, decision-making. Moreover, the chances of success of this type of strategy considerably increase if the companies that implement it consider establishing these strategies as a regular and internal tool to their company and not as a temporary program to use at needs. Implementing a strategic use of data will soon show the returns on investments after implementing the project. Monitoring these improvements and returns will make it simpler to understand the potential of data governance and will ease the involvement of all the members and sectors of the company.

Lastly, an outlook at future prospects reveals how this discipline is regularly and continuously changing and evolving. In an increasingly digitalised world, where billions units of data are generated every day, this is undoubtedly highly exposed to dangerous situations. As a consequence, the data governance sector is advancing towards the development of technologies that aim to proactively protect data to safeguard companies and people alike.

Bibliography

- AA.VV. (2010). *Best practices for a successful MDM implementation*, White Paper, September.
- AA.VV. (2011). *Enterprise information management: best practices in data governance*, Enterprise Architecture, May.
- AA.VV. (2016). *Data governance combined with analytics for powerful business insights*, Protiviti.
- AA.VV. (2017). *Data management and use: governance in the 21st century*, The Royal Society, British Academy.
- AA.VV. (2018). *Data management, privacy and security in connected systems*, Interact Lighting.
- AA.VV. (2018a). *Effective data governance*, in www.infosys.com.
- Abadi, D. (2009). *Data management in the cloud: limitations and opportunities*, Bulletin of the IEEE Computer Society Technical Committee, 12.
- Abu Saed, K. et al. (2018). *Data governance cloud security assessment at data center*, International Conference on Computer and Information Sciences.
- Akter, S. et al. (2016). *How to improve firm performance using big data analytics capability and business strategy alignment*, International Journal of Production Economics, 182.
- AICPA (2018). *An overview of data management*, in www.aicpa.org.
- Al Badi, A., Tarhini, A., Khan, A. (2018). *Exploring big data governance frameworks*, Procedia Computer Science, 141.
- Alhassan, I., Sammon, D., Daly, M. (2016). *Data governance activities: an analysis of the literature*, Journal of Decision Systems, 25, 51.
- Alofayasan, S. et al. (2014). *The significance of data governance in healthcare*, Proceedings of the International Conference on Health Informatics.

Al-Ruithe, M. et al. (2016). *A conceptual framework for designing data governance for cloud computing*, Procedia Computer Science, 94.

Al Ruithe, M., Benkhelifa, E. (2017). *Analysis and classification of barriers and critical success factors for implementing a cloud data governance strategy*, Procedia Computer Science, 113.

Arass, M., Souissi, N. (2018). *Data lifecycle: from big data to smart data*, Conference Paper

Baker, S., Sjoberg, P. (2018). *Intelligent data governance*, John Wiley & Sons, Hoboken.

Begg, C., Cairra, T. (2012). *Exploring the SME quandary: data governance in practice in the small to medium sized enterprise sector*, The Electronic Journal Information Systems Evaluation, 15, 1.

Berson, A., Dubov, L. (2011). *Master data management and data governance*, McGraw Hill Education, New York.

Board Briefing on IT Governance, 2nd ed., IT Governance Institute, Rolling Meadows, IL, 2003

Brennan, R. (2017). *Challenges for value driven semantic data quality management*, Proceedings of the 19th International Conference on Enterprise Information Systems, 1.

Brous, P, Janssen, M. et al. (2011). *Coordination decision-making in data management activities: a systematic review of data governance principles*, Springer Verlag, Berlin

C. Batini, C. Cappiello, C. Francalanci, A. Maurino, "Methodologies for data quality assessment and improvement" ACM Computing Surveys (CSUR), vol. 41, p. 16, 2009

Cai, L., Zhu, Y. (2015). *The challenges of data quality and data quality assessment in the big data era*, Data Science Journal, 14, 2.

Chakravorty, R. (2014). *Data governance within risk management*, CCR Magazine, March.

Chen, Y.-C., & Hsieh, T.-C. (2014). *Big data for digital government*.

International Journal of Public Administration in the Digital Age

- Cheong, L., Chang, V. (2007). *The need for data governance: a case study*, Australasian Conference on Information System, December.
- Crawford, M. (2019). *Governing the future: better data governance requires better standards*, Mowat Center
- D. McGilvray, *Executing Data Quality Projects*, Amsterdam: Morgan Kaufmann Publishers, 2008
- Deloitte (2013). *Developing an effective governance operating model*, in www.deloitte.com.
- Dumbill, E., (2012), *What is big data? An introduction to the big data landscape*, available at: www.oreilly.com
- Erwin Data (2018). *The value of data governance for the data driven enterprise*, in www.microway.com.
- E. U. Commission, "Commission recommendation of 6 may 2003 concerning the definition of micro, small and medium-sized enterprises" Official Journal of the European Union, L, vol. 124.
- Fleissner, B. (2014). *The importance of data governance in healthcare*, Encore Health Resources, White Paper.
- G. Dimitri, M. Patrice , G. Edward, C. Cecilia, F. Michael, B. Gabriela, et al., "*Annual Reports on European SMEs 2012/2013*," 2013
- Government, A. *The National Cloud Computing Strategy*. National. Broadband New. 2013, 2013, 36
- G. Reese, *Cloud Application Architectures: Building Applications and Infrastructure in the Cloud (Theory in Practice)*, Sebastopol, CA: O'Reilly Media, Inc., 2009
- Gregory, A. (2010). *Data governance – protecting and unleashing the value of your customer data assets*, Journal of Direct, Data and Digital Marketing Practice, 12, 3.
- Hameed, K., Al Ruithe, M., Benkhelifa, E., (2018). *Data governance taxonomy: cloud versus non cloud*, Sustainability, 10, 95.

- Hendey, L. et al. (2018). *NNIP's resource guide to data governance and security*, NNIP, September.
- Herbst, N.R.; Kounev, S.; Reussner, R. *Elasticity in Cloud Computing: What It Is, and What It Is Not*. International Conference on Autonomic Computing, San Jose, CA, USA, 26–28 June 2013
- J. Jeston, *Business Process Management: Practical Guidelines to Successful Implementations*, 2nd ed., Burlington, MA: Butterworth-Heinemann, 2008.
- Kim, H., Cho, J. (2018). *Data governance framework for big data implementation with NPS Case analysis in Korea*, Journal of Business and Retail Management Research, 12 (3).
- Klievink, B. et al. (2017). *Big data in public sector: uncertainties and readiness*, Information System Frontiers, 19.
- Knight, M. (2018). *The future of data governance: balancing data governance and data management*, Data University, 13 November
- Korhonen, J. et al. (2013). *Designing data governance structure: an organizational perspective*, Journal on Computing, 2, 4.
- Kourik, J., Qang, J. (2017). *The intersection of big data and the data life cycle: impact on data management*, International Journal of Knowledge Engineering, 3, 2.
- Ladley, J. (2012). *Data governance*, Elsevier, Usa.
- Lee, S., Zhu, L., Jeffery, R. (2017). *Data governance for platform ecosystems: critical factors and the state of practice*, Twenty First Pacific Asia Conference on Information Systems, Langkawi.
- Loshin, D. (2018). *Enterprise data architecture and data governance: use metadata to get to the starting gate*, Business Intelligence Solutions, 301.
- Lundberg, J. (2003). *Understanding data lifecycle management*, Veritas Architect Network.
- Manoj, M. (2015). *Big data governance frameworks for data revolution for sustainable development*, The Centre for Internet & Society.

- Martijn, N. et al. (2015). *Determining the effects of data governance on the performance and the compliance of enterprises in the logistics and retail sector*, International Federation for Information Processing, 37, 7.
- Martijn, N., Jonker, R. (2015). *The effects of data government in theory and practice*, Compact, I.
- Mullich, J. (2013). *Closing the big data gap in public sector*. Bloomberg Businessweek.
- Neera Bhansali (2013). *Data Governance: Creating Value from Information Assets*. CRC Press
- Nwabude, C., Begg, C., McRobbie, G. (2014). *Data governance in small business*, International Conference on Business, Management and Governance, IPEDR, 82.
- Oracle Corporations (2017). *All the ingredients for success: data governance, data quality and master data management*, in www.oracle.com.
- Panian, Z. (2010). *Some practical experiences in data governance*, World Academy of Science, Engineering and Technology, 62.
- Patel, J. (2015). *Data quality management, tools and techniques*, IJCS, 7, 1.
- Puri, C. et al. (2018). *Universal metadata repository: integrating data profiles across an organization*, IEEE Computer Society.
- Purohit, P. (2014). *Master data management*, Conference Paper, March.
- R. A. G. Monks and N. Minow, *Corporate Governance*, Chichester, England: John Wiley and Sons Ltd, 2008
- Rosenbaum, S. (2010). *Data governance and stewardship: designing data stewardship entities and advancing data access*, Health Services Research, 45, 5.
- Rossi, R., HIRAMA, K. (2015). *Characterizing Big Data Management*, Informing Science and Information Technology, 12.
- Russom, P. (2013). *Managing big data*, TDWI Research, Fourth Quarter.
- Salido, J. (2010). *Data governance for privacy, confidentiality and compliance*, Isaca Journal, 6.

- Shankaranarayanan, G., Even, A. (2004). *Managing metadata in data warehouses: pitfalls and possibilities*, Communications of the Association for Information Systems, 14.
- Scholl, H. J., & Scholl, M. C. (2014). *Smart governance: A roadmap for research and practice*. In iConference 2014 Proceedings
- Siddiqua, A. et al. (2016). *A survey of big data management: taxonomy and state of the art*, Journal of Network and Computer Application, April.
- Sidi, F. et al. (2014). *Data quality: a survey of data quality dimensions*, Journal of Quality Management, 4.
- Simonet, A., Fedak, G. (2017). *Active data: a programming model to manage data life cycle across heterogeneous systems and infrastructures*, Future Generation Computer Systems, April.
- Singh, G., Bharathi, S., Chervenak, A., Deelman, E., Kesselman, C., Manohar, M., Patil, S., & Pearlman, L. *A metadata catalog service for data intensive applications*. In Proceedings of the ACM/IEEE SC2003 Conference on High Performance Networking and Computing
- Smale, N., Unsworth, N. et al. (2018). *The history, advocacy and efficacy of data management plans*, International Research Institutional, 56.
- Sweden, E. (2008). *Data governance. Managing information as an enterprise asset*, Nascio Governance Series.
- Zhang, N., Yuan, Q. (2016). *An overview of data governance*, Economics Paper, December.
- W. H. DeLone, "Firm Size and the Characteristics of Computer Use," MIS Quarterly, vol. 5.
- Watson, H., McGivern, M. (2016). *Getting started with business driven data governance*, Business Intelligence Journal, 21, 1.
- Weber, K., Otto, B., Osterle, H. (2009). *One size does not fit all – a contingency approach to data governance*, Journal of Data and Information Quality, 1, 1.

Wende, K., Otto, B. (2014). *A contingency approach to data governance*, Conference Paper, November.

Wibbenmeyer, K. (2015). *How to successfully implement a corporate taxonomy*, Global Journal of Computer Science and Technology, 15, 5.

Sitografy

www.aicpa.org

www.deloitte.com

www.dmreview.com

www.forbes.com

www.gartner.com

www.infosys.com

www.oreilly.com

www.microway.com

www.oracle.com

Acknowledgment

Reaching the end of my university course, I would like to dedicate a thought to all the people who, over the years, have been close to me and supported me, physically and otherwise.

First of all, I would like to thank my parents, Giorgio and Roberta, who allowed me to reach this very important goal by never stopping to believe in me and pushing me to always give my best, and my brother Pietro, who I am sure will always be there for me as I will be for him.

The deepest gratitude goes to my girlfriend Maria who never stopped supporting me and who bared with me during this long period of research and writing, and pushed me not to give up and to always try to reach every goal I set my eyes to, wisely advising me; I will be always grateful.

Furthermore, I would like to address a thought to all my friends, who have never stopped making me laugh and make me feel part of something bigger than my own self. I would therefore like to name all the "Partigiani": Daniele, Alberto, Guido, Enrico, Gianluca, Alberto and Marco; and all the "Partigiane": Alessandra, Elena, Daria, and Roberta. "Cata e i suoi Amiketti": Matteo, Giovanni, Alvisè, Nicolò, Gianluca, Alessandro, Nicolò and Norberto.

A mention to my university mates, with whom I shared joys and sorrows Nicola, Pietro, Ilaria and Giacomo.

A special thank you goes to Fabio and Valeria for welcoming me to their home and family and to Cecilia, who helped me make this paper grammatically presentable.

I would also like to thank my supervisor and professor Giovanni Vaia, for having accompanied me on this final thesis writing path.

I also want to thank my company tutor Rossano Zara, for giving me the opportunity to be included in a very important firm like "Centro Porsche Padova", making me fully experience the greatness of the brand and making me know fantastic people.

Finally, I would like to thank my Grandparents Bruno and Gabriella, and turn my thoughts to my grandfather Ercole who unfortunately left us a few months ago.

You all mean so much to me, I will never stopped saying it.

Thank you all.

Alvisè

