

POLITECNICO DI TORINO

Master of Science in Management Engineering

Master Thesis

**A Method of Evaluation and
Prioritization of Data Governance
Activities in Big Data Projects**



University Tutor

Ing. Valentina Gatteschi

Prof. Gonzalo Ruz Heredia

Company Tutor

Ing. Marcelo Mercado

Author

Giorgio

Pulvirenti

Year 2019-2020

*To people that
are able to see the beauty*

ACKNOWLEDGMENT

I would like to take advantage of these few lines to express all the gratitude I feel towards the people who have contributed to this achievement (and not only ...)!

So I thank my father Martino, who has always supported me with great dedication; my mother Nancy who always urged me not to give up; Chiara who has always accompanied me during any decision taken; Moro who has always recommended the use of the most innovative software; Chiara without whom I would never have undertaken this thesis path and finally Lele and Matteo who does not know why but must be present in all Thanksgiving.

Special thanks go to Professor Gonzalo Ruz Heredia, to the corporate tutor Marcelo Mercado and to the whole Data Scientist team who have followed and guided me along the way and have always been present during all the meetings held.

SUMMARY

The General objective of the thesis is to help the adoption and the implementation of Data Governance Big data departments.

We initially hypothesized and later tested that an instrument of measurement of the activities of the Data Governance and the implementation of a prioritization of the activities in the backlog can help the adoption and the implementation of the Data Governance.

In order to reach our General Objective two specific objectives have been done:

1. Providing the structure and the content for organizing the evaluation of Data Governance for business intelligence and business analytic projects belonging to the Big data department.
2. Create a tool to prioritize Data Governance activities in the backlog-

The main problem encountered and in fact the reason why the framework is developed is that there is a lack of standardization in the Data Governance implementation process. This problem has repercussions at the macro level in the Data Governance program and at the micro level, in the Data Governance activities performed in each company project.

For these reasons this text will deal with presenting a framework for evaluating the implementation of the Data Governance program in each individual company project, belonging to a big data department. A set of activities have been identified, which each project should implement, which are of fundamental importance from the perspective of Data Governance

The final result of the thesis is to increase the efficiency of DG implementation in terms of activities done respect all the activities defined in the Data Governance. Also, the framework helps to respect Data Governance requirements

and business requirements.

The strength of the framework is to continuously measure the performance of each project from the point of view of Data Governance. The thesis proposes the framework according to a Data Governance structure that is created ad/hoc for our aim but it is also replicable in every Data Governance model. The document provides a description of the work done and the followed methodologies in order to better understand the use of the framework, however, it's important to note that these steps are presented in a linear way but some the "Non linear" methodologies were followed like Crisp-DB in order to have a continuous feedback with the stakeholders. Hopefully, this work will continue improving itself (considering that the world related to big data is in continuous changing). The evaluation and the prioritization phases will be processes of cooperation between the team responsible for the project and the Data Governance team. In order to evaluate and validate the framework we conducted a Testing phase in a limited number of projects but we hope that future works will cover a bigger amount of data doing some extensive tests in order to validate it also from a statistical point of view. This work supports the hypothesis that Lean Thinking, Process, Portfolio, Program and Project Management, and the Work System Theory, can complement the models and standards of Governance and Management of Enterprise IT, with an approach not existing in these models and standards of Enterprise IT, and suggests a Lean Framework that can support Governance and Management of Enterprise IT.

Contents

ACKNOLEDGMENT	1
SUMMARY	2
1 Introduction	7
1.1 Problem or opportunity	8
1.2 Hypothesis	8
1.3 The General objective	8
1.4 Specific Objectives	8
1.5 Keywords	9
1.6 Thesis structures	9
2 Introduction to Data Governance	11
3 Description of Big Data Department	15
3.1 Big Organization picture	15
3.2 Strategy	16
3.3 Relations with other Countries	17
3.4 Human Capital	18
3.5 The first step of Data Governance	19
4 Literary Review	21
4.1 Big Data Management	22
4.2 Rights and responsibilities	23
4.3 Dama Model	23
4.4 Starting with Data Quality	24
4.5 First steps to start a Data Governance Program	24
4.6 Our starting point for the evaluation model	25

4.7	Policy Based	25
4.8	Lean Thinking	26
5	Work Methodology	27
5.1	Agile	27
5.2	Scrum	28
5.3	What methodologies is being done	28
5.4	Agile for Big Data Governance	29
5.5	CRISP/DM	29
5.6	Huemul Big Data Governance	31
6	Methods and Materials - Framework	33
6.1	An Environment Study	34
6.1.1	Internal Business Study	34
6.1.2	External Business Study	37
6.2	Creation of the Framework	37
6.2.1	Data Governance Program Steps	37
6.2.2	Phase 1 Evaluation Model	38
6.2.3	Phase 2 Prioritization Model.	62
6.3	Updating the Model	72
6.3.1	Change of Business Requirements	72
6.3.2	Modification of the Related Business Requirements Weights	72
6.3.3	Modification of Activities	72
6.3.4	Interaction between Business Requirements and Activities	73
6.3.5	Update of Completion Indicators	73
6.4	Use of the Model	73
6.4.1	Activity Diagram	73
6.4.2	Usage of the Model	75
6.5	Test of the Model	77
6.5.1	Difference between two Projects in terms of Data Governance	77
6.5.2	Procedure of the Test	78
6.6	Test Results	79
7	Conclusions	81
7.1	Discussion	83

7.2 The impact of protests in Chile’s organization and the Data Governance role	85
List of Figures	89
List of Tables	91
Bibliography	93
Sitography	97

Chapter 1

Introduction

The framework for evaluating the Data Governance program in each individual company project, has been developed in Enel Chile in the BigData platform. Through an internship it has been possible to develop and use the framework. In parallel, I was assigned the role of coordinator of the team of consultants who was responsible for managing the data governance of the department. In this way it has been possible to be in close contact with Data Governance activities and compare them with theory.

The state of the department, where the framework has been developed, 3 years ago presented a map of applications at risk of obsolescence and heterogeneity that did not allow access to best practices or had the capacity to adapt to meet the needs of digitization and business growth. So, two years ago, this path began in the BigData platform division, which has allowed the development of a set of business intelligence and business analytics products that support and improve business processes. The implementation of the platform, also, reduces complexity and enables digital transformation by intelligently occupying and structuring the Data Lake (DL) LATAM. Currently, the Data Lake LATAM contains a large part of customer information from all LATAM countries, where part of the data is from Synergia, Epic, and other sources charged to the Data Lake for various needs. Some information is still outside the DL (for example, large customer data - SAP) and the division between all the objectives wants to have all the data in the DL and started a work of planning of loading new sources in the Data Lake. In order to achieve that the department invested a lot of money in a Data Governance program. Currently there is a team of consultants that is

in charge of all the Data Governance projects that interfaces with the Big Data platform team to present the results and advance with the requests. This team is developing different solutions but in a non-organic way and without having a clear vision of what are the goals that they want to achieve. In this context, Big Data platform managers do not know also how to evaluate the Data Governance team and do not have tools to evaluate whether Data Governance practices are present in the department's projects.

1.1 Problem or opportunity

The main problem encountered is that there is a lack of standardization in the Data Governance implementation process. This problem has repercussions at the macro level in the Data Governance program and at the micro level, in the Data Governance activities performed in each company project.

1.2 Hypothesis

The hypothesis is that one instrument of measurement of the activities of the Data Governance and the implementation of a prioritization of the activities in the backlog can help the adoption and the implementation of the Data Governance providing all the benefits that the Data Governance has.

1.3 The General objective

The General objective is to help the adoption and the implementation of Data Governance implementation.

1.4 Specific Objectives

1. Create a tool to evaluate Data Governance activities for each internal project in the department.
2. Create a tool to prioritize Data Governance activities.

How?

A measuring instrument will be applied for the activities of the Data Governance as well as an activity prioritization tool for the activities still in the backlog.

Desirable result

Increases the efficiency of DG implementation in terms of activities done respect all the activities defined in the Data Governance. Helps to respect Data Governance requirements and business requirements.

1.5 Keywords

Having a theoretical description of the main concepts is of fundamental importance to continue with the analysis of the following framework. We suggest to the lector to have a knowledge of some important keywords regarding Digital transformation:

Big data, Business Analytic, Business intelligence, Data Culture, Data Driven, Data Strategy.

1.6 Thesis structures

In this chapter an introduction to the thesis has been presented explaining what are the hypotheses from which we started, what are the problems encountered and the objectives set for solving the problem.

In the second section, a high level introduction on data Governance will be made for all those readers who have just started on the topic.

The third section will give an overview of the department in which the internship took place in order to describe the context in which the framework was applied.

In the fourth section we will talk about the state of the art of Data Governance.

The fifth section will focus on the methodology used in the construction and application of the framework.

The sixth section represents the body of the work, here will be discussed: the creation of the model, the use, updating and testing of the model.

In the seventh section the conclusions will be presented in which it will be discussed what added value the framework brings.

In the eighth section we will report the discussions in which advice will be given on how to use the framework in different business contexts. In this section an interesting chapter will focus more about the use of Data Governance in certain emergency situations.

Chapter 2

Introduction to Data Governance

In business, it is, most of the time, about timing. Whether it is better customer service, improved marketing ROI, discovering new business opportunities, or enhancing business processes, the Business Intelligence department can deliver results if and only if the right business user gets the right data (intelligence or insight) at the right time. Data, which is inaccessible, inaccurate, or untrustworthy, is of no use to an average business user.

In the best-case scenario, probably both the data and the BI environment will require continuous governance and monitoring. To get the needed *competitive advantage*, enterprises will need to know exactly what data is available wherein both internal and external data pipelines. This will become more a reality as data volumes and complexity continue to rise. Every organization wants to turn its stores of data from business byproducts into valuable assets. As big data has grown and technology has rapidly evolved, companies of every size have explored ways that data analysis can help them leverage their data for strategic business decision making and competitive differentiation.

A critical tool for making this a reality is enterprise data governance, which ensures the health of any data supply chain as information is deployed and processed across an organization's data ecosystem. As the data travels through the data fabric of an organization, it gets transformed, enriched and the new data elements/metrics are added. The governance for the data domains should account for this shared ownership and accountability. Both information systems (IS)

researchers and practitioners consider data governance as a promising approach for companies to improve and maintain the quality of corporate data, which is seen as critical for being able to meet strategic business requirements, such as compliance or integrated customer management. Both sides agree that Data Governance primarily is a matter of organization.

However, hardly any scientific results have been produced so far indicating what actually has to be organized by data governance, and what data governance may look like. The paper aims at closing this gap by developing a framework that helps companies to evaluate the effectiveness of the data governance program in their projects.

The thesis aims at closing this gap between governance and management on the basis of a comprehensive analysis of the state of the art both in science and in practice. At the same time, it is beneficial to practitioners, as companies may use it as a guideline when organizing and evaluating data governance. Regarding the quality of their data, companies see themselves confronted with a number of strategic business requirements, such as compliance with legal and regulatory provisions or the need for customer-centric business models. In this context, data governance is seen as a promising approach for companies to improve and maintain the quality of their data. At its most fundamental level, data governance consists of people, processes and technology. All working together to manage data, establish data quality, and facilitate the usage of data to support an organization's objectives and ensure (Analytics & Marketing Insights – Ryan Chase):

- Consistently, accurate data
- Repeatable processes for efficiency and quality
- Holistic (data) analytics strategy
- Improved decision-making (based on data not on gut)
- Democratization of data to achieve the highest levels of analytics maturity and support self-service

An overall policy for how data is formatted and managed must come from an observation of how the business works. This policy not only dictates the processes of collecting and formatting data in a usable way, but it is also dependent on the observation of business processes. It's a good idea to have IT people involved in

developing policy and data processes shadow the business units they're supporting so they can see how data processes can support business processes.

Each person must be able to speak the same language when it comes to the pillars of data governance, such as data quality and data lineage, he said.

In order to start our travel in the data governance there are four guidelines that we should follow (DAMA-DMBOK2):

- Only work on what is needed
- Set a deadline
- Get the right people involved
- Check back frequently with your stakeholders.

Chapter 3

Description of Big Data Department

In this chapter, we will make a brief description of the department in which the internship took place and in which the framework was developed in parallel. It is of fundamental importance to have a descriptive picture of the department because the framework has been developed in a technological context where we work with Big Data and therefore this changes the traditional business paradigms. Furthermore, seeing in which position of the organization we find ourselves allows us to see from what point of view we see the data Governance and how we must interface with it.

3.1 Big Organization picture

From an organizational point of view, Enel is structured in a Business line such as: Renewable Generation, Thermal Generation, Infrastructure and Network, Retail (Market), Enel X, etc ... One of these business lines is Market and we worked in here. In this business line, there is a sector called Global Digital Solution Southam. GDS is present in all the countries of South America: Chile, Argentina, Peru, Brazil, and Colombia. In Chile GDS manages and maintains the Data lake and therefore has a function of coordination also of the other GDS of the other countries that access the data lake. We, therefore, find ourselves in GDS Market Southam and here work different teams dealing with different platforms such as Billing, CRM, Big Data. In the Big Data Platform within GDS Market Southam,

the author of this thesis did the internship. In the department, the projects are divided in Business Intelligence and Business Analytics projects.

Figure 3.1 shows the division of the projects in Business Intelligence and Business Analytics and also shows the new Data Governance projects that are being born.



Figure 3.1. Business intelligence, analytics and governance in the Big Data Platform. “Image taken from GDS Market Southam”

3.2 Strategy

This path of Digital revolution began in Enel Market Southam 3 years ago, which has allowed them to develop a set of BI / BA products that support and improve business processes. In addition, in these years, there were several changes of personnel and this has impacted much what was the know-how created in the big data platform. We can therefore say that this team that manages this platform works relatively from a little time on the platform and therefore the projects in progress or finished are also few.

This will slightly affect us in the final testing phase of the model. One of the pillars of the global strategy of the Enel Group is to move towards a Data-Driven Company where data forms the basis for making decisions and measuring the results of those decisions. The Big Data Platform, takes the data as the epicenter

of processes and decision making, putting itself at the service of people and platforms. It also allows concentrating on different sources, all the data of the company in a Data Lake. Here the Data is organized in process structures to facilitate its exploitation in a more efficient way, creating a more valuable information asset. Big Data Platform is at the service of people and applications and is a base information asset for the creation of Reporting (BI) and Analytical Models (BA) for decision making.

Managing and governing data as Strategic Assets means have these benefits:

- *Data Reliability*
- *Consistency*
- *Completeness*
- *Data Availability*
- *Accessibility*
- *Opportunity*
- *Data Collaboration*
- *Visualization*
- *Self-Service*
- *People*
- *Competencies*
- *Cultural change*

3.3 Relations with other Countries

It must be considered that GDS Chile has a coordination function towards GDS Colombia, Brazil, Peru, Argentina. There is, therefore, to be considered a whole theme of cooperation between the different countries. The big data platform, in fact, seeks to develop all its projects in common and in collaboration and cooperation with all 5 countries. Obviously, this point depends very much on the

project, in fact for some it makes sense, others do not, in a mixed way. But in general, when it is desirable to design in cooperation with all countries. Because in the end, they are the same ones who will then use the results of the projects. The uniformity allows better management and standardization. In this regard, Cooperation and Collaboration here are some points that they are trying to push:

- GDS Chile wants to maximize **cost synergies** by building **centralized solutions** that meet local needs.
- **Same products for all the geographies.**
- **Standard processes**, addressed by data and with local adoption.
- **Common architecture** for the region that allows disseminating best practices and capturing synergies.
- **Best practices Automation processes and scalability.**
- **A joint space between business and IT** (product managers, product owners, and platform managers). Business defines the "what", IT defines the "how".
- Perform a **“smart” integration** with Enel X and make use, above all, of new instruments like Financial (Fintech). In the medium term, it may be a source of information for other Business applications.

3.4 Human Capital

Enel is a company that works a lot with consulting companies. Most of the staff working inside Enel are in fact outsiders contracted by consultancy agencies. It turns out to be a point of reflection because it has a series of repercussions. Surely the working environment is dynamic and there are elements with great capacity and having different backgrounds. In the same way, however, the know-how is lost a little between all these consultancy companies and also that which is the corporate identity is sometimes not recognized. In this department, there are different BI/BA projects but now we focus on the governance platform evolution project that is outside of the BI/BA segmentation. In fact, for about 6 months, a team of consultants has been contracted to see all those issues related to Data Governance. This team has not developed a program and works by developing

the requests provided to it by the Big Data Platform team. There is, therefore, a lot of work to be done in the construction of the Data Governance program that to date is not yet implemented, planned and well structured.

3.5 The first step of Data Governance

From the perspective of Data Governance, below will expose some activities that have been followed and some of the points that the same team has promised to carry out in the future.

What the Big Data Platform Department has done so far will be summarized as follows:

- Ingestion and Data Processes
- Continuous optimization of infrastructure and processes
- Data intake monitoring
- Monitoring of data transformation processes
- Construction of data consumption services between platforms
- Development of good practices at a functional level and Data technicians
- New working methodologies to reinforce the potential use of Data Lake through Data Insights & Delivery

What they are moving on?

- Improve Data Catalog in Data Lake
- Conclude the implementation of Sandbox for Users
- Data Quality Process Improvement in Data
- Data Governance and Data Security Improvement
- Cover new areas of information (for example CARE)
- Data enrichment with new sources

- Combine technical and business skills
- Disseminate and stimulate the use of the platform through ambassadors.
- Data Governance is also working to check that any data flow follows the architecture established in the department. Figure 3.2 shows the systems architecture in the department

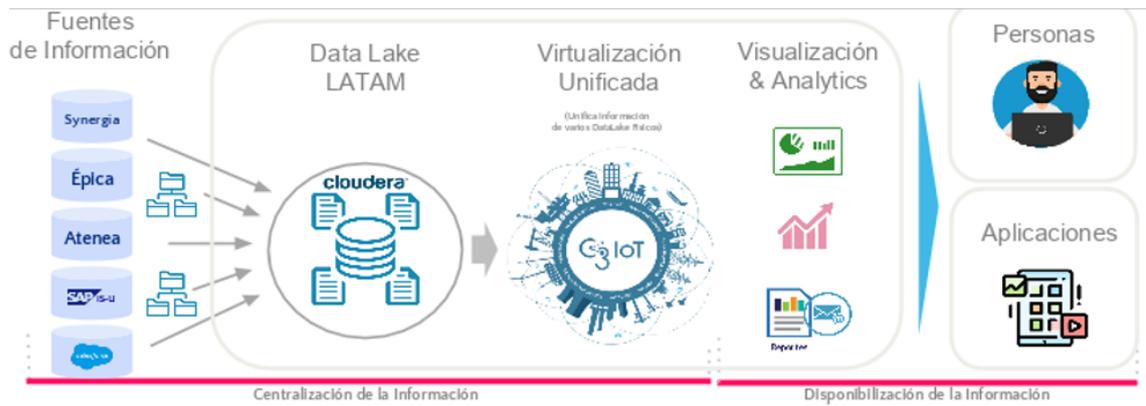


Figure 3.2. High Level BigData platform architecture

Chapter 4

Literary Review

In order to create our framework, we analyzed a big literature of the Data Governance until now. For example, we analyzed the main frameworks to create a Data Governance program or some interesting applications of the data governance program in practical cases. We saw that a well designed governance process clearly defines how data and analytic work successfully towards generating business outcomes in an organization. Implementing governance can be challenging and requires considerate all the work done from the others in order to implement new things and grow in the field. Also, we should do a careful assessment, choices, and prioritization of the most valuable roles and activities that all the data governance framework proposes. We must not forget also to consider all processes and people of our organization and how they interact. As we can see an increasing number of firms are accelerating the deployment of their big data analytic initiatives with the aim of developing critical insight that can ultimately provide them with a competitive advantage [1]. Exponential data growth has placed information governance as a critical issue for senior IT and business management [2]. More in general IT governance has been at the center of attention for both IT researchers and practitioners for over two decades and empirical evidence suggests that by establishing appropriate governance schemes and practices, the implementation of IT strategy can be executed in alignment with the business strategy, which ultimately leads to firm performance gains [3].

4.1 Big Data Management

If we came across data governance then surely, we would have wondered what the relationship is with the data management systems and specifically with big data management. To be sure, data governance is closely correlated with data management and indeed part of the outcomes of Data Governance are the same as data management. However, they differ because data management is created and used with the sole purpose of management and optimization. Data governance, on the other hand, has a more optic of governance, that is, application of safety rules, management of responsibilities and quality. Big data management is a new discipline, where data management techniques, tools, and platforms "You can't manage what you don't measure"[4]. Exploring the Big data management, we will also understand the importance of incorporating coordination theory into a governance research framework [6].

Just to remember that Big Data does not mean no privacy. It is also known that data governance is of particular importance when it comes to the management of master data, i.e. data about materials and products, customers and suppliers, for example [7]. The Centre for International Governance Innovation addresses the contemporary issue that gets the most attention: balancing the exciting uses of big data with the desire to maintain a high, or at least acceptable, level of privacy [6]. There are studies that aim to explore the existing Big Data governance frameworks and their shortcomings and propose a new framework [8]. The proposed framework consists of 4 macro components and 18 micro components. As a framework validation, the proposed framework has been compared with the ISO 8000 data governance framework and, as we will see, has a lot in common with it. To examine the impact of information governance in contemporary organizations, researchers analyzed survey-based data from 158 firms that have embarked on big data initiatives(Information Governance in the Big Data Era: Aligning Organizational Capabilities) Analyzing the Data Governance from a high level, are provided principles, definitions and a model for governing bodies to use when evaluating, directing and monitoring the handling and use of data in the organizations [9]. This document is a principles-based advisory standard.

4.2 Rights and responsibilities

If we start looking for Data governance literature we will find out that there is only limited research on it. Before continuing with rights and responsibilities we want to say that for our research, we also incorporate data governance sources from consultants, analysts and practitioners [10]. As mentioned above, it's common sense that data governance is basically about assigning decision rights and responsibilities when it comes to data management in companies [11]. Furthermore, practitioners and researchers agree upon the fact that implementing data governance is an organizational design task [12].

However, a holistic perspective on the entirety of the several organizational concepts is still missing with regard to data governance. Therefore, in the following a conceptual framework the different dimensions of “organization” as a theoretical concept is derived from studies already done [13]. We will talk deeper about DAMA guide later. Introducing roles and committees for data governance typically involves hierarchy-overarching and hierarchy-complementing measures. In organizational design, such measures are the result of an effort to mitigate the dysfunctional aspects of the primary organization [14].

There have been numerous suggestions, particularly from the practitioners' community, regarding roles and committees to which decision rights and responsibilities are assigned in the course of data governance. Precision and unambiguity of terms is still insufficient, though. When understanding on data stewardship, for example, suggestions range from “data steward” [15] to “data custodian” [15] to “data governance manager” [16] to “chief steward” [17].

Talking about roles and permission gives users the opportunity to also try to close this gap by developing a morphology of data governance organization on the basis of a comprehensive analysis of the state of the art both in science and in practice [18]. Executives and managers have a big role in data governance, so a novel construct there is a theoretical framework for the governance of data [19].

4.3 Dama Model

We already mentioned the Data Management Body of Knowledge [20] that presents a comprehensive view of the challenges, complexities, and value of

effective data management. DAMA model recognize that managing data is central to the success of the organizations and they recognize data has value and they want to leverage that value. As our ability and desire to create and exploit data has increased, so too has the need for reliable data management practices. The second edition of DAMA International's Guide to the Data Management Body of Knowledge [21] updates and augments the highly successful DMBOK1. An accessible, authoritative reference book written by leading thinkers in the field and extensively reviewed by DAMA members, DMBOK2 brings together materials that comprehensively describe the challenges of data management and how to meet them by: DAMA-DMBOK2 provides data management and IT professionals, executives, knowledge workers, educators, and researchers with a framework to manage their data and mature their information infrastructure, based on some principles.

4.4 Starting with Data Quality

As we will see data quality is an important theme when we talk about data governance and one of the most important things it's the integration with the data management program. Integrated DQM (Data Quality Management) is required in order to address both organizational and IT perspectives. In the literature there are a few Data Quality Management (DQM) approaches dealing with accountability. Successful DQ programs identify the organizational processes behind DQ [22]. Some can think that data governance is a subtype of Data quality management. We totally disagree in this framework because we will consider data quality as a dimension of Data Governance, but this is useful in order to have a bigger picture of Data Quality [23]. Data Governance and specifically quality's not losing efficiency, in fact, methodologies and some tools reduce Analytics cycle time while improving quality [24]. We also will talk later about DataOps principles.

4.5 First steps to start a Data Governance Program

In order to start to think about Data Governance an interesting step it's to fill the framework Data governance in one page [25]. This model evaluates and

compare IT governance with a proposed matrix representing IT governance in one page, mapping how key decisions are made. Some white paper helped us to understand the impact of Data Governance in big organizations. For example, the principles and best practices for data governance in the cloud [26]. This white paper provides guidance and best practices for data governance as we move our data into the cloud. It provides a framework for data governance in the cloud, deep dives into how implement data governance in our organization and outlines the business benefits of a robust data governance implementation. We can't start planning a Data Governance program if we didn't explore the idea that corporate data needs a comprehensive strategy [27].

4.6 Our starting point for the evaluation model

Our framework for data governance includes five interrelated decision domains: Data principles; Data quality; Metadata; Data access and Data life-cycle. Some of these decisions domains are well described in [28], a paper that shows organization's decision-making about its data assets. We will use this framework as a starting point of our dimensions. Data Governance it's not a project it's a program and this means that we need also a plan for continuous improvement thanks to our well-designed framework, we will be able to begin implementing and managing our governance process with a concrete plan [29]. Next, we will talk about dimensions belonging to Data Governance. One of these is the Data Access which means to define an IAM program and a well-defined strategy and governance model to guide the life cycle phases Insights on governance, risk and compliance [30].

4.7 Policy Based

In a big Company like Enel, we also have to understand what Policy-based management means [31]. This book aims at providing an overview of recent work that identifies and addresses some challenges. The design of suitable policy management systems requires addressing many research challenges:

- As data volumes increase and devices start getting imbued with artificial intelligence.

- Allowing for increased autonomy, managing them, especially managing their security challenges.
- Becomes increasingly difficult. Policy-based management has proven itself useful in simplifying the complexity of management in domains like networking.
- Security, and storage, and it is expected that many of those benefits would carry over to the task of managing big data and autonomous systems.

4.8 Lean Thinking

When we talk about governance we also talk about processes and how the processes are executed. It's important to present and analyze the main concepts of Lean Thinking, Lean IT, standards for Governance and Management of Enterprise IT, Portfolio, Program and Project Management, Processes Management, and Work Systems Theory [32]. But Lean and agile methodologies are old methodologies and we also look in something that we can apply to the data Governance [33]. The goal of this ongoing research is to academically elaborate DataOps as a new discipline. This paper explores the body of knowledge and presents a working definition of DataOps and an initial research framework. DevOps is a methodology contained in DataOps that also redefine and optimize the way organizations develop and deliver applications by having development and engineering teams collaborate more closely to accelerate delivery [34]. As we said it's important to see how the data governance works in daily practices [35].

Chapter 5

Work Methodology

A few years ago, the CIO of Enel Chile established some high-level strategy for the entire Enel group. The entire Enel group and also the Enel Latam group have to put all the resources to implement the strategies. One of these is to work with an Agile methodology. So, with this objective, many investments were made to begin this great change. Enel Chile is a company with a culture that passes from a state company that does not allow it to be easily prone to change. The organizational structure and human capital are not accustomed to change and are accustomed to working with a Waterfall methodology. So, step by step is trying to implement agile in all projects without but forcing the methodology but pushing it lightly. It is very important to consider what work methodologies are currently used from the perspective of Data Governance because it will be the program that will be built in relation to this. Data Governance can also dictate a methodology that is based on the Agile strategy of Enel Global but with specific practices focused on the data.

5.1 Agile

Agile is a methodology for project development, but not like any methodology, but one that has demonstrated much faster and more flexible than any other. It is treated, in a certain way, in a different way to work and carry out any project, and to organize. One of the exceptional things that Agile has is that it divides a project into small parts that allow the work teams to be able to complete them in a few weeks and with complete efficiency. In agile, there is an incremental and iterative development approach. The work is prioritized according to the

value of the business or the client. There are multi functional teams that work on product iterations for a period of time. Each iteration focuses on the production of a work product.

Agile is specifically used by high-level work teams, something that is very necessary if we want to make the most of this tool. Agile's goal is to develop high-quality services and products at an extreme speed, all focused on responding to the needs and priorities of customers in a more precise and effective way.

5.2 Scrum

Agile is a way of thinking and doing things, it is governed by the agile manifesto where there are 4 main values and 12 principles, and there are various frameworks, frameworks, or methodologies that apply that Agile culture, but they have unique practices that they differ from other frameworks, that is, Scrum and Kanban have the Agile culture but each of them has different practices that differentiate them, same values, different practices, different approaches. Scrum is a tool that is part of the agile tools portfolio. Scrum is a method to work in teams based on iterations or Sprints. Thus, Scrum is an agile methodology, so its objective will be to control and plan projects with a large volume of last-minute changes, where uncertainty is high. It is usually planned for weeks. At the end of each sprint or iteration, the validated work of the previous week is reviewed. Based on this, the activities in which we will invest our resources in the next Sprint are prioritized and planned. The Scrum methodology focuses on adjusting its results and responding to the real and exact demands of the client. Hence, each deliverable is reviewed, since the requirements vary in the short term. The minimum time for a Sprint is one week and the maximum is four weeks.

5.3 What methodologies is being done

The Market / GDS / Supplier inter-functional Group have started working under Agile methodology using the Scrum tool. They actually have:

- **PRODUCT OWNER**, responsible for the definition and scope of the project. Determines the backlog and MVP to be developed and validates each of the deliveries.

- **BUSINESS ANALYST**, provides the functional knowledge of the process and supports the development team. work hand in hand with the P.O.
- **SCRUM MASTER**, responsible for compliance with the values and principles of the scrum framework. it facilitates the resolution of squad impediments
- **DATA COMPETENCE**, technological coach that is responsible for providing advice and guide the development of solutions.
- **DIGITAL SOLUTIONS REFERENT**, leader in charge of advising technically to the work team.
- **DEVELOPERS**, responsible for building and developing the solution, based on the definition and scope determined by the P.O.

5.4 Agile for Big Data Governance

We said before that the department of Big Data Platform as all the Enel departments cannot apply the agile methodologies blindfolded. We should Implement it Step by Step and adapting it according to the department in which we find ourselves.

5.5 CRISP/DM

One of the agile tools largely in the world of data analysis is Crisp-DM. It is built from the book DAMA-DMBOK2 (Data Management Body Of Knowledge) and allows us to organize and control Big Data projects at the corporate level. The CRISP-DM methodology — which stands for Cross-Industry Standard Process for Data Mining — is a product of a European Union-funded project to codify the data mining process. Just as the agile mindset informs an iterative software development process, CRISP-DM conceptualizes data science as a cyclical process. The software development process can be summarized in the Figure 5.1.

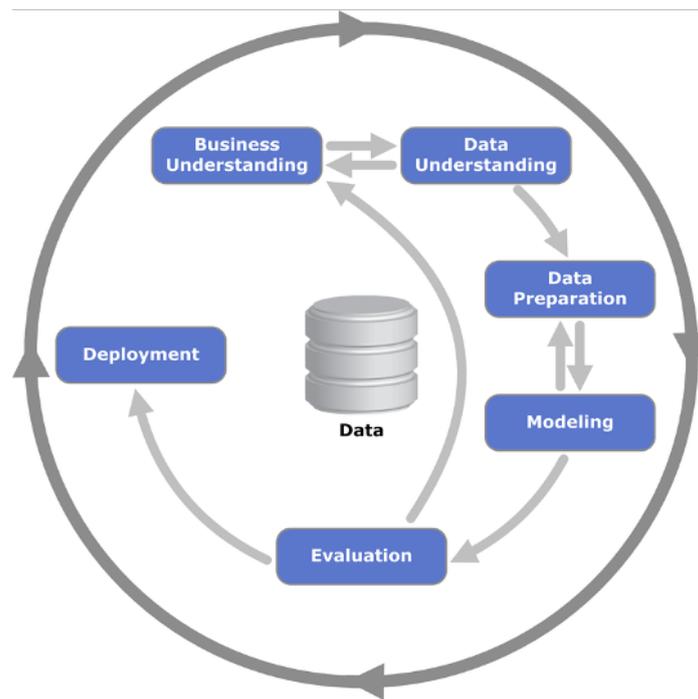


Figure 5.1. Crisp-DM phases.

Source:

https://es.wikipedia.org/wiki/Cross_Industry_Standard_Process_for_Data_Mining#/media/Archivo:CRISP-DM_Process_Diagram.png

While the cycle approaches on iterative progress, the process doesn't always flow in a single direction. In fact, each step may cause the process to revert to any previous step, and often, the steps can run in parallel. The cycle consists of the following phases:

- Business understanding
- Data understanding
- Data preparation
- Modeling
- Evaluation
- Deployment

5.6 Huemul Big Data Governance

The implementation of a unique data corporate strategy, based on good Data Governance practices, is possible thanks to Huemul: a library that works on Spark, Hive and HDFS.

The Huemul BigDataGovernance design is designed to optimize the development time of data analysts, and at the same time increase the quality and governance of the data. Using a solid methodology that allows our team's engineers to focus their efforts on the definition of the input interfaces, the definition of the tables and the construction of robust mastering processes. Huemul is an operation Model Based on Good Data Governance Practices.

The construction of Huemul BigDataGovernance is based on the good practices described in the DAMA-DMBOK2 (Data Management Body Of Knowledge) and allows to speed up the development of Big Data projects at the corporate level.

Chapter 6

Methods and Materials - Framework

This is the body of the thesis and in this Chapter we will provide in detail all the procedures performed to build the framework, to update it and to use it.

Before going into the framework, we want to give guidelines for implementing the General Data Governance program. We already have given some suggestion regarding the methodology of work to use in the previous chapter 6 Work methodology, but now we will enter a little bit in the detail of the Data Governance Framework. Following the in-depth study of the state of the art of Data Governance for big data, a certain degree of knowledge of the Data Governance program has been gained. Therefore, some advice is addressed to the Data Governance team in order to implement the Data Governance Program in an optimal way. Once this initial phase is over, we will enter into the body of our thesis, that is into the evaluation and prioritization framework.

I believe it is useful to understand this document by the Data Governance team even if they do not use the proposed tool. This is because, to create the framework, we have considered all past studies on the subject of Data Governance and this document, therefore, also acts as a synthesis of the state of the art. All the guidelines to create and implement Data Governance will be shown below. These procedures should be done as a first step also for the creation of a program of Data Governance. In the case that they were not made, they should still be done even if during or afterwards the creation of the Data Governance program. Below we will analyze 4 macro sections that indicate the 4 macro steps performed in

the Development of the Model: environment study, creation of the model update of the model.

6.1 An Environment Study

In this section we will give guidelines on what is the study to do before creating a Data Governance framework like ours. Obviously, we need to analyze the environment in which we are going to position our framework.

6.1.1 Internal Business Study

Of fundamental importance for a team that deals with Data Governance is to do an internal company study. Internal benchmark is based on the observation and learning of positive behaviors, but from the company's own departments. To carry out an internal benchmark we must analyze the different departments in which our company is divided, observing in detail the productivity of each one of them. From among the whole, we must select the one that stands out for its greater efficiency and analyze the reasons that make it especially productive.

Once the conclusions have been extracted, it will be time to apply them to the rest of the departments, in order to make everyone look like the best and most productive and efficient of them; The one selected as a model. Figure 6.1 explains, from a geographical point of view, where the data governance in the department is positioned also in relation to the data strategy and the data culture.

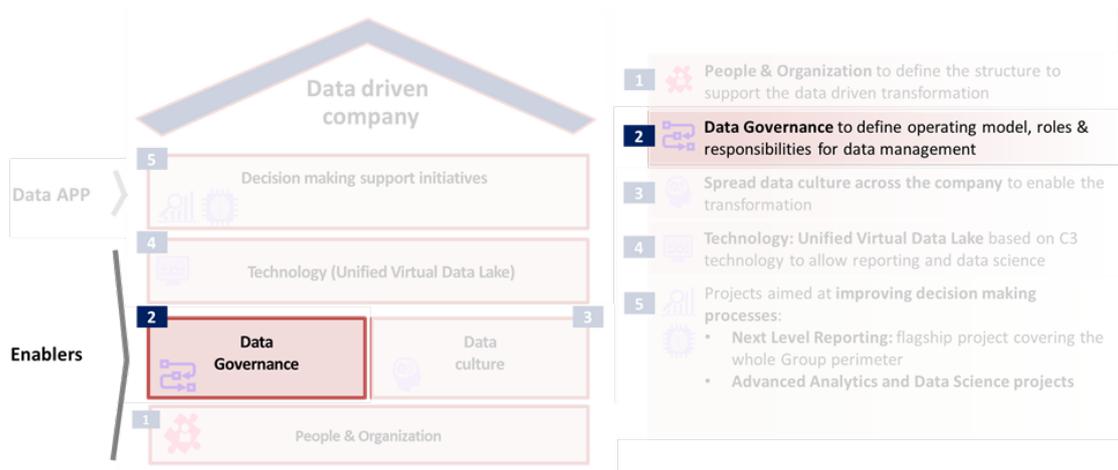


Figure 6.1. Data Governance in the Data Driven company Enel.

“Image taken from GDS Market Southam”

Evaluation of Corporate Policies

Company internal regulations are a tool that every organization that aims for efficiency should be equipped with. And if we think about it, it is absolutely not something strange.

For example, in our case Enel Global, based in Rome, Italy, dictates policies that have a vertical impact on the entire organization, thus also reaching GDS.

Once we have in hand all the policies that apply to our department, we must begin to evaluate the critical issues found, understand what is not clear and what is not done in the ways and terms that we expect and what are the behaviors within company that create difficulties and are an obstacle to productivity and the achievement of set goals. In our specific case, different security policies, access and data quality apply. All these will be considered in the framework created. The policy that we will take as a reference is policy 440 which in fact speaks of the organization of Data Governance.

Analysis of new National Regulations

An important part is the analysis of national regulations and laws on the use of personal data that they apply to our department. In our case we must consider the Chilean law n.1628 and we focus on: "Title I De la utilización de datos personales". We must not consider all the points of the law but only what

affect our division. The right to privacy or private life has been developed both by doctrine and legislation. In Chile, its legal development is found on the protection of private life, which configures, in this respect, a separate statute, with regulatory autonomy. In accordance with this, personal data are those relating to any information concerning natural persons, identified or identifiable. The processing of personal data in our legislation recognizes limitations, built by a catalog of rights recognized in favor of their owners and whose respect is guaranteed by actions expressly provided by law.

Study of the Working Environment and Working Methods

It is of fundamental importance to analyze the work environment from a psychological and methodological point of view. The psychology of work and organizations is the study of people's behavior in the workplace and in the performance of their professional activity in relation to interpersonal relationships, the tasks to be performed, the rules and the organization's functioning. Furthermore, it is necessary to analyze which are the methodologies used in the development of the projects, which are the methodologies being implemented and which are the desired ones. This point is well analyzed in chapter 7 Work methodology.

Analysis of Business Problems and Needs

This study of the department in which we need to develop the Governance data will surely bring to light some problems and some business needs. This must be taken into consideration and analyzed from the point of view of Data Governance. Later we will see that an important part of our framework is the construction of the Business requirements. These are built precisely from the problems and business needs.

Consulting Role

The last point of the internal company evaluation is the evaluation of the role of business consultants. Big Data companies are large companies and the percentage of consultants who work in internal projects is not indifferent. It is therefore necessary to see what their role is as consultants and how they relate to projects administered by internal company workers. For example, in our case the Data Governance team is a team composed of only consultants and an internal company

project manager. The way in which they relate to the project manager and how the tasks are entrusted to him is of fundamental importance.

6.1.2 External Business Study

An important step is to study the external business environment. Over the years Data Management technologies has been growing fast. Technology is often changing our methodologies to make things easier and better. So, it is important to follow this change and continuous evaluate the change being open to the world. We should evaluate new methodology of Data Governance and new tools. In addition, an interesting tool that is having a large success in the market of Metadata Management tools of Data Governance is “Coollibra”.

6.2 Creation of the Framework

In this section we will show all the procedure for the creation of the framework. This section is very important in order to personalize the framework considering our purpose. Once we understand in which activity of the Data Governance program we find ourselves we start to describe the creation of the framework. Each company should take the following model as a reference, and should adapt it according to specific business needs

6.2.1 Data Governance Program Steps

The aim of this document it’s to define the framework of evaluation and prioritization of the Data Governance activities in Big Data Projects. In order to understand when we have to use this framework we should understand what are the main steps of the General Data Governance program.

These phases are of a high level but they give us an overall view of where our framework is.

Typical Data Governance program steps are:

1. Define goals and understand benefits;
2. analyze current state and delta analysis;
3. refer to road-map;

4. convinces stakeholders and budget project;
5. develop and plan the data governance program;
6. implement the data governance program;
7. monitor, control and evaluate.

We then analyze the last step of the Data Governance program “Monitor, control and Evaluate”. Not surprisingly they look very much like the phases described in ISO 800 (execute monitor control). Our aim is not to define completely this step of the Data Governance program. But it is simply to provide a useful tool for the evaluation. In the next chapters 6.2.2 Phase 1 evaluation model 6.2.3 Phase 2 prioritization model, we will talk about the creation of the two central parts of our framework.

6.2.2 Phase 1 Evaluation Model

From the State of the Art to the 4 Pillars

An in-depth study was carried out on the state of the art and also on the various data governance frameworks implemented globally as can be seen from chapter 7 literary review. Therefore, the fundamental pillars of which the data governance program should be composed were identified.

Starting from these five pillars, the data governance model was defined for internal company projects. We need to apply a change of perspective, in fact, now we look at data governance from the point of view of the individual project. This means going to analyze the specific activities related to data governance that each project should implement to comply with the guidelines of the data governance program (Figure 6.2).

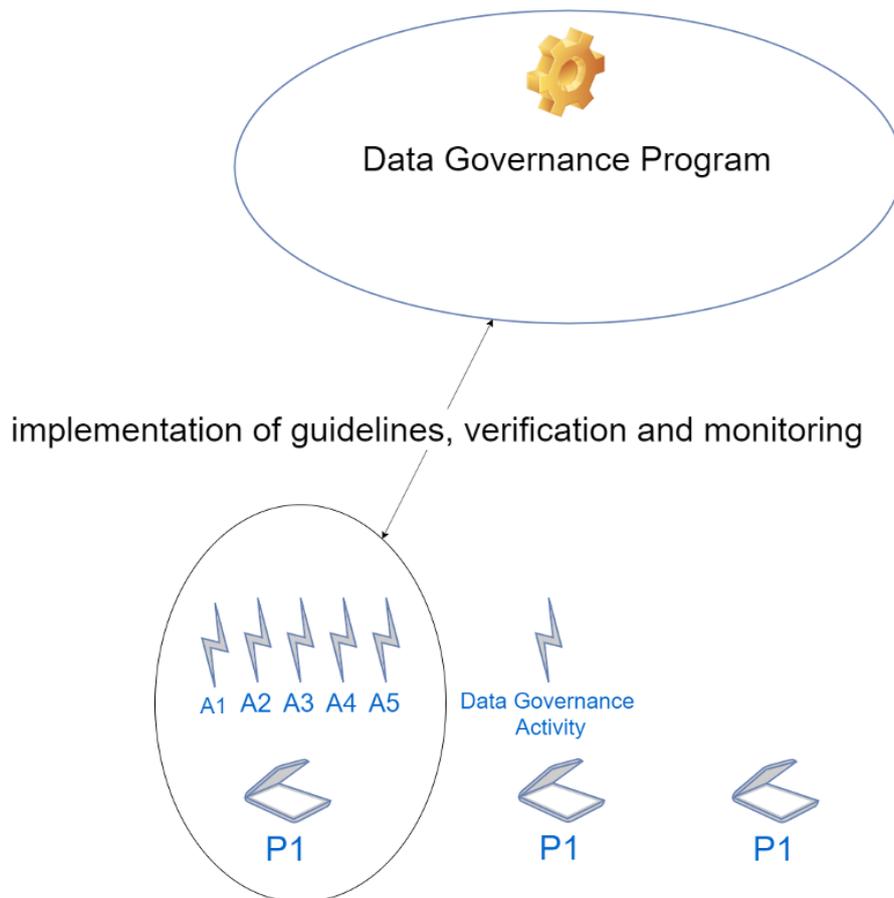


Figure 6.2. Relation between Data Governance Program and Data Governance activities of the projects.

For each Project, 4 fundamental pillars were identified (figure 6), each of which contains 5 dimensions corresponding to 5 macro activities, with the exception of the last pillars for which only 3 macro activities have been identified. For the definition of the activities we have considered all the previous internal and external study, in our case of Enel. Below we will describe the 4 pillars encountered in order to understand better what is involved. It is desirable to think that the Data Governance team implements programs starting from each of these pillars. This is not the goal of our text so we will provide a description of the pillar and only an introduction to the activities of the program.

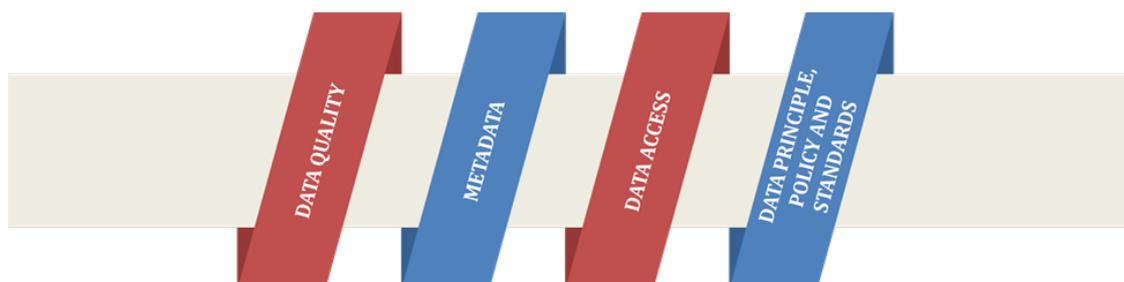


Figure 6.3. Pillars of the framework.

Data Quality

It is necessary, for meeting regulatory compliance requirements, ensuring data quality and getting value from big data implementations. Data-driven company needs to encourage everyone to detect issues and propose data-quality improvements. Of course, any proposed initiative will have to be analyzed, reviewed, and emphasized. A data quality office can oversee the analysis of data issues, evaluate the impact that it has on the business, propose solutions to cleanse existing data, and either mitigate or prevent future occurrences. In multi-domain MDM, it is natural to have certain data-quality requirements and activities already presumed since they are intrinsic to consolidating and synchronizing master data, as explained earlier in this chapter.

Therefore, the data-quality roadmap is somewhat embedded to the overall MDM roadmap. Although MDM provides the opportunity to start a data-quality practice, DQM needs to be built as a strong practice with well-supported capabilities and leadership. To be sure, they need to be collaborating functions. Data quality is the ability to supply accurate, timely and complete data, which can be translated into information, whenever and wherever this is required. Data quality is vital to effective decision making at all levels of the organization and also to promote a common quality model across Group, providing guidelines in terms of Roles, Processes and Metrics according to a “proactive approach. Clear quality status is based on objective and shared indicators for an effective decisions-making process. Data Quality main process design and implementation are:

- **Strategy** - Develop a Data Quality Strategy

A DQ strategy is required to manage and direct all DQ activities in line with the overall business strategy. The strategy includes the strategic objectives

which are pursued by DQM, how it is aligned with the company’s strategic business goals and its overall functional scope

- **Organization** - Design the Data Quality Organization

Designing the DQ organization includes defining roles and responsibilities, determining information needs, defining metrics and standards, and designing data processes.

- **Information Systems** - Design the Data Quality IS Architecture

The DQ architecture is aligned with and supports the overall enterprise architecture. Designing the DQ architecture encompasses creating a business data repository and defining the information systems in line with data quality requirements. Data quality tools may support the information quality improvement process.

Metadata

Simply, metadata is “data about data,” and it generally defines the content of a data object. Metadata within the data governance practice has the primary responsibility of enabling policy and providing access to data. These policies include those concerned with data definition, data usage, data security and data lineage and heritage. It is important to remember that although governance and policies are created to determine the appropriate actions to be applied to a given data object, ultimately, they must be applied to the physical storage of the information as well. Metadata will assist in both business and technical instantiating of the data, making it a very powerful part of the data governance practice’s set of tools. As more and more people are using and producing data, managing metadata is no longer just the busy work of a few isolated data management technologists – the number of data stakeholders has grown dramatically! While the need for and value of metadata management is growing, implementing it has become more complex due to the increasing diversity of data sources and stakeholders. Organizations will benefit from a comprehensive view of their metadata, and of metadata management. If an organization fully understands the value of metadata management, they will implement the proper management and technical solutions that enable the discovery and collection of all forms of meta data - both business and technical. Metadata can be the center of the

data governance effort, since understanding the context of the data content is the central concept of data stewardship. To achieve the business benefits of enterprise data management, the connection between the data instances and the various forms of metadata associated with each instance of data becomes an asset to be managed for competitive advantage. Some forms of metadata that may be overlooked include business rules, calculations, algorithms, data usage patterns - these are as important as the basic definitions and data types / formats most usually associated with the term "metadata."

Metadata provides the linkage between the business need or desire (policy) and the information or data value. The effective management of metadata is one of the essential activities of a data steward within a governance practice, thus enabling data management policy and access to information. Metadata management refers to the activities associated with ensuring that meta data is created / captured at the point of the data's creation and that the broadest possible portfolio of meta-information is collected, stored in a repository for use by multiple applications, and controlled to remove inconsistencies and redundancies. In short, data governance uses metadata management to impose management and discipline on the collection and control of data.

Metadata Management includes phases related to both Data Catalog structure definition and content information management.

So, the general Activities of Metadata Program should be:

- Data Catalog - to develop a unique Data Catalog, describing all data of interest in the Unified Virtual Data Lake (UVDL).
- To define the information set to transform data into knowledge.
- Making data and related assets available to all users to foster data democratization and avoid redundancy.
- Metadata is the set of information that transforms data into knowledge by providing specific contextual information to enable.
- Data objects understanding for better decisions making.
- Data relationships and data transformations impact.
- Map all divisional Data Sources and algorithms.
- Data Catalog development.

Data Access

Every organization depends on information, whether it's financial data, intellectual property or sales forecasts. Data is everywhere, and it's essential to the success of any organization to enable collaboration across the business. Not all data is created equal, nor is it managed in the same ways. However, one thing is clear - all access to data must be managed and protected.

Managing access to data stored in files is a growing problem. In the past, sensitive data was relatively protected within structured systems and applications. As businesses have become more collaborative users have exported large volumes of data to create new content in the form of PDFs, spreadsheets, documents, and presentations. This user-generated content is then stored in often unprotected file shares, collaboration portals such as SharePoint, cloud storage systems including OneDrive and Box, email systems and other repositories. The amount of data stored in files has increased exponentially over the past few years and is projected to grow 800% in the next five years. Sensitive data stored in files exposes organizations to significant security and compliance risk. For example, organizations are very focused on controlling access to sensitive data within their payroll application (structured data), but with a simple “export” button, that information can be migrated into a spreadsheet, which can then be stored in a variety of ungoverned locations such as SharePoint or Box. Very quickly, that highly sensitive data finds its way into many unstructured files with virtually no oversight. Without the proper visibility and controls, it is difficult to know where these files reside, what sensitive information they may contain, who has access to them, and what they are doing with the data.

Adding to these security concerns, many enterprises are challenged to maintain compliance with the growing number of data privacy laws as these files proliferate across the organization. The data access standards (and the associated service level agreements) can be based on the definition of “unacceptable” uses of data and external requirements for auditability (the ability to track who / what has accessed / modified data), privacy and availability. Data access decisions also provide standards at the physical and logical level.

Below some activities that Data Governance program should implement related to Demand and Supply:

- Manage data provisioning requests.

- Clear accountability on access control.
- Demand and Supply Management includes phases related to both data accessibility and incident management.

Principles Policy and Regulation

Effective data principles establish the linkage with the business. For example, the organizational decision to standardize business processes implies that there should be a clearly defined business owner of data assets (data principle). By delineating the business uses of data, data principles therefore establish the extent to which data is an enterprise wide asset, and thus what specific policies, standards and guidelines are appropriate. In keeping with the notion of data as an asset, data principles also establish / foster opportunities for sharing and reusing data. Each principle is supported by a rationale and a set of implications. Data principles consider the usage of external data, such as, customer data from third-party service providers. An organization's data principles also take into consideration the regulatory environment that could influence the business uses of data. Data principles therefore define the desirable behaviors both for IS professionals and business users. For example, the notion of business owners of data implies that business users have an important role in managing data quality as well as its lifecycle, interpretability and access. On the other hand, IS professionals play the role of data stewards instead they employ IT tools (such as, DataFlux, Informatica Data Quality) that help surface quality issues for the business owners (or data owners / trustees).

We will talk later about DataOps methodology but here there are a few best practices associated with implementing a DataOps strategy:

- Establish progress benchmarks and performance measurements at every stage of the data lifecycle.
- Defines semantic rules for data and metadata early on.
- Incorporate feedback loops to validate the data.
- Use data science tools and business intelligence data platforms to automate as much of the process as possible.

- Optimize processes for dealing with bottlenecks and data silos; this typically involves software automation of some sort.
- Design for growth, evolution and scalability.

From the Pillars to Dimensions

Each project, therefore, from the point of view of Data Governance should activate 18 macro activities related to Data Governance that means 5 for each pillar except the last one. From now on we will use the term dimension in the same way as activities because each pillar defines 5 dimensions that eventually translate into 5 activities (with the exception of the last pillar). The Dimensions have been identified based on an in-depth study of the state of the art and the available frameworks (see Chapter 6 "Literary Review" and in particular the sections c; d; e; f).

As for the activities or "Dimensions" included in each pillar, these were decided based on an in-depth study of the various frameworks in circulation, company policies, the operation of the Big data department and also on the basis of a quantitative analysis carried out in based on a small test. We will see in the chapter test of the model that one of the model projects of the department was compared with one of the projects gone wrong. We have analyzed the delta between the two projects, which means that we have compared the activities related to the Governance data made in the project that went well and confronted the activities done and not done, related to the DG, of the project gone wrong. In this way it was possible to confirm the activities taken into consideration by the previous theoretical study. It should be noted that in these dimensions we describe the activity to be done from a high level. This model just wants to give only a framework to work on but it is then up to each company and each department in which it is to be applied to establish the procedures and specific operations to complete the activity of the dimension. In the following, all the dimensions relating to the pillars described above will be described one by one.

Data Quality

METRICS

Different data consumers may have different data quality requirements, so it is important to provide a means to document data quality expectations as well as techniques and tools for supporting the data validation and monitoring process. Quality will be evaluated and monitored through clear and objective metrics / KPI based on Quality Dimensions. Intrinsic KPI have been defined suitable for all data; Contextual KPI will be detailed according to metrics provided by the data office Use case by Use Case For each dimension, Data Quality KPI are elaborated upon standard metrics that enable to measure or evaluate performance, progress, efficiency or other data characteristics. The elaboration of quality metrics into KPI is defined in relation to each Use-Case specific requirements. Intrinsic metrics represent the minimum set of data quality requirements that need to be satisfied for mapping data on UVDL. This activity consists in define and calculate the Quality metrics related to the project. Ideally just one metrics set for each project. The evaluation of these metrics thresholds is provided by the relevant Data Owner:

1. Completeness
2. Validity (Accuracy)
3. Consistency
4. Timeliness
5. Uniqueness

This metrics are well described in the Data quality metrics model of Enel (Figure 6.4) and below there is an example (Figure 6.5) of use case in order to better understand the structure and the use.

		Description	Scope	Applicability	Examples
	Completeness	Measure of the absence of empty values (null string or empty)	Ensure to have enough data representing the information needed	Entity Attribute	A given Table is characterized by 5 mandatory fields. If a record has a missing value is not "complete"
	Validity	It refers to compliance of each field with correspondent syntax rules (format, type, range)	Ensure that values of data accurately match defined rules	Entity Attribute	Date values must be stored in the same defined format. I.e. "dd/mm/yyyy"
	Accuracy	Measure expressing the accuracy in representing a business object	Ensure data objects reliability in terms of business and technical expectations	Entity Attribute	Representation of costs in a different currency than the expected one.
	Consistency	Refers to the coherence and uniqueness of data representations across systems	Ensure that data values are unique and equally visible by all users. [Critical to standardize data representations and manage shared concepts]	Entity Attribute	Data available in different "Tables/Systems" (i.e clients) are always characterized by the same "Value"
	Timeliness	Measures punctuality, actuality and freshness of data	Ensure fresh data by setting rules related to time data, such as frequency of updates, actuality and more.	Entity Attribute	Weekly sales data must be updated every week to be considered reliable
	Uniqueness	Ensure the absence of duplicates in data	Guarantee that data is unique. Solves problems related to size and complexity of data sources	Attribute	Employees serial numbers must be unique and no duplicated values are accepted

Figure 6.4. Data quality metrics set. "Image taken from GDS Market Southam"

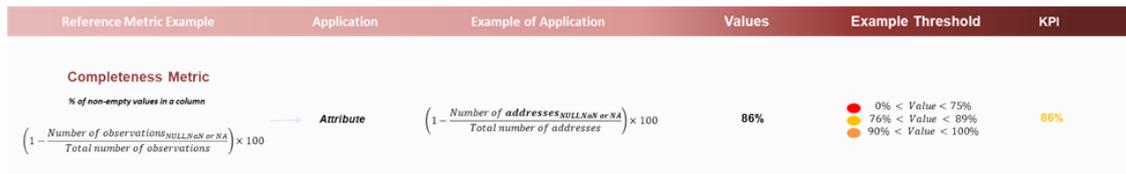


Figure 6.5. Example metric use case. "Image taken from GDS Market Southam"

TOOLS

Data quality tools are the processes and technologies for identifying, understanding and correcting flaws in data that support effective information governance across operational business processes and decision making. The packaged tools available include a range of critical functions, such as profiling, parsing, standardization, cleansing, matching, enrichment and monitoring. A supporting quality tool will be defined by the Data Governance Program in order to provide all the instruments necessary to manage the processes (workflow engine), to build up and to monitor Metrics / KPI according to Roles and Responsibilities defined (Profiling). This activity consists in the implementation of the Data Quality tools defined in the Data Governance program. In our case the tools are:

hive, talend, spotfire, control m

Hive allows you to create the infrastructure necessary to manage a complete on top of Hadoop data warehouse. Hive is positioned within the batch layer of Hadoop and allows you to manage both structured and unstructured data stored within the Blob Storage. Use an SQL-like language called HiveQL. It is ideal for any type of non-transactional analytics.

Integrate Hive data with standard components and data source configuration wizards in Talend Open Studio.

Talend is an open source data integration platform. It provides various software and services for data integration, data management, enterprise application integration, data quality, cloud storage and Big Data. Talend allows:

- Clean and reliable data production.
- Immediate verification of data quality.
- Compliance with data protection regulations.

Spotfire is an analysis tool to get insight from your data. It allows anyone to quickly and easily to create simple measurement dashboards, predictive applications, or dynamic real-time analysis applications, with numerous scalable features, including visual analytics, data wrangling, predictive analytics, location analytics, and streaming analytics. In Spotfire, inline and AI-powered data wrangling helps users spot issues, fix data quality problems, inconsistencies and deficiencies as they interact with data. Users don't need to go outside Spotfire to prepare data before visualizing. They can combine, shape, clean, enrich and transform data all while doing data exploration. (<https://www.businessprocessincubator.com/content/tibco-spotfire-named-a-leader-in-the-forrester-wave-data-preparation-solutions-q4-2018/>)

PROCESSES

The right processes for data quality management will provide measurably trustworthy data for analysis. This dimension defines structured Data Quality Management activities for managing, measuring and monitoring data quality. Also, it provides guidelines to define metrics and KPIs for validating and managing data quality along the entire life-cycle. Following the processes described by the policy we can efficiently give the access to the data to who needs it and with a

clear and transparent protocol. This activity consists in Implement Data Quality management processes described by the policy 440 when happen a Data Quality case:

- **Use-Case Data Quality Validation**
The Data Quality Validation describes the activity of preliminary understanding of data quality conditions and metrics required for data-driven initiatives within a GBL / Staff and Service Function / Digital Hub / Digital Solutions; the activity aims to certificate data quality within data sources before being mapped on UVDL.
- **Vertical Data Quality Remediation**
The Data Quality Remediation phase consists in the operational procedure for correcting any data quality issues arising within a GBL / Staff and Service Function / Digital Hub / Digital Solutions.
- **Use-Case Data Quality Contextual KPI Definition**
The activity describes the workflow of introducing new contextual data quality metrics (KPI) within a use case for measuring the adherence to technical and business requirements.
- **Vertical Data Quality Monitoring**
The vertical monitoring activity is an ongoing and periodic data quality control performed within each GBL / Staff and Service Function / Digital Hub / Digital Solutions.
- **Central Data Quality Monitoring**
The central monitoring activity is an ongoing and periodic data quality control that defines the internal controls for monitoring the state across GBL / Staff and Service Functions / Digital Hub / Digital Solutions 'data sources mapped on the Unified Virtual Data Lake.

BUSINESS VALUE

Data governance means better, leaner, cleaner data, which means better analytics, which means better business decisions, which means better business results. Stewards will work with IT departments, analysts, and end-users to keep the program aligned with business goal. In order to reach it we have to stay in

continuous contact with the business. This can empower business and IT teams to interact with data — with both the agility the business demands and the data security IT needs. This activity consists in a continuous interaction between business and IT in order to Evaluate the Business value for each increment of the Quality KPI metrics considered by the project correlated to the quality metrics. Practically this means calculate the economic impact of the increasing or decreasing of each data quality metrics. This activity kind of force us to stay in continuous contact with the business.

ISSUES

Several issues can arise when an organization's data governance framework is not proper. But there are ways of sorting things out. From the perspective of the single project, during the development, it is common that happen many quality issues. This activity consists in govern them doing two activities:

1. Find as soon as possible the solutions of the past and present issues that are in the waiting list;
2. identify the possible future issues.

So, we should activate a process of continuous analysis of past and present and future issues. Practically this activity consists of Identifying Possible futures quality issues of the project. It also consists of analyzing well the past quality issues solved and organize well the resolutions of the present and past issues pending. Assign each issue to the person with fair responsibility and understand when and how he/she will solve it.

Metadata

PROFILES AND OBJECTS

Once our data assets are assessed and classified, it is crucial that we document our learning so that our communities of data consumers have visibility into our organization's data landscape. We need to maintain a data catalog that contains structural metadata, data object metadata. The following objects described below aim to provide the useful data that has to be managed by the

data catalog. Define the structure of the data catalog and the tools to manage it it's not the purposes of this document and it should be one of the purpose of the data Governance big program. In order to help the data governance group to maintain update, in time, the data catalog this dimension (Profiling and objects) and the following dimension Data Lineage are super useful. This activity consists in providing detailed information by the objects of the project:

- **Data definition** (formal Description) that the project will define.
- **Data goal** (formal description) that the project wants to achieve.
- **Data sources** (Data Base, File System, Data Lake) that the Project will use.
- **Entities** (Table or specific views) that the Project will use or create.
- **Attributes** (Columns) that the Project will use or create.
- **Models** (Data Science Models) that the Project will use or create.
- **Metrics** (Metrics and KPIs) that the Project will use or create.

DATA LINEAGE

We should realize that all data moves through life-cycle stages is central to designing data governance. For example, from the perspective of data in an electronic health record (EHR) maintained by a hospital, the uses and consequently the value of the diagnostic information of a patient admitted in the hospital changes as the patient undertakes surgery, moves to an acute care center, is discharged, receives a follow-up consultation, and transitions from sick-care to wellness-care. By understanding how data is used, and how long it must be retained, organizations can develop approaches to map usage patterns to the optimal storage media, thereby minimizing the total cost of storing data over its life cycle. Many organizations do not know what data they have, how critical that data is, the sources that exist for critical data, or the degree of redundancy of their data assets. In order to manage inventory of data as well as its various data sources, information chain manager develop an understanding of different types of data that are the most / least prevalent, their storage requirements, and

the growth trends. This activity consists in providing detailed information by the Data Lineage (the process of the data) of the project:

- Data Acquisition and Capture: Define your Data Suppliers.
- Data Processing: Define your project activity related to the Data, Data responsible, Data shareholders
- Data deliverable: Defines consumers of the data processed or created by the project, data stakeholders
- Data Maintenance: Implement a Robust Maintenance Plan

PROCESSES

We should define structured Metadata Management activities for managing and monitoring Data Catalog Metadata within Unified Virtual Data Lake. Also, we should provide guidelines to make data and related assets visible to both business and technical users, identifying the relevant data perimeter. Following the processes described by the policy we can efficiently give the access to the data to who needs it and with a clear and transparent protocol. This activity consists in Implement Metadata management processes described by the policy 440 when happen a Metadata management use case:

- Data Catalog Design
Activities for the creation / modification of data catalog structure (e.g., change set of extended characteristics of catalog entity object)
- Metadata Insert
Activities for inserting new entries within the data catalog (e.g., adding new C3 Type and corresponding catalog entity Metadata)
- Data Catalog Information Retrieval
Activities for the retrieval of data catalog information (e.g., find a specific Type mapped in the catalog).
- Metadata Update
Activities for updating data catalog entries (e.g., change or delete current Metadata fields definition for a C3 Type).

- Metadata Monitoring-Remediation
Activities for supervising Metadata consistency, issues, usage and security across Digital Hub / Digital Solutions and systems.

TOOLS

The need for a framework to aggregate and manage diverse sources of Big Data and data analytics and extract the maximum value from it is indisputable. Metadata management is designed to address this task. It provides powerful tools that put information assets to work more effectively including ratcheting up governance and compliance while reducing risk.

Metadata management solutions typically include a number of tools and features. These usually include metadata repositories, a business glossary, data lineage and tracking capabilities, impact analysis features, rules management, semantic frameworks, and metadata ingestion and translation. This activity consists in the implementation of the Metadata tools defined in the Data Governance program. In our case the tools are: HUE, Talend.

ISSUES

Several issues can arise when an organization's data governance framework is not proper. But there are ways of sorting things out. From the perspective of the single project, during the development, it is common that happen many metadata issues. This activity consists in govern them doing two activities:

1. Find as soon as possible the solutions of the past and present issues that are in the waiting list,
2. identify the possible future issues.

So, we should activate a process of continuous analysis of past and present and future issues. Practically this activity consists of Identifying Possible futures metadata issues of the project. It also consists of analyzing well the past metadata issues solved and organize well the resolutions of the present and past issues pending. Assign each issue to the person with fair responsibility and understand when and how he/she will solve it.

Data Access

PROCESSES

The first aspect is the provisioning of access to available assets. It's important to provide data services that allow data consumers to access their data, and fortunately, most cloud platform providers provide methods for developing data services. These platforms define structured Demand and Supply activities for data accessibility and the related access provisioning and provide guidelines for managing incident related to data management and the related escalation. Following the processes described by the policy we can efficiently give the access to the data to who needs it and with a clear and transparent protocol. This activity consists in Implement Data Access processes described by the policy 440 when happen a Metadata management use case:

- Data Accessibility
Activities for managing request to access to specific data perimeter.
- Incident Management and Escalation
Activities for managing unplanned or critical issue that cannot be remediated through standard processes and may escalate to the Data Office.

ROLES AND PERMISSIONS

One important aspect is prevention of improper or unauthorized access. It's important to define identities, groups, and roles, and assign access rights to establish level of managed access. This dimension involves managing access services as well as inter-operating with the cloud provider's Identity Access Management (IAM) services by defining roles, specifying access rights, and managing and allocating access keys for ensuring that only authorized and authenticated individuals and systems are able to access data assets according to defined rules.

We must set roles and responsibilities for Metadata Management activities that involves both GBLs / Staff, Service Functions, GDS relevant units and Project team. This activity consists in the Identify the roles and permissions they have to access different project data and prioritize access actions. This practically means identifying for each person within the project or outside the project that interacts

or that can interact with the data the permissions they have and the prioritizations.

SECURITY

Despite the efforts of information technology security groups, in the companies, to establish perimeter security as a way to prevent unauthorized individuals from accessing data, perimeter security is not, and never has been sufficient for protecting sensitive data. Attempting to prevent someone from breaking into our system has limited success, but at some point, our data may become exposed. We might sustain a security breach, or even insider exfiltration. It is important to institute additional methods of data protection to ensure that exposed data cannot be read, including encryption at rest, encryption in transit, data masking, and permanent deletion.

This activity consists in the application of the Security guidelines synthesized by these 2 points:

1. Follow the **Cyber Security Framework** (Enel policy n.17) Implement the following processes:
 - (a) Processes aim at describing cyber security entire lifecycle.
 - (b) Processes aim at describing cyber security activities carried out on a rolling basis.
2. Managing Logical Access to information systems (Enel policy N.25) Implement the following processes:
 - (a) Role categories and classification.
 - (b) Logical access management.
 - (c) Policy enforcement and periodical controls.

BACKUP, RECOVERY AND ARCHIVING

All too often, digital archiving projects get pushed to the bottom of the priority list when planning digital workplace initiatives. People make the mistake of thinking that low storage costs and a strong search engine justify retaining all of their data. Without a good Backup, Recovery and archiving practice in place,

organizations open themselves up to a number of possible risks. This activity consists in the application of these 2 points:

1. Program and Do the Backup and Recovery. Implement a Robust Recovery Plan
2. Retention and archiving policy. Implement the processes described in the policy of Enel in our case.

ISSUES

Several issues can arise when an organization's data governance framework is not proper. But there are ways of sorting things out. From the perspective of the single project, during the development, it is common that happen many access issues. This activity consists in govern them doing two activities: 1 Find as soon as possible the solutions of the past and present issues that are in the waiting list, 2 identify the possible future issues. So, we should activate a process of continuous analysis of past and present and future issues. Practically this activity consists of Identifying Possible futures access issues of the project. It also consists of analyzing well the past access issues, already solved, and organize well the resolutions of the present and past issues pending. Assign each issue to the person with fair responsibility and understand when and how he/she will solve it.

Principles Policy Regulation

Principles

Data governance is a continuous process and it has to be managed properly over the years. The core principles identified should be kept simple and understandable to all the layers of an organization. Overall, participation and adherence to the key points are the guiding factors for a successful data governance implementation. The principles that we came up with represent a simple view of what his organization set out to accomplish by deploying a formal Data Governance program. Every project should implement these principles in order to help the Data Governance team to govern the organization. It is our thought that organizations should consider these principles as an easy way to describe the basics of Data Governance for their organization. Data analytics teams that can

successfully meet the requirements for rapid delivery of new analytics will play a high-visibility role in helping their organizations compete in the on-demand economy. Improving the speed and robustness of analytics can be achieved using a process and tools approach called DataOps. DataOps draws from process innovations in software development and lean manufacturing. Organizations that implement DataOps correctly have experienced significant improvements in the ability to produce robust and adaptive analytics. DataOps may be implemented in seven simple steps without discarding an organization's existing analytics tools:

Data Ops:

- Add Logic Tests (Testing of inputs, outputs, and business logic must be applied at each stage of the data analytics pipeline)
- Use a Version Control System (like Git)
- Branch and Merge
- Use Multiple Environments
- Reuse & Containerize
- Parameterize Your Processing
- Work Without Fear (Value and Innovation Pipeline)

Policy

Policies and procedures are an essential part of any organization. Together, policies and procedures provide a roadmap for day-to-day operations. They ensure compliance with laws and regulations, give guidance for decision-making, and streamline internal processes. However, policies and procedures won't do our organization any good if our employees don't follow them. Employees don't always like the idea of having to follow the rules. But policy implementation is not just a matter of arbitrarily forcing employees to do things they don't want to do. Following policies and procedures is good for employees and our organization as a whole. In such a way as to reinforce the need to implement the policies we summarize all the policies of our company in this case Enel in

such a way as to give greater weight to the policies in our framework. Each activity of the project should implement the same guidelines and rules of the policies:

1. Data Governance policy 440. Implement the following processes:
 - (a) Metadata Management
 - (b) Demand and Supply Management
 - (c) Data Quality Management
2. Cyber Security Framework policy n.17. Implement the following processes:
 - (a) Processes aim at describing cyber security entire lifecycle
 - (b) Processes aim at describing cyber security activities carried out on a rolling basis
3. Managing logical access to information systems. Policy n.25. Implement the following processes:
 - (a) Role categories and classification
 - (b) Logical access management
 - (c) Policy enforcement and periodical controls
4. Retention and archiving policy. Implement the processes described in the policy of the company, in our case Enel.

Regulation

There is growing recognition of the risks of exposing individuals' personal and private information corresponding to emerging indignation over corporations using and selling what is believed to be personal or private information. Couple that with the increasing number, volume and breadth of data breaches, and it leads to yet more interest in governmental intervention and protection. The directives for protecting "personal" information are raising awareness of general concepts of protection for any kind of "sensitive" data. Finally, once we start thinking about data governance for data protection (for the sake of regulatory compliance), it becomes critical to consider other aspects of governance that are

required for regulatory compliance.

In Chile, its legal development is found in the protection of private life, which configures, in this respect, a separate statute, with regulatory autonomy. In accordance with this, personal data are those relating to any information concerning natural persons, identified or identifiable.

CPLT is the body that will ensure compliance with the law of protection of personal information. The ARCOP Model tries to synthesize the fundamental activities that the user can do. This helps us because these activities can be related to the activities of our project, so we have to analyze them. Verify if any of the following point is touched by the project considered. And if we have to notify and discuss it with the Data Governance.

- A: right of **access** of the user to his information
- R: **rectification** that I can request from my data to Enel
- C: right to **cancel** or request that they erase the data they handle from me if I consider it convenient
- O: **oppose** the exchange of data without the client's authorization
- P: **portability** of the same data

Creation of Performance Indicators

We are always in phase 1 of our framework and our goal is the evaluation of Data Governance in every project of the department. This phase is very delicate as for each activity described above it is necessary to evaluate the level of completion of the same. This level of completion must consider several factors, in fact it is not enough that an activity has been done but it is also necessary to consider how, in what time, if it is unique etc. (Completeness, Uniqueness, Timeliness, Validity, Accuracy, Consistency). In fact, what we want to have are 18 final indicators that summarize the level of (actual) completion of the activity. For the creation of these completion indicators, reference was made to the model of quality metrics of the data provided by the same Enel Global (Completeness, Uniqueness, Timeliness, Validity, Accuracy, Consistency). For each activity, some of the indicators described in the metric were created, those that made sense. Subsequently, an indicator was created that could summarize the previous indicators created in the best possible way. At this moment we can

have a completion metric that can best reflect the degree of completion of our activity. We therefore obtain 18 indicators that provide us with a summary of the percentage of completion of each dimension. At this point there will therefore be activities such as: 100% (perfectly completed), 80% (almost completed), 60% (partly completed), 0% (not even started) etc.

The table below shows all the pillars, the dimensions belonging to the pillars, a short description of the activity of the dimension and the metric used to evaluate the % percentage of completion.

Pillars	Dimension	Metric Name	Metric Formula
Data Quality	Metrics	% metrics set weel provided	Validated metrics/total metrics
	Tools	% Tools correctly implemented described in the Dgprogram	Tools well implemented /total defined tools
	Processes	% processes well implemented/total processes	Well implemented processes/total processes
	Business Value	% of correct business value for each increment of the metrics	correct business value for increment of metrics / incrementing or diminishing of quality metric set
	Issues	% of issues identified	present and past quality issues solved from the last evaluation/present and past issues founded from the last evaluation
Meta-data	Profiles and Objects	% of valid info for the objects of the project	Valid objects/total objects

data

	Data Lineage	% of valid data lineage of the project	Valid data lineage / total data lineage
	Processes	% processes well implemented/total processes	Well implemented processes/total processes
	Tools	% Tools correctly implemented described in the Dgprogram	Tools well implemented / total defined tools
	Issues	% issues identified	Present and past metadata issues solved from the last evaluation/present and past metadata issues founded from the last evaluation
Data Access	Processes	% processes well implemented	Processes well implemented / total processes
	Roles and Permissions	% roles and permissions identified	Identified accesses for each identified roles / possible accesses for each possible roles
	Security	% security guidelines well implemented	Security guidelines well implemented / Total security guidelines
	Backup, Recovery, Retention and archiving	% Steps well implemented	Steps well implemented / total steps

	Issues	% of issues identified	Present and past access issues solved from the last evaluation/present and past access issues founded from the last evaluation
Principles, Policy, Regulation	Principles	% Data Ops followed steps	Followed DataOps steps / DataOps steps
	Policy	% Policy implemented	Policies well implemented / Total policies
	Regulation	% Considered regulation points	Considered regulations points / Regulation points

Table 6.1: Model Phase 1.

6.2.3 Phase 2 Prioritization Model.

Once the phase 1 is completed, the result we get are 18 indicators that give us the percentage of completion of each dimension. Our goal at this stage is to give an advice on prioritizing the activities not yet completed. This means that if all the activities are not 100% complete, which almost always happens, we want to give advice on which activities to prioritize and which are not. This data-back is provided by the Data Governance group to the project team. Obviously, this advice will have to consider the relative importance of each activity and at the same time the level of completeness achieved. We will then see how this feedback is released.

HOQ Modified and Adapted to our case

With the aim of finding the relative importance of each activity, we chose to take the HoQ quality house as a reference. This choice was based on the fact that the house of quality is an excellent tool that schematically connects functional

requirements with business requirements. This in view of a continuous collaboration with the business and at the same time in order to find the relationships of our activities with the business. We point out that the house of quality is not only a technical tool but represents the result of a series of phases, consisting of investigations, discussions and technical analyzes. We in our case use only the analytical part of the tool and at the same time we heavily modify it. In fact, we not only adapt it to our case by mapping various concepts that are actually different between them, but we also add algorithms that are useful for our purpose. Therefore, the dimensions of each pillar and the business requirements are considered as functional requirements. Subsequently in the correlation matrix we have seen the correlations between the business requirements and the dimensions to see the impact that each activity of the Data Governance has on the business.

HOQ

The House of Quality (House of Quality) is a graphic tool belonging to the Quality Function Deployment (QFD) methodology. This approach, still not widespread in Chile, makes it possible to translate customer needs into specific techniques to create or improve products and services. the QFD focuses on expectations and needs of the customer, in an attempt to reduce the possibility the expectations are then combined, through the effective representative synthesis of the House of Quality, with specifics, objectives and priorities in order to achieve the expected values. The house of quality is a matrix of several zones or "rooms". Its structure may vary according to the type of information to be represented. In principle, it consists of 8 rooms, each of which is used as a function. Through the activities of the house of quality it is therefore possible to overturn the business requirements directly on the process variables (functional requirements), focusing attention on the requirements that deserve greater attention and structuring one's own action according to priorities for action.

Creation of Business Requirements

Business requirements means all the requirements that the business communicates and converses with IT but it could also be seen as future strategic plans. In our case we created business requirements based on our personal experience as a trainee and comparing them with business and GDS stakeholders. We analyzed

various problems encountered, we derived the real goals that we can't reach because of these problems and finally we got the business requirements (Tables 6.2, 6.3 and 6.4). This is an important phase of the model as it goes to define the reasons why we will advise to perform one activity before another. Although it is a qualitative phase, it is important to follow an analytical protocol and try to associate our qualitative choices with quantitative reasons.

PROBLEMS:
1) Architecture problems (A the library is not controlled and neither is scalable B Today the Tools we have to compute are limited and are rarely bottlenecks C Lasado is the computation with the hdfs data volume if we want to increase the Computers would have to increase the machines 4 We have a 24/7 structure and there are cost problems and there are cost problems, we could parallelize but could not scale so much)
2) The agile step is difficult
3) It works many hours and few efficient
4) Do not check continuously with stakeholders
5) There is no transparency: open Company
6) Few security controls
7) Little team-working and little seems forced
8) Few people available
9) IT the bad guys of the movie
10) People do not know the advances of others

Table 6.2: From the problems

GOALS:
1) Efficiently use more type of computations which include Interactive Queries and Stream Processing, reducing the number of read / write operations to disk
2) Definitely implement Agile
3) Efficiency and better organization and work planning
4) Continually check with stakeholders - get it live immediately and receive real customer feedback (not management and PO can decide what works for the customer, it the customer who decides)
5) Transparency and open Company - Communicate work progress to others and detect connection points. Cooperation between different teams
6) Security guidelines and ensure compliance with laws and regulations
7) Mob programming - all work TOGETHER and show themselves the results
8) Scalar - scalar through SCALE methodologies (scaled agile framework)
9) Cooperation between business and IT (GDS go on)
10) Progress - Everyone shows their progress on the product instead of moving tickets or email around
11) Distributed teams - reduce amount of necessary communication by through an intelligent architecture with clean interfaces and restructure your organization
12) Continuously Identification of the most critical areas for improvement

Table 6.3: From problems to the goals

BUSINESS REQUIREMENTS:
1,10) Continuous improvement - Continuous innovation for the big data architectures that efficiently use more type of computations which include Interactive Queries and Stream Processing, reducing the number of read / write operations to disk
1,2) Efficiency and better organization for work planning - Permanently implement Agile and make agile step candidly
4) Live feedback - get it live immediately and receive real customer feedback (not management and PO can decide what works for the customer, leave it to the customer, .0 who decides)
5.10) Transparency and Cooperation - Communicate work progress to others and detect connection points. Work TOGETHER and show themselves the results Everyone shows their progress on the product instead of moving tickets around
6) Laws and regulation - Establish Security guidelines and ensure compliance with laws and regulations
7) Coordination and Distribution of Teamwork - Reduce amount of necessary waste communication through an intelligent architecture with clean interfaces and structured organization
8) Scale - scale projects and people involved through SCALE (scaled agile framework) methodologies
9) Cooperation between Business ad IT - Cooperation between business and IT (GDS go on)

Table 6.4: From the goals to the requirements

Identification of Business Requirements WWights

Then we need to identify the real weights of each business requirement. This step, depending on how we proceed, can be carried out by the business, by IT or better still in cooperation between the two. This step is very delicate as it is the input of what will subsequently be the weights of our dimensions. Interesting support is provided by the part relating to the comparison with competitors on business requirements (see section: Comparison of business requirements with other competitors). We can in fact see how our competitors behave in relation to business requirements with respect to another and finally decide to follow the strategic choices of the best competitors.

Identification Correlation Business Requirements and Functional Requirements

As an important step we find to define the degrees of correlation. The matrix of the degrees of correlation, in our case, allows us to see the relationships between the activities and the requirements of the business. The correlation matrix contains within it three qualitative symbols:

●, ○ and ▼

that are mapped respectively in numerical quantities: 9.3 and 1. In this way it is therefore possible to directly overturn the needs of the business, focusing the attention to the requirements that deserve more attention. This is established by the data governance team together with the business working together in close cooperation. Ideally it is a continuous working team aimed at meeting the most exact values possible of the correlation matrix. The part relating to the comparison with competitors on the business requirements gives us a slight support again.

Mapping of Degrees of Completion into Multiplicative Coefficients

An important part of our phase 2 of the model is the multiplicative factor. This part, as already mentioned, consists in modifying the relative weights of each Dimension according to the percentage of completion of the activity. It is therefore a question of considering how much the different % of completion affects: in fact, if an activity is completed 80% it does not have the same importance (for the same weight) of a 50% completed activity. It is therefore necessary to map each completion percentage with a multiplicative factor. And then this multiplicative factor will be used to further adjust the weights of our dimensions. This mapping as it is logical to think changes according to the activity. In fact, the rule that a half-completed activity has more importance than a completed 80% is not always true. Some activities work in binary and therefore only make sense if they are 100% completed and therefore an activity at 80% is more urgent than 20%. It is therefore necessary to differentiate according to the activity and the sense that is encountered in this mapping. In our specific case we used, for simplicity, an average mapping for all activities. We assessed that at a general level it is unimportant to complete activities with a high degree of completion, it is very important to complete the activities left in the middle and it is usually important to complete the activities just started. We therefore evaluated that on average the activities with a percentage of completion $< 30\%$ are medium to low in importance. Activities with a completion rate $> 30\%$ and $< 80\%$ have a high degree of importance and activities with a completion rate $> 80\%$ are of low importance. This operation is not present in the house of the original quality and has been added as well as others considering our specific objective.

Calculation of the Relative Importance Level of each Dimension

Our goal is to prioritize activities that are not 100% completed yet. In order to find this prioritization, we considered the weighted average of the different degrees of correlation between functional and business requirements. We must also consider the different degree of completion of each activity. This is an important variable because, as already mentioned, an activity completed at 80%

does not carry the same weight as a completed 20%. The multiplicative coefficient resulting from the mapping will be applied to the relative weight previously found and will change the same weight. The result will therefore be a vector of rating for each dimension considered. It will depend on us, then, to normalize these rating in order to translate everything into importance percentages.

Quality Roof

Optionally it is possible to activate a last phase in which the so-called "quality roof" is used. Therefore, by analyzing the degrees of correlation between the various dimensions, it is possible to further modify the relative weights. In our case, for example, the activities' rating has been further modified by adding for each activity the rating of the activity with which it correlates multiplies with the degree of correlation. As we have seen, the House of Quality was taken as a reference, but this has been extensively modified and adapted to our specific case. In fact, it was not adequate to take an ancient framework and apply it with blinkers.

Comparison of Dimensions with other Projects

An important part is the comparison of the percentages of completion of the other projects belonging to the same department. In this way it is possible to see the progress of colleagues and perform internal benchmarking. Furthermore, if other projects have reached a higher completion rate than ours, we can converse with them and cooperate to advance even with our % of completion. Maybe we can also ask for advice or suggestions on the practices used, tools etc.

Comparison of Business Requirements with other Competitors

Another important part is the comparison of the level of effectiveness of our business requirements with those of other companies. Obviously, this is not the purpose of our project and above all it is not analyzed by our department. But it can be used to decide the weight of the business requirements that we remember

being an activity to do together with DGgroup and business. In fact, if we see that we are very different from our competitors, this can be an important variable that allows us to increase the weight of a business. Furthermore, the DG group works in different fields and is a multidisciplinary program, so having an overall view of our competitors can be interesting.

Calculation of the Relative Importance Level of each Dimension

At this point it is possible to use our model by evaluating the degree of completion of the dimensions and prioritizing the activities not yet 100% completed. We would like to remember that to get to these results various important operations have been done. But these operations have all been justified as they are the result of prior studies of reliable sources and recognized frameworks. In any case, to further confirm the proposed model we will deal with justifying part of the choices made through tests that verify the effectiveness of the model. A small application of the model to some corporate projects of the Big Data department will take place and the results will be evaluated. The figure 6.6 shows the model used to process the second phase of the framework. Like we can see it's similar to the HoQ but if we deep look in it we modified a lot according to our goal.

6.3 Updating the Model

The updating of the evaluation and prioritization model is of fundamental importance as all the information becomes obsolete after a certain period of time, especially in the world of data where, within a few months, technologies, methodologies, frameworks etc. change. The ideal so it would be to update after a certain period of time set by the Data Governance Team or in some cases just before running the model.

6.3.1 Change of Business Requirements

In this phase it is necessary to re-evaluate the previous business requirements as these can be changed or obsolete for various reasons. The possible reasons are: obsolescence to pass the time, change of the management, change of the data strategy, organizational change, etc. This revision must be done periodically and, in the event, that reasons are found for which it is necessary to modify the business requirements, an activity similar to that of the creation of the business requirements is activated. Data governance together with the Business cooperatively analyzes all the needs, the problems of the department, the needs of the business, the data driven initiatives and future project plans. Through this analysis it comes to the business requirements.

6.3.2 Modification of the Related Business Requirements Weights

As in the previous step, with a view to continued cooperation with the business, we need to update the relative weights of the business requirements always in relation to the needs of the business, data driven initiatives and future project plans. Data governance being a multidisciplinary team with open doors a bit everywhere should be able to obtain the most reliable weights possible.

6.3.3 Modification of Activities

This phase is very delicate as the dimensions as already described have been found based on a deep analysis of the state of the art. It is therefore necessary to evaluate well if there is a real need for a non-previous DG activity considered or if we need to modify some of them. Each department has the freedom and

the duty to modify the activities chosen by the undersigned based on the specific requirements and the specific organization. But as already mentioned, being a delicate phase because the whole model changes, all the steps indicated in the creation must be followed. A technique to meet the activity that is obsolete could be to activate phase 2 of the model without considering the % completion of the dimensions. In this case we could consider the option of replacing the activities with % of minor importance with the new activities considered.

6.3.4 Interaction between Business Requirements and Activities

The degrees of correlation must also change over time and also in relation to unexpected events. These interactions must be re-evaluated by the Big Data department and the business. With a view to transparent, continuous and positive cooperation.

6.3.5 Update of Completion Indicators

The indicators of completion should be changed over time and also in relation to unexpected events. Some indicators may be obsolete and therefore should be reviewed through cooperation between data governance team and project team.

6.4 Use of the Model

The model described is used in the evaluation and prioritization phase of data governance activities related to a project. Below we will see at what point in the "Monitor, control and Evaluate" phase described in the chapter: "Data Governance program steps" we find ourselves. The activity diagram (Figure 6.7) will help us in this, which well describes all the phases that must be passed and the relationships between the same phases. The activity diagram also gives us an overall view of the actors involved and the relationship between them.

6.4.1 Activity Diagram

Before starting the project must implement the requirements described by the data governance necessary to start the project. We do not deal with defining

these requirements or assessing the effectiveness of these requirements. We focus on the next phase of the project. In this phase the project team will implement the 23 dimensions described by the 5 pillars. And the DGgroup will take care of evaluating the completeness of these activities. The model is then used on three occasions (Figure 6.7):

1. The project team announces that it has implemented the dimensions and is ready for an evaluation.
2. A timer expires that forces the DGgroup to evaluate the dimensions even if the Project team is not ready.
3. The project ends before the timer expires and then the DGgroup receives a notification of the fact and applies the model later.

At this point the Data Governance applies the model and based on the results of the model receives the green light to start from the data governance. Specifically, the model is applied and if the result of the application of Phase 1 and after Phase 2 indicates a $\% < 70$, then a prioritization board is given and we go back to the phase of implementation of the DG rules. Differently if the $\% > 70$ the evaluation activity ends even if it is possible to give the prioritization suggestions anyway.

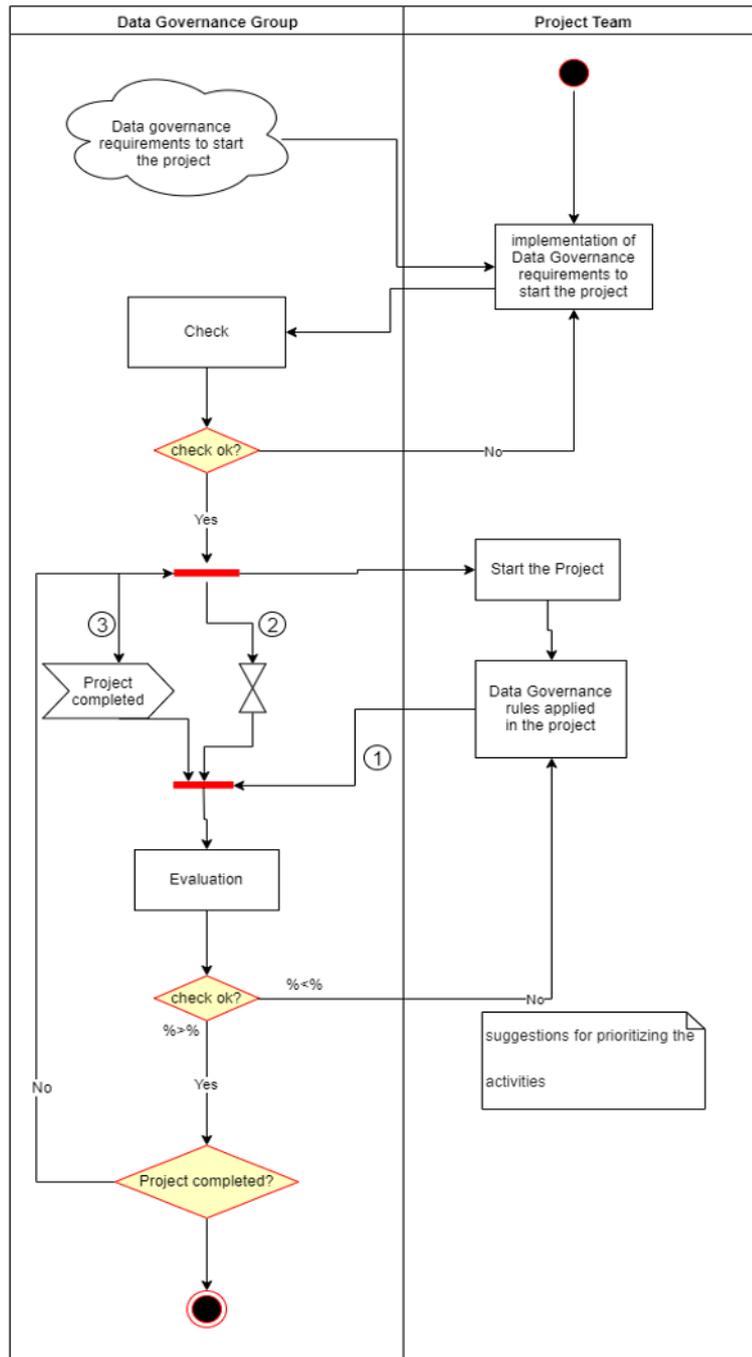


Figure 6.7. Data Governance Activity Diagram

6.4.2 Usage of the Model

The application of the model is carried out on a specific project and the activities carried out are evaluated. Phase 1 and then Phase 2 will be carried out as

described above. This chapter relating to the application of the model will be fairly concise as the big work is in the creation and updating phase. In the application phase it is sufficient to simply establish the numbers of the completion metrics, run the model and analyze the results.

PHASE 1

The DGgroup together with the Project team meets physically and / or remotely and establishes the numbers to be included in each indicator. If complaints arise at this stage, these are written and analyzed at a successive date, during the update phase.

PHASE 2

The model is executed with the data previously established in the creation and / or update phase. Subsequently the weights are inserted relative to the first phase of the framework and the sum of the importances of the dimensions relative to the completion percentage is calculated. If this number is lower than the sum of the ratings of the dimensions considering the 100% then it is necessary to go back to apply the DG rules and give a prioritization advice based on the results of the relative weight of each of our dimensions (end of phase 2). Moreover, if the sum of the importances of the dimensions relative to the completion percentage is much lower than the sum of the ratings of the dimensions considering 100% then an alarm bell is triggered which should immobilize the whole department. This is to analyze the situation well and understand what is not going and what comes from. This protocol is the result of the test performed on some corporate projects of the department of Big Data Governance. Specifically, we have analyzed some projects that went well and others that have gone wrong and we have seen what the number, from the point of view of Data Governance, is that, on average, indicates the good or the bad success of the project. Obviously to understand if a project will be good or bad there are numbers of variables to consider and that probably are not considered in Data Governance. Surely, however, Data Governance is one of the most important variables. In our specific case we have seen that this threshold is 40%, following the analysis made between working projects and non-functioning projects. It is not excluded that this number, depending on the department, can obviously change justifying through prior study and tests.

6.5 Test of the Model

Some tests will be described below in order to confirm some assumptions made in the model and also to practically apply the framework in some practical projects and see the results. The tests carried out below want to be a starting point for future tests and subsequent studies. It is hoped that the model will be used in other companies and departments where large amounts of data are processed. The department where we performed the internship has undergone several changes over the years with various strategic organizational changes dictated by Enel Italia. The Big Data platform has been around for about two years and in the meantime, there have been various changes to all the staff working in the department. The team we worked with has been stationed in this department for about 2 years so it has a limited portfolio of projects to work on. In fact, we are talking about 5 projects, some of which are completed, others in progress and others in kick-off. For this reason, the projects considered in the following tests are relatively limited but are useful as a starting point for analyzing model performance and future tests. We also would like to remember that the model has a strong theoretical basis based on a state of the art that is broad and full of recognized scientific publications. It is therefore a practical application to the department in which we work to show above all how the model is used.

6.5.1 Difference between two Projects in terms of Data Governance

In this test, two projects belonging to the Big Data platform were considered. The projects that were considered are two projects that have had two different and important results for our study. The first project (Data Archiving) had a negative result, which means that it was only in part that it was possible to reach the goals set at the beginning of the project. The second project (Care), on the other hand, was able to achieve the intended results and has the function of a good project. Within the same department, when starting a project, the Second project is taken as an example, as it describes the quality standards that all projects should follow in terms of: good practices, safety, protocols, etc. We will not enter in what the project specifications are, and we will treat the projects from a high level. Our aim is simply to analyze all the Data Governance activities performed by both projects and see the differences.

PROJECT 1

Data Archiving - Notifications (a project without base table and as MVP it is decided to build a board from the Raw Data layer, and then work it for a next stage on Governance criteria). They support salesforce objects within the data lake (they inject it to the data lake and free it from salesforce to free up space). Development of reports that allow them to monitor this information.

PROJECT 2

Format Care BI Latam, BI project on the field of Emergencies - Contactability - Notifications - Care. The objective of the project is to provide a unique, best of platform breed, which supports attention and emergencies through all channels and also facilitate the implementation of improvements or evolutions in existing capacities. The main benefits of the project are Minimize operating times in key processes for customer service and migrate interactions with the customer, whether by attention or emergencies, to low-cost digital channels.

6.5.2 Procedure of the Test

The analysis started by taking into consideration all the activities performed by both projects that could have a relationship with the Data Governance program. We then evaluated the differences between the two projects, considering which of the activities described in our framework were performed in one project and not in the other and vice versa. To find these activities all the documents produced by the project in the development phase were taken into consideration. PPT, WORD, Methodologies, Meeting histories, Data occupied in the incoming data lake, data occupied in the outgoing datalake, logs etc. In addition, focus groups where project managers of the projects and stakeholders were actively focused were made. Furthermore, all the procedures described in the "Use of the Model" chapter were considered. The results of these analyzes will be shown below.

1. Data Archiving has performed the following activities of our framework with the following completion rates:

Tools	0%
Issues	40%
Roles and Permissions	10%
Security	100%
Backup, Recovery and Retention	100%

Figure 6.8. Data Archiving Activities 1

2. Care has performed the following activities of our framework:

Tools	100%
Processes	25%
Business Value	50%
Issues	50%
Objects and Profiles	80%
Data Lineage	80%
Processes	25%
Roles and Permissions	10%
Security	100%
Backup, Recovery and Retention	100%
Principles	60%
Policy	25%

Figure 6.9. Care Activities 1

6.6 Test Results

From the Figure 6.8 we can deduce that 7 activities were performed in the project that went well and were not carried out in the project that went wrong. Furthermore, some of the activities carried out in both projects were done with different percentages. In fact, the activities done in the project that went well have always been completed with higher percentages compared to the project that

went wrong. Subsequently two indicators were calculated which give an overall picture of the percentage of activities completed in the two projects compared to the ideal situation. The first indicator makes a direct comparison between the two projects in terms of the level of completion of the activities. The second indicator is a comparison of the level of completion of the activities of each project with respect to the ideal level which therefore translates into two indicators each for each project. We want to remind that the ideal completion level is the same for both projects and will be the same for any project unless we update the model. With these two values it is also possible to see in terms of data Governance how much this value differs in the two cases.

1. Percentage constituted by the ratio between the sum of the importance of the activities completed by the Data Archiving project on the sum of the importance of the activities completed by the Care project. We therefore see that this indicator is 42%.
2. Percentage constituted by the ratio between the sum of the importance of the activities completed by each project on the sum of the importance of the activities completed ideally. We therefore see that for the Data Archiving project this indicator is 19% and for the Care project this is indicated as 44%.

Chapter 7

Conclusions

At this point we can say that we have created a tool to evaluate Data Governance activities for each internal project in the department and, also a tool to prioritize Data Governance activities not yet 100% complete.

Also, we have reached to support the Data Governance implementation because we followed the main guidelines of the Data Governance and we developed a framework that monitor constantly each project in the department in terms Data Governance activities. Also, we have seen that there is a strict correlation between reach the genera goal of a project and apply our framework.

Finally, we confirmed the hypothesis that one instrument of measurement of the activities of the Data Governance and the implementation of a prioritization of the activities in the backlog can increase the implementation of the Data Governance and provide all the benefits that the Data Governance provides.

We have therefore shown practically and through examples how to use the framework. We must consider that the framework should not be taken as a black box and used in any company department to evaluate projects. The basic concept of the framework, that is, the pillars and dimensions hopefully can be applied to the different areas of planning, but always with a critical eye and aimed at customizing the model. For example, when we define the completion metrics and when we describe the activities of each dimension, we must always consider and customize everything based on the department in which we find ourselves. The Data Governance team intelligently evaluates what to modify, insert or delete the framework depending on the application that must do it, the time or the situations in which it is located.

Based on the results of our work it is possible to affirm that it is possible to affirm that the Data Governance has a big impact in what is the success of a project. Furthermore, we must not forget that each of the indicators that we created in the test phase has a direct impact on the business requirements and that consequently has a direct economic impact.

We are aware that the data are not sufficient to justify the model but we have to consider that the department in which the internship was carried out is quite recent and consequently having history on whom to say this project went well and the other has gone bad is very relative. In any case, this is an excellent starting tool to effectively implement Data Governance. Moreover, we do not want to say that the success or failure of the project depends only on Data Governance but certainly, as already amply explained, it has an important role. We can therefore say that there is a large component of data governance activities performed (or in part) by the project that went well and not performed (or not sufficiently) by the project that went wrong. In this way we can therefore confirm that the activities that have been chosen are both the result of an in-depth study of the state of the art and the frameworks in circulation, and also the result of a small test aimed at verifying what was done previously. This study represents only a first step and above all focused specifically in the department of Big Data Governance.

It is hoped that, in the future, the following framework can be used by and properly modified by companies. Furthermore, we hope that we will carry out subsequent massive tests aimed at statistically consolidating the actual usefulness of the framework.

If we consider that each of the indicators, calculated in the Test chapter, has a direct impact on the business requirements and that the business requirements have a direct economic impact, we can put forward an idea for a future Test.

In fact, with a process of "Reverse engineering" we can go back and see how much they affect our business requirements. based on the degree of completion of our activities we could go back to the impact we have on our business requirements with respect to the ideal case. Subsequently, translating every business requirement in economic terms we can arrive at the economic delta that we are losing by not performing the activity as the norm. To do this test, we should have the percentage of completion of the activities of an ideally large number of projects

available and we should also have an estimate of the economic value that has to complete each business requirements.

7.1 Discussion

We have presented a data governance framework that can be used by practitioners to evaluate every project of their department from the point of view of Data Governance. This model can be used also as a starting point to develop a data governance strategy and approach for managing data as an organizational asset. Below we only want to give some guidelines for creating a Data Governance program. This is not the purpose of our project but following a careful analysis of the state of the art we feel we can give some advice. We have identified four decision domains, presented arguments for why each of these domains is important, described some key decisions to be made for each domain, and provided some examples of organizational positions that may be given accountability. We also have proposed a cooperation between different levels of organization, and we also want to affirm that when we talk about Governance we touch different areas of the company, so in some cases, shared decision rights may be appropriate for different decision domains in the same organization. Similar to Weill and Ross [36], we also suggest that a “one page” (Figure 7.1) design matrix may be useful for communicating a given organization’s data governance approach. The proposed framework also provides a common terminology that can be used by researchers to share their findings with other members of the information technology community.

Domain Archetype	IT Principles	IT Architecture	IT infra- structure Strategies	Business Application Needs	IT Investment
Business Monarchy					
IT Monarchy					
Federal					
IT Duopoly					
Feudal					
Anarchy					
Don't Know					

© 2004 MIT Sloan Center for Information Systems Research

Figure 7.1. IT Governance in One Page

We suggest also some activities in order to share some best practices that Data governance team should promote:

- Launch of Data room in every GBL/Function and staffing of roles per governance model defined or needed training
- Deep dive meetings and sharing meetings with data driven companies for top managers
- Program on entrepreneurship and data driven organization for managers and mid-managers
- Development of a communication plan to reach the company population, training pills and function/division ad hoc cultural awareness programs
- Migration to UVDL of at least 50% of Data sources
- Identification of 10 use cases per GBL/Functions (including minimum 3 use case cross GBL)
- Mapping of data needed for identified use cases in Data Catalogue and UVDL and development of needed analytics
- Next level reporting development of near real time reporting for each GBL

We must not forget that Data governance is about the three Ps: Policy, Process and People. An overall policy for how data is formatted and managed must come from an observation of how the business works. Data Governance not only dictates the processes of collecting and formatting data in a usable way, it is dependent on observation of business processes. It’s a good idea to have IT people involved in developing policy and data processes shadow the business units they’re supporting so they can see how data processes can support business processes. Each person must be able to speak the same language when it comes to the pillars of data governance, such as data quality and Metadata. We finish these thesis presenting four guidelines that we should ever follow. At first glance these concepts seem obvious but if we pay close attention to them and if we think about them, every company has small or big problems in the implementation:

1. Only work on what is needed
2. Set a deadline
3. Get the right people involved
4. Check back frequently with your stakeholders.

In order to ensure the framework is accurate, it is essential to gain community consensus for the Framework, for this reason we hope will become the basis for future work, studies, and applications.

7.2 The impact of protests in Chile’s organization and the Data Governance role

Having done the training in Enel Chile during the protests in Chile in 2019, (also occasionally called "The Spring of Chile", "Chile woken up", "Crisis in Chile 2019", "Revolt of October" and "Revolution of the 30 pesos ») allowed us to elaborate a profound reflection on how Data Governance should interface with this kind of situation. It seems to me, very important to make a little reflection, considering having lived in a historic moment for Chile, on what these kinds of situations have impacted organizations like Enel Chile. Furthermore, we will try to look at what is the role of Data Governance in these cases. We will not go into what is the policy of these events but we will see everything from an organizational point

of view. The wave of peaceful and violent demonstrations that led this South American country to face the greatest social crisis of the last three decades has left disastrous consequences in its "miraculous" economy. All this, in Chile, has disastrous economic consequences in small and medium-sized enterprises but also destructive consequences in large companies. After the social epidemic, many industries have been paralyzed. In this scenario of uncertainty, the economy is one of the most affected. And, in addition to the dollar, there are other numbers that keep Chilean economic authorities on the alert. One of these is the growth of gross domestic product (GDP). The low growth, meanwhile, can have devastating consequences for employment, this being one of the factors that most worries the Chilean authorities. Many companies have started to lay off some of their employees and today it is feared that by the end of the year up to 300,000 jobs could be lost. The Chilean economy is in an extremely complex situation, partly because it has a fairly massive paralysis of activity. Violence on the streets has hit big companies like Enel whose main building was located at the center of protests. Among the chaotic scene of the protests, the corporate building of ENEL, in the heart of the capital, was hit by a great fire. The fire spread through the building's stairs and made the entire main building of the company unusable, which has an impact on the organization of the multinational. All the employees have had to work for a long time between remote and physically. The projects were delayed and communication between people was incredibly affected. From a psychological point of view, it was also difficult because it touched on the routines of people used to physically going to the same place years and year ago and so on. Data Governance should therefore consider such situations and adapt effectively in case of need. The whole Data Governance program must therefore be reviewed considering these possible variables. Specifically, we must start from the 4 dimensions: Data Quality, Metadata, Data Access and Standards and make sure that each program considers this eventuality. The most affected pillars are (1) Data Quality, (2) Data Access and (5) Principles policies Standard. This is because analyzing the quality of the data from a high level is affected by unstable organizational situations that do not allow us to give the right level of attention. The access allowed to the data and the responsibilities assigned must be redesigned considering the unstable location and how people interact in an unstable environment. Finally, Principles policies and Standard are particularly affected as it is difficult to follow the principles defined by the

Data Governance program. We conclude this thesis further reinforcing concepts such as cooperation and transparency. These are concepts used many times in the before mentioned Data Governance Framework which, however, can also be adapted to the social situation currently being experienced by Chile. In fact, if the institutions, organizations and people cooperated more closely together through dialogue and greater cooperation, many problems would be avoided.

List of Figures

3.1	Business intelligence, analytics and governance in the Big Data Platform. “Image taken from GDS Market Southam”	16
3.2	High Level BigData platform architecture	20
5.1	Crisp-DM phases.	30
6.1	Data Governance in the Data Driven company Enel.	35
6.2	Relation between Data Governance Program and Data Governance activities of the projects.	39
6.3	Pillars of the framework.	40
6.4	Data quality metrics set. “Image taken from GDS Market Southam”	47
6.5	Example metric use case. “Image taken from GDS Market Southam”	47
6.6	Model Phase 2	71
6.7	Data Governance Activity Diagram	75
6.8	Data Archiving Activities 1	79
6.9	Care Activities 1	79
7.1	IT Governance in One Page	84

List of Tables

6.1	Model Phase 1.	62
6.2	From the problems	64
6.3	From problems to the goals	65
6.4	From the goals to the requirements	66

Bibliography

- [1] Constantiou, I.D. and J., Kallinikos, *New games, new rules: big data and the changing context of strategy*, Journal of Information Technology, 2015
- [2] Luftman, J. and T., Ben-Zvi, *Key issues for IT executives 2011: Cautious optimism in uncertain economic times*, MIS Quarterly Executive, 2011
- [3] De Haes, S. and W., Van Grembergen, *An exploratory study into IT governance implementations and its impact on business/IT alignment*, Information Systems Management, 2009.
- [4] McAfee and Brynjolfsson, *Big Data: The Management Revolution*, Harvard business review , 2012.
- [5] J. Alberto Espinosa, Frank Armour ,*The Big Data Analytics Gold Rush: A Research Framework for Coordination and Governance* , IEEE.
- [6] Rohinton P. Medhora, Susan Ariel Aaronson, *Data Governance in the digital age special report*, CIGI.
- [7] Dreibelbis et al., *Enterprise Master Data Management: An SOA Approach to Managing Core Information*, IBM Press,2008.
- [8] Ali Al-Badi, Ali Tarhinia, Asharul Islam Khana, *Exploring Big Data Governance Frameworks*, Procedia Computer Science, January 2018.
- [9] Patrick Mikalef, John Krogstie, *Information Governance in the Big Data Era: Aligning Organizational Capabilities*, Norwegian University of Science and Technology,2018.
- [10] Weber, Kristin, *A Model for Data Governance - Organising Accountabilities for Data Quality Management*, Alexandria 2007.
- [11] Khatri and Brown, *Designing Data Governance*, Communications of the ACM, 2010.
- [12] Abraham, Rene, Brocke, Jan vom Schneider, Johannes, *Data Governance: A conceptual framework, structured review, and research agenda*, International Journal of Information Management, 2019.

- [13] By Shannon Kempe, *Assessing Data Management Maturity Using the DAMA DMBOK Framework – Part 1*, Dataversity, 2011.
- [14] Galbraith, *The Star Model*, Jossey Bass, 2002.
- [15] Cheong Lai, Chang Vanessa, *The need for data governance: A case study, Proceedings - 18th Australasian Conference on Information Systems*, 2007.
- [16] McGilvray, *Data Governance: A Necessity in an Integrated Information World, Part 2 ; Look for those who have the ability to represent all data usages - not just those within their business area*, DM Review, 2007.
- [17] Weber E.U., *Decision under uncertainty. psychological, economic an neuroe-conominc risk of explanation*, Weber et al., 2009.
- [18] Boris Otto, *A morphology of the organisation of data governance*, 19th European Conference on Information Systems, 2011.
- [19] Tomi dahlberg, Tiina nokkala, *A framework for the corporate governance of data – theoretical background and empirical evidence*, Business, Management and education ISSN.
- [20] Dama International, *DAMA-DMBOK: Data Management Body of Knowledge: 1st Edition*, Dama International, 2010.
- [21] Dama International, *DAMA-DMBOK: Data Management Body of Knowledge: 2nd Edition*, Dama International, 2017.
- [22] Bitterer & Newman, *Organizing for Data Quality*, Gartner Research, 2007.
- [23] Izham Jaya, Fatimah Sidi, *A review of data quality research in achieving high data quality within organization*, Journal of Theoretical and Applied Information Technology, 2017.
- [24] Nick Bracy, Andrew Sadoway, *The DataOps Cookbook*, DataKitchen, 2019.
- [25] Peter Weill, Jeanne W. Ross, *IT Governance on One Page*, Massachusetts Institute of Technology (MIT) - Center for Information Systems Research (CISR), 2005.
- [26] Google Cloud , *A thought-leadership whitepaper*, Google Cloud, 2018.
- [27] Sas , *The 5 Essential Components of a Data Strategy*, Sas, 2018.
- [28] Khatri and Brown, *Designing Data Governance*, Communications of the ACM 2010.
- [29] Ironside How To Guide, *How to build a modern data and analytics Governance Framework - An Ironside How To Guide*.
- [30] Ernst & Young, *Identity and access management Beyond compliance*, 2013.
- [31] Seraphin Calo, Elisa Bertino, Dinesh Verma, *PolicyBased Autonomic Data*

- Governance* , Springer-Nature New York Inc, 2019.
- [32] Mauro Gonçalves, Mehran Misaghi, *Proposal of a Framework of Lean Governance and Management of Enterprise IT*, 2014.
- [33] Julian Ereth, *DataOps – Towards a Definition*, University of Stuttgart, Stuttgart, 2018.
- [34] Mit Technology review, *DataOps and the future of data management*, 2019.
- [35] Zeljko Panian, *Some Practical Experiences in Data Governance*, World Academy of Science, 2010.
- [36] Weill and Ross, *IT Governance: How Top Performers Manage IT Decision Rights for Superior Results*, 2014.

Sitography

- https://www.sas.com/en_id/insights/articles/analytics/a-framework-for-analytics-governance.html
- <https://www.pitneybowes.com/us/customer-information-management/case-studies/five-steps-to-smarter-data-strategy.html>
- <https://www.ironsidegroup.com/2018/06/07/five-essential-capabilities-master-data-management/>
- <https://www.ironsidegroup.com/2017/02/15/understanding-data-governance-tools-documents-processes-people/>
- <https://blog.aimultiple.com/data-governance/#best>
- <https://www.ibm.com/blogs/cloud-archive/2017/08/looking-data-governance-wrong/>
- <https://www.readitquik.com/articles/data/why-data-governance-is-crucial-for-big-data-environments/#>
- <https://www.pitneybowes.com/us/customer-information-management/case-studies/five-steps-to-smarter-data-strategy.html>
- <https://www.youtube.com/watch?v=KvGsyKtNEAk&list=PLVbsAdgZXvtyy6HVKCP0HChjC>
- <https://en.wikipedia.org/wiki/DataOps>
- <https://medium.com/data-ops/dataops-is-not-just-devops-for-data-6e03083157b7>
- <https://www.datascience.com/blog/what-is-dataops>
- <https://sdtimes.com/data/is-dataops-the-next-big-thing/>
- <https://searchdatamanagement.techtarget.com/definition/DataOps>

- <https://www.cio.com/article/3237694/what-is-dataops-data-operations-analytics.html>
- <https://blogthinkbig.com/dataops-tecnologia-datos-emergente>
- <https://streamsets.com/why-dataops/what-is-dataops>
- <https://www.slideshare.net/kellynpotvin/screw-devops-lets-talk-dataops>
- <https://medium.com/@raycad.seedotech/devops-methodology-and-process-dde388eb65bd>
- <https://theagileadmin.com/what-is-devops/>
- <http://www.getchef.com/blog/2010/07/16/what-devops-means-to-me/>
- <http://www.kartar.net/2010/02/what-devops-means-to-me/>
- <https://theagileadmin.com/2010/10/15/a-devops-manifesto/>
- http://itpi.3dcartstores.com/BookseBooks_c_11.html
- <http://velocityconf.com/velocity2008/public/schedule/detail/1525>
- <https://aws.amazon.com/es/devops/what-is-devops/>
- <https://caylent.com/theory-of-constraints-devops>
- <https://medium.com/rightdata/a-modern-enterprise-data-management-framework-44957d557874>
- <https://kanbanize.com/blog/kanban-vs-scrum-infographic/>
- <https://www.scaledagileframework.com/implementation-roadmap/>
- <https://www.scaledagileframework.com/story/>
- <https://kanbanize.com/blog/kanban-vs-scrum-infographic/>
- <https://www.planview.com/resources/articles/what-is-safe/>
- <https://www.scaledagileframework.com/>
- <https://www.softwaretestinghelp.com/scaled-agile-framework/>
- <https://www.planview.com/resources/articles/what-is-lean/>
- https://www.q-perior.com/en/fokusthema/the-q_perior-data-management-design-framework/

- <https://www.fayerwayer.com/2018/05/senado-aprobo-ley-proteccion-datos-personales-chile/>
- <https://aws.amazon.com/es/devops/what-is-devops/>
- <https://www.businessprocessincubator.com/content/tibco-spotfire-named-a-leader-in-the-forrester-wave-data-preparation-solutions-q4-2018/>
- <https://www.tibco.com/it/products/tibco-spotfire>