POLITECNICO DI TORINO

Faculty of Engineering

Master of Science in Engineering and Management

Master's Thesis

Data quality and data governance in insurance corporations

Supervisor
Prof. Elisa Ughetto

Candidate
Michele Chiantera

Academic Year 2022/2023

“ Non est ad astra mollis e terris via.”

(Lucius Annaeus Seneca, *Hercules furens*)

Abstract

This paper will focus on the topics of data governance and data quality in the insurance industry. Nowadays, in fact, more and more companies are choosing to invest in order to have a solid data analysis and management structure, in order to ensure better customer service, cut costs, and have more robust information available to prepare financial statements.

In addition, within the paper, an attempt will be made to bring out the idea that data governance and data quality are closely related, and investment in one area implies an eye for the other; consequently, although they are distinct, they are "two sides of the same coin." Since the information processed and managed by these companies is purely financial in nature, the first part of the paper will be a general overview of the general principles of insurance financial reporting and international accounting standards; in particular, there will be a special focus on IFRS9 and IFRS17, as they are more closely related to the insurance context and are of particular interest to the project covered during the internship period.

The following parts of the thesis will focus on data governance and data quality in general, using scientific literature as a source to provide benefits, costs, challenges and key aspects.

Finally, this information will be useful for a better understanding of the last chapter, which focuses particularly on the work I did during my curricular internship at EY, concerning a project for a client operating in the insurance and asset management sector. The information provided in this section will not be extremely detailed (due to confidentiality constraints on sensitive data and the identity of the customer) but it will still be enough to relate what was described in the previous part and to understand why insurance companies need to have a solid data governance structure linked to a solid data quality policy.

Table of Contents

1. Data and reporting in insurance corporations

1.1. Introduction: is the insurance sector data-driven?

Although the insurance market is known to be one of the most traditional and conservative sectors in the use of innovative technologies, in recent years, more and more companies are adopting a data-driven approach, based on the collection, analysis and interpretation of data; in this regard, it can be stated that the goal of this change is to increase operational efficiency and provide customers with increasingly personalized service. A report published in March 2023 by Capgemini Research Institute shows that, globally, 40% of insurance companies currently use data to access new markets, while 43% have modernized their risk management algorithms.

However, a survey conducted in 2023 by IVASS[1] and reported by *Il Sole 24 Ore*, shows that insurance companies have not yet fully adopted such tools (only 27 % of corporations use at least one machine learning algorithm in processes involving direct effects on customers[2]). In addition, in the insurance sector, frequencies of internal data collection and updating are quite rare, therefore, data-enrichment processes over internal data and "external" sources (which are integrable in real time) assume great importance.

Currently, many firms in the sector are increasingly investing in data-driven technologies, such as artificial intelligence (AI) and machine learning (ML), in order to improve their ability to analyze large amounts of data and improve their decision-making processes; nevertheless, there are still many challenges to be faced, such as the lack of technical expertise and the difficulty of integrating data from different sources. According to Capgemini Research Institute only 18% of companies have the technical capabilities, culture and business practices to support data-driven programs; the consequence is the difficulty of making the most of the growing volume of data, with inevitable consequences in terms of efficiency, service quality and customer satisfaction.

The Capgemini Research Institute report also highlights ways in which the data-driven approach is a source of competitiveness that can make a real difference in the insurance industry.

---

[1] *Istituto per la Vigilanza delle Assicurazioni,* Insurance Supervisor Institute.
[2] The survey focuses on the Italian scenario; it was commissioned to University of Milano-Bicocca and Doxa, and funded by MISE.

A major benefit is a greater understanding of the risks associated with specific activities, customer behaviors, market trends, and weather conditions that can affect the likelihood of a claim. This information can help insurance companies create "tailor-made" policies for their customers that cover exactly certain risk categories, avoiding the overpricing or underpricing of certain situations, which affect the company's profitability.

Other benefits of the data-driven approach in the insurance industry are the possibility of fraud prevention and better claims management in the market; it is also possible to identify new markets, offer innovative products based on the data collected, and optimize business processes, using data to make conscious business decisions. The figure below shows, as stated earlier, that only one in five insurance companies can easily and profitably use a data-driven approach.

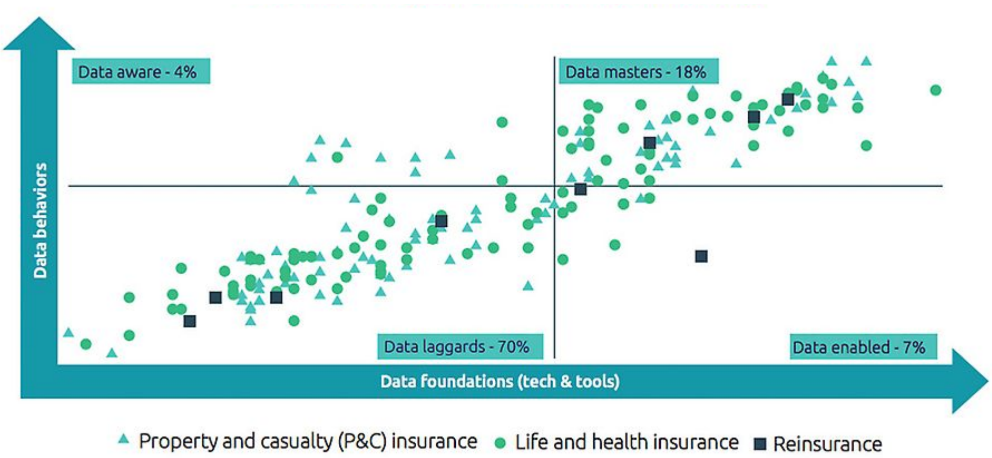

Fig. 1.1. Data maturity of the insurance industry (source: Capgemini Research Institute)

Companies in the insurance sector and, in general, in the service and customer-oriented world are making increasing investments in data management, albeit at different speeds. In particular, according to Busch (2020), effective data governance will become increasingly crucial in the future in order to link the service offered by a company to the actual needs of its customers; this factor, in the insurance sector, may therefore translate into greater meticulousness in the information reported on financial statements, as well as a consistency of sensitive data on customers and the services offered (such as, for example, the type of insurance cover). These factors are confirmed by several trends, which link investments in data governance to the development of new business models, such as UBI (a type of insurance based on the actual use of an asset rather than its estimated use) or a more self-service-based insurance model (with customers expecting to autonomously choose a more immediate and direct service, without the intermediation of a live operator). These data are confirmed by an EasySend report from 2023, according to which, in many cases, it is precisely customers who “anticipate” companies and push for more efficient information management; according to this report, 89% of insurance company customers expect a substantial increase in the forthcoming digital services offered and the efficiency (and transparency) with which sensitive financial data are processed.

However, in order to ensure effective data governance, the role of an adequate data quality policy is essential; the latter is an important pillar in the data quality framework, as information cannot be managed by companies if it is incomplete or inconsistent. Unlike data governance (focused on the management aspect of the information and the company as a whole), data quality is related to the integrity and intrinsic value that the data want to convey (Smith, 2021); nevertheless, data quality is a cornerstone of effective data governance, so much so that in many insurance companies, the units that dictate data governance guidelines are often integrated within the data quality offices themselves.

Moreover, testifying to the close correlation between the two components is the fact that investments in information management and governance “go” first through the data quality structures; according to the 2018 Gartner Data & Analytics Summit, in fact, companies consider it a priority to invest in data quality as the management of erroneous or inconsistent information costs companies in the financial sector losses of $15 million per year.

According to M. Goetz (2014), companies also consider it important to invest in this direction in order to increase productivity, as one third of analysts say they use 40% of their working time to validate and check potentially incorrect information before it can be used in decision-making. Consequently, all trends seem to confirm the link between data governance and data quality: *you cannot have the first one unless you start investing in the second.*

Finally, considering that in the insurance sector most of the data processed and handled is financial and inherent to the financial statements, it is important to know, first of all, how the financial statements of these companies are prepared; in this regard, the international financial recording standards, which will be discussed in the following paragraphs (focusing in particular on IFRS9 and IFRS17, as they relate to the insurance world, the concepts of data governance and data model, and the final project carried out during the internship period), assume particular importance.

1.2. Accounting standards: some historical background

Considering what stated above, it is not surprising that the need to use a data-driven approach affects insurance companies' accounting ways, and the necessity to work with robust, up-to-date and consistent information. Therefore, to effectively introduce the concept covered, it is necessary to present the basics and main features of insurance accounting.

Speaking of a general context, in the US all corporate accounting and reporting is ruled by a common set of standards, known as Generally Accepted Accounting Principles, or GAAP, set by the independent Financial Accounting Standards Board (FASB); the latter is a 14-member committee based in London that works primarily on the drafting of accounting standards and the convergence of the various national accounting standards spread around the world. Another committee is the IFRIC (International Financial Reporting Interpretations Committee), which periodically monitors the application of standards by suggesting their correct interpretation and proposing the most appropriate treatment for cases not covered by the standards. The Standards Advisory Committee (SAC) is an advisory committee composed of members from around the world that is responsible for making proposals to the IASB on activities to be carried out. These three entities form the IASC (International Accounting Standards Committee) foundation, which is responsible for issuing IAS (International Accounting Standards); the IASC foundation also consists of an administrative body (the so-called Trustees), a 22-member committee that appoints the members of the IASB, IFRIC and SAC, approves the foundation's budget and decides its overall strategies.

The EU issued a series of so-called "endorsement" regulations[3] in the early 2000s to regulate the concrete application of IAS/IFRS in the European legal system; in particular, with Regulation No. 1606 of 2002, the European Union made the adoption of international standards mandatory in the consolidated financial statements of listed companies starting with the financial statements of the fiscal year as of January 1, 2005, as well as for banks and insurance companies. Italy extended the obligation to the annual financial statements of the same corporations for the year 2006 (in addition to the option for the consolidated financial statements of the other corporations starting with the financial statements of the year 2005).[4]

[3] Regulation (EC) No. 1606/2002, which was followed by Regulation (EC) No. 1725/2003.
[4] Legislative Decree No. 38 of 2005, "Exercise of options under Article 5 of Regulation (EC) No. 1606/2002 on International Accounting Standards".

In 2001, the IASB replaced the IASC with the task of achieving convergence among national accounting standards through the development of global accounting standards; so, the committee began to work on International Accounting Standards and International Financial Reporting Standards (IFRS, will be discussed more in detail in the following paragraphs). At the same time, the European Union (EU) began to work on Solvency II, a framework directive to simplify and strengthen solvency requirements throughout the EU in order to create a single insurance market; the thinking behind this move was that a set of universal accounting standards would facilitate global capital flows and reduce the cost of raising capital. In the current scenario, about 100 countries require or allow the application of international standards developed by the IASB. In 2021, during the COP26 of the United Nations Framework Convention on Climate Change in Glasgow, the IFRS foundation announced the formation of the new International Sustainability Standards Board (ISSB).

It should be underlined that IAS/IFRS accounting standards are not instantaneously applied in the European Union, as they undertake an preliminary technical review by a committee of experts called the European Financial Reporting Advisory Group (EFRAG) and a policy assessment by a committee of government representatives called Accounting Regulatory Committee (ARC). For EU endorsement, the document must also pass scrutiny by Standards Advice Review Group (SARG), appointed by European Commission Decision 2007/73/EC, whose function is to advise the Commission on the neutrality of EFRAG's opinions. When the last checks are successfully passed, the accounting standard is approved through regulation by the Union ministers and takes immediate legal effect in all member states.

In the United States, the Securities and Exchange Commission (SEC) requires companies filing financial statements to follow GAAP or IFRS, depending on whether they are US issuers or foreign private issuers. Over the years, the FASB has evaluated and partially aligned its standards with International Financial Reporting Standards through a joint project or decided in other cases to not follow them.

Those previously described are just the innovations introduced in recent years, but accounting standards have evolved steadily over time. Historically, prior to the 1930s, corporate accounting and reporting focused on management and creditors as end users; since then, investors have increasingly used GAAP in order to evaluate and compare financial performance from period to period and across companies. In addition, GAAP emphasized "transparency," meaning that financial statements and reports must be understandable to aware people, the data included in financial statements must be reliable, and companies must fully disclose all relevant and meaningful information. In this regard, scientific literature (Stein et al., 2017) entrusts the concept of "transparency" with the task of constructing meanings that give sense and rationality to complex, ambiguous, and uncertain events such as accounting scandals and subsequent financial crises; in order to restore credibility to financial markets, regulators should therefore offer viable alternatives to senior managers' reports on transparency and financial reporting.

In addition, special accounting standards have been developed for industries that have a fiduciary responsibility to the public, such as banks and insurance companies. In order to protect insurance policyholders, in fact, state insurance regulators began to monitor the solvency of insurance companies. This led to the development of special accounting standards for insurance, also known as SAP (Statutory Accounting Principles and practices); in this regard, the term "statutory accounting" means that SAP incorporates practices prescribed or permitted by law; thus, the main purpose of SAP is to provide information about an insurance company's solvency, and it focuses more on the valuation of assets and measurement of liabilities on the balance sheet using more conservative criteria than GAAP. In the US, for example, insurance corporations report to SEC using GAAP, but report to insurance regulators and tax authority using SAP.

1.3. Insurance accounting

In general terms, it is relevant underlining that insurance corporations bear and manage risk in exchange for a premium. The computation of premiums varies dependently to policies, also according on past data grouped from analogous contracts. In this regard, it can be easily understood that the real expense of each policy to the guarantor is unknown until the conclusion of the deal period (actually, in the case of particular coverage services the cost is a question mark even after the termination of the policy period), when the price of claims can be computed conclusively.

According to collective knowledge, the insurance industry can be splitted into two sub-sectors: property/casualty (general or non-life insurance) and life/health. Generally, property/casualty policies cover homes, auto, and businesses; life/health insurance includes life, long-term care and disability insurance, annuities, and health insurance. It should be noticed that insurers present financial reports to state regulators by means of compulsory accounting standards, but significant differences between the accounting practices of property/casualty and life insurers due to the nature of their products exist(e.g., the duration of the insurance contract).

Normally, a lengthy and detailed statement constitutes the annual financial statements of an insurance company is recognized. In this regard, a typical insurance balance sheet is characterized by the peculiar pattern of cost-effectiveness, but there are many critical parts; these are often related to the convergence process between environmental changes in the reference context, supranational supervisory regulatory requirements (e.g., the Solvency II Directive) and new international accounting standards (Cappiello, 2016).

In legal reporting principles, the primary segment comprises a balance sheet, an income statement, and a section known as the Capital and Surplus Account; the latter states the main policyholders' surpluses and changes in the account during the year. Generally speaking, balance sheet delineates a profile of the company's financial position at a given point in time, while income statement records the company's operating results occurred during the fiscal year. An insurance company's policyholder surplus (defined as its assets minus its liabilities) serves as a hedge against catastrophic losses and to finance expansion; in this regard, regulators require that insurers have sufficient surplus to support the policies they issue. The greater are the risks assumed, the greater is the amount of policyholders' surplus required.

Thus, it can be said in general that the main characteristics of an insurance company's financial management are related to a high concentration of investments in real estate, government bonds or other bonds otherwise characterized by low risk; however, in recent years, a gradual increase in the use of structured and/or derivative finance instruments with higher risk content can be observed (Torchiani, 2023).

As already mentioned, the investment objectives for an insurance company are profitability (preeminent in life insurance) and liquidity (preeminent in non-life insurance), to be followed through ALM[5] logics (increasingly relevant also in consideration of IVASS

[5] Assets and Liabilities Management (ALM) is a model introduced in 1970s for measuring for all banks' financial operations the level of rate risk and making explicit the potential for loss or profit from fluctuations in market rates. The business analysis methodology is widespread in the banking and life insurance world, while it is less so in other sectors, such as non-life insurance and pensions. In order to provide values that are compatible with the definition of the economic value to be represented, an ALM model must be useful, meaningful, implementable, able to provide values consistent with all asset and liability classes, calibrated according to price observation, uncontroversial, specifiable, and *auditable*, which means suitable for review (Santomero & Babbel, 1997).

measures and regulations, including Regulation No. 36 of January 31, 2011, concerning guidelines on investments and assets covering technical provisions).

The life insurance dealing cycle, unlike that peculiar to non-life insurance, has a extended duration; the typical investments of a life company are therefore long-term, safe and sufficiently profitable investments, guided by ALM logic.

The guiding values of ordinary financial statements are also to be emphasized; these principles apply to the financial statements of all organizations, so they also extend to insurance companies. In the Italian context, the preparation of annual financial statements is regulated by Article 2423 of Civil Code. Among the principles, the most relevant are:

- *Clear, true and fair representation of the reported data*[6]; this is related to a direct exposition of the parts that make up the operating budget. The main purpose is to make the data in the financial statements easily comprehensible for the goal of easier reading and understanding by corporate stakeholders. The financial statements should also cover analytical data and be complemented by information in the explanatory notes that be able to ease understanding of the totals shown; from a practical point of view, only reliable information that has a significant effect on decision making should be showed.

- *Prudence and business continuity*[7]; the firm is normally considered to be able to continue its operations in the foreseeable future. This standard requires that the values recorded in the financial statements are considered on the assumption that the company will continue its business in its normal course, with no intention or need to place the company in liquidation or cease operations or subject it to bankruptcy proceedings. The principle of prudence also requires that the valuation of financial statement items is conducted using caution in estimates made under conditions of uncertainty; specifically, only profits realized as of the end of the fiscal year can be disclosed, and losses affecting the fiscal year should be considered even if they are recognized after the end of the fiscal period.

---

[6] *"The financial statements must be drawn up clearly and give a true and fair view of the company's financial position and results of operations for the year"* (Art. 2423 (2), Civil Code).
[7] *"The valuation of items should be made prudently and with a view to the continuation of the business"* (Article 2423a (1), Civil Code).

- *Accrual principle*[8]; accrual is the time basis by which elements should be recognized. Accordingly, they are attributed to the fiscal year to which they are referred, not to the year in which the corresponding receipts and payments materialize.

- *Prevalence of economic substance over legal form*[9]; operating events relevant to the company are recognized with direct reference to their economic substance, and not only in relation to their legal form. If substantive and formal aspects do not coincide, all information necessary to express the economic substance of the transaction should be provided in the notes to the financial statements.

- *Separate evaluation of assets*[10]; Compliance with this principle means having the obligation to specify separately in the notes to the financial statements the items being grouped; the goal is for recipients to have evidence of the assets in the production process (a principle applicable mainly in manufacturing companies).

- *Continuity in the preparation of financial statements*[11]; this principle provides for continuity of application over time of the accounting and valuation criteria, to facilitate the comparison of results achieved in different periods. Any changes must be justified and described in the supplementary note.

The main items of an insurance balance sheet and their evaluation criteria are described below.

---

[8] "*Account shall be taken of income and expenses pertaining to the fiscal year, regardless of the date of collection or payment; account shall be taken of risks and losses pertaining to the fiscal year, even if known after the close of the fiscal year."* (Article 2423a (1), Civil Code).

[9] *"The evaluation of items must be made [...] taking into account the economic function of the asset or liability under consideration"* (Article 2423a (1), Civil Code).

[10] *"Heterogeneous elements included in individual categories should be evaluated separately"* (Article 2423a (1), Civil Code).

[11] *"Evaluation criteria cannot be changed from one fiscal year to the next"* (Article 2423a (1), Civil Code).

1.3.1. Assets

For insurance companies, asset management is the collection and the maintenance of listed and unlisted financial instruments over time; as mentioned earlier, the goal is to get the best possible return associated to a particular risk threshold. It is partially constructed on diversification by assets and geographic areas, partially on timing choices and the ability to dynamically change the composition of the portfolio.

According to the 12th annual insurance survey compiled by Goldman Sachs Asset Management, titled Balancing Performance and Inflation Uncertainty, it is stated that environmental, social and governance (ESG) factors remain at the forefront of portfolio considerations. As also shown in the figure below, 90% of insurance companies take these aspects into account in their investment process; however, a growing percentage of corporations (68%) believe them that it is rising yields that will have the greatest impact on their asset allocation decisions in the coming years.

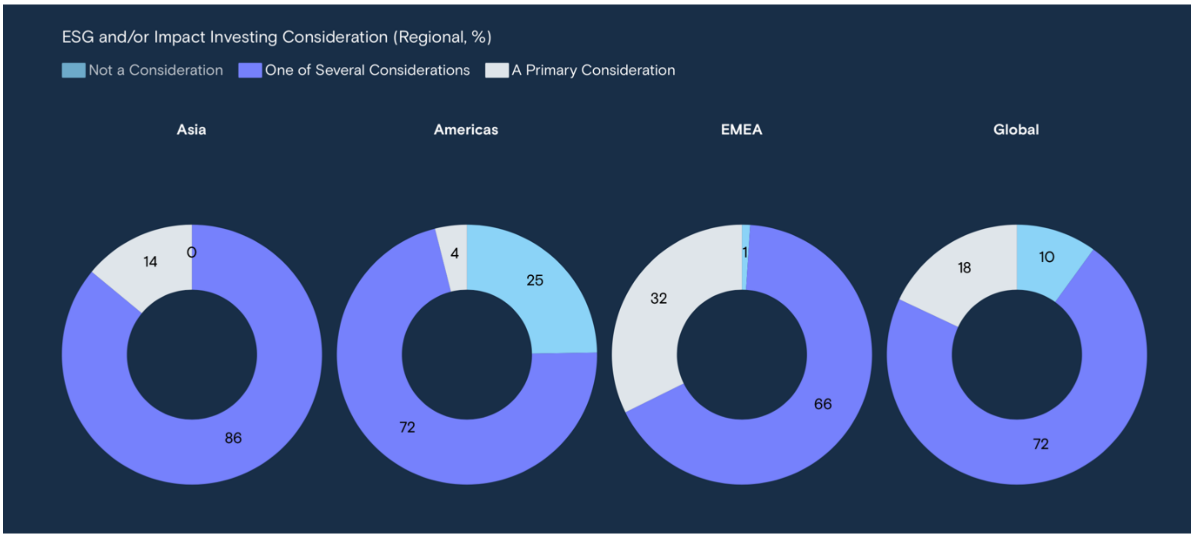

*Fig. 1.**Errore. Nel documento non esiste testo dello stile specificato.**2. Impact of ESG principles on insurance companies' asset management (source: Goldman Sachs Asset Management)*

According to the Insurance Information Institute, most of insurance firms' assets are highly secured government and corporate bonds that pay an income and are generally held to maturity; this factor is related to companies' need to be able to pay claims promptly and to raise cash quickly in case of unforeseen events, disasters or natural catastrophes.

Under SAP, bonds are valued at amortized cost rather than current market cost; such evaluation criterion ensures that the value of bond holdings remains relatively stable over the years and reflects the expected use of the asset.

Yet, when predominant interest rates are higher than bond coupon rates, amortized cost overestimates the value of the asset, producing a value higher than market value. Using the amortized cost method, the difference between the cost of a bond at the date of purchase and its face value at maturity is recognized by gradually changing the bond value; the result is an increase in value over the original price when the bond was purchased at a discount, and a decrease when the bond was purchased at a premium.

According to GAAP, on the other hand, bonds can be evaluated at market price (if the insurer is planning to hold them until the maturity date) or recognized at amortized cost (if they will be made available for sale or active trading).

A very important asset category for property/casualty companies are preferred and common shares, which are valued at market price; in this regard, life insurance companies generally hold a small percentage of their assets in preferred or common stock. Another class of assets are reinsurance receivables, which are defined as the amounts owed by the company's reinsurers; the latter are insurance companies that insure other insurance companies in order to share the risk of loss. If the receivables prove uncollectible, they reduce the surplus and are reported on the liability side of the balance sheet.

It should also be underlined that some assets are not documented under SAP, although they are recorded under GAAP. Another special topic concerns fixed assets; real estate and loans usually represent a small part of a property/casualty insurance company's business because they are moderately illiquid. On the other side, life insurance companies, whose liabilities are long-term commitments, have a larger share of their investments in residential and commercial mortgages.

1.3.2. Liabilities and capital

Academic literature distinguishes the liabilities of an insurance company into obligations to policyholders and claims from creditors.; the first component is considered the most relevant.

Specifically, the Insurance Information Institute, classifies the reserves (necessary to meet obligations to policyholders) available to insurers into two classes:

- *Reserves for unearned premiums*; from a technical standpoint, they represent the company's liability for unexpired insurance coverage (expiration is referred to the policy period). As consequence, they are the part of the policy premiums that the company would have to return to the policyholders, in case the firm ceases its operations, or customers decide to drop the insurance coverage[12]. This amount, subsequently, is not formally collected by the corporation until the insurance has expired.

- *Reserves for claims and adjustments;* these are defined as the obligations the insurance company has incurred as a result of claims that have been (or will be) reported for exposures protected by the insurer. The provision of these funds is necessary for the company to pay adjusters, legal counsel, investigators, and incur other expenses associated with the settlement of claims. In this regard, the estimation and settlement of some claims, such as fire damage, proves to be faster. Others, however, such as some workers' compensation claims, may have settlement times that extend long after the policy expires; in particular, the most difficult losses are related to events that have already occurred but have not been reported to the insurance company, known as "incurred but not reported" (IBNR). Examples of IBNR loss cases include workplace accidents that occurred due to inhalation of substances harmful to health, but were not reported until years later, following the diagnosis of an eventual illness. Claims costs, including IBNR claims, are usually estimated basing on insurers' experience. Later, as companies collect more data and statistics on accidents over the years, reserves are adjusted, with a corresponding impact on profits.

Reserve management is one of the main drivers of an insurance company's profits; however, the main sources of profit are investment results and policyholder premiums.

---

[12] In accordance with regulations, if the policy is cancelled before maturity the customer is entitled to a refund of part of the premium paid in advance. In Italy, the customer's right to terminate the insurance contract, is regulated by Act 40/2007, according to which all multi-year insurance contracts, including policy contracts, can be terminated at any time as long as there is prior notification. Article 21 from Act 99/2009 also stipulates that insurance companies, in addition to being able to enter multi-year contracts, if they provide a discount on the contract linked to the expiry of the policy, may require that the policy not be cancelled before five years have elapsed since it was signed.

Under GAAP, profit is generated consistently over the life of the contract; in addition, policy acquisition expenses are deferred and expensed on a pro rata basis, usually in line with premiums earned.

Under SAP, on the other hand, policy acquisition expenses are recognized as costs as soon as the policy is issued, while premiums are collected over the life of the policy.

1.4. International Financial Recognized Standards

Having set out the basics of insurance accounting, it is necessary to describe in more detail the International Financial Recognized Standards (IFRS), which have already been mentioned above; these standards were issued by the IASB with the aim of promoting transparency and ensuring that financial statements drawn up in different countries are comparable according to a single internationally recognized set of rules.

This body of standards gradually replaced the International Accounting Standards (IAS) that had been in place since 1973. As written in the preceding paragraphs, the European Union started working on IFRS in 2002 and definitely introduced them in 2005; in 2006, the standards were also ratified by Italy.

Nowadays, according to data published by *ESG360* in February 2023, IFRS principles are adopted by 143 countries, including 98% of European countries and 92% of Middle Eastern countries.

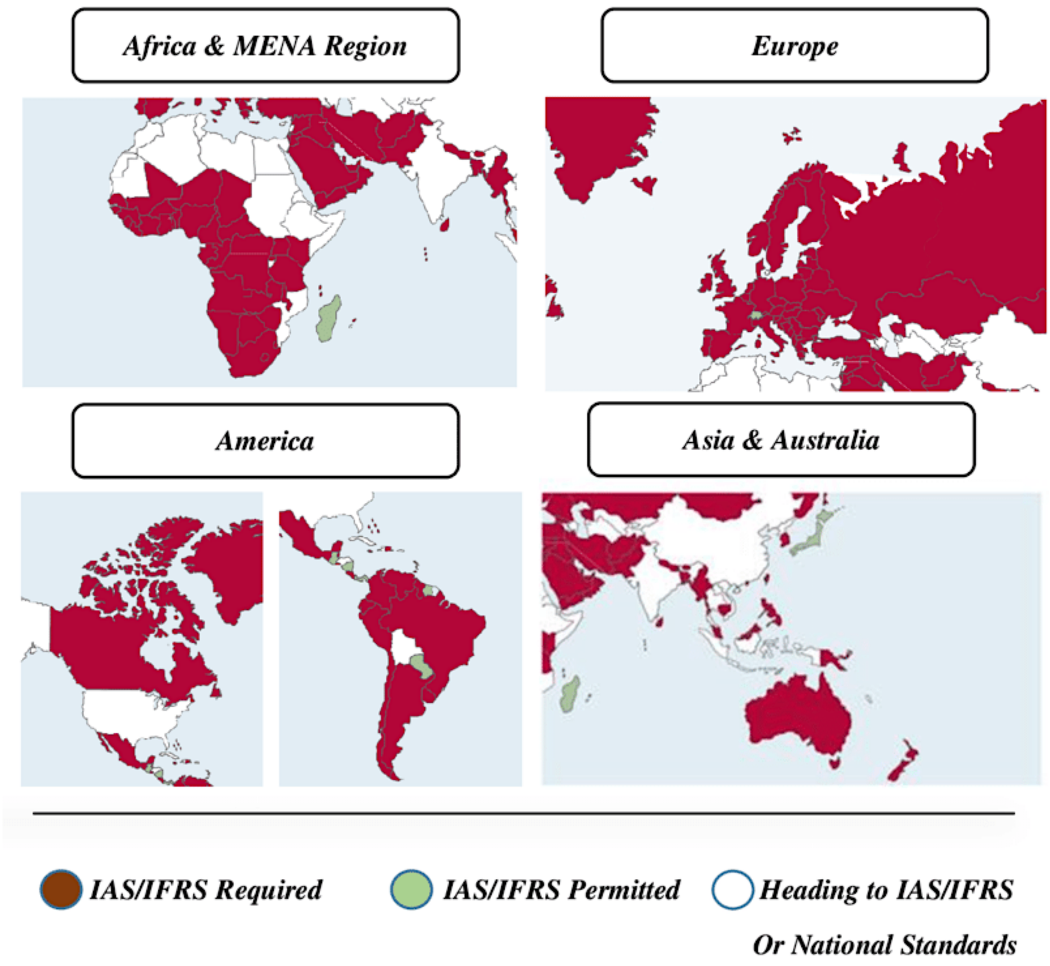

Fig. 1.3. IFRS/IAS adoption around the world in 2022 (source: ResearchGate)

Current regulations stipulate that the obligation to use the standards set out in the IFRS to present financial statements primarily concerns all European listed companies and large international companies operating in the European Union. This obligation also applies to banks and financial intermediaries subject to supervision, companies issuing popular financial instruments, unlisted insurance firms with reference to consolidated financial statements only, and listed insurance corporations; SMEs, meanwhile, have the option of using a simplified standard, the so-called "IFRS private entities" or "IFRS SME", but implementing the principles is not mandatory.[13]

[13] In Italy, the 2019 Budget Law No. 145, that came into force on January 1, 2019, defined the entities that have the option, not the obligation, to apply the standards. Article 2 of Legislative Decree 38/2005 indicates the entities required to apply international accounting standards.

Companies not included in the aforementioned list may continue to apply their domestic regulations (including the rules of the civil code); however, according to the Italian *Fondazione Nazionale di Ricerca dei Commercialisti* (National Research Foundation of Accountants), these regulations will be amended in the next years to bring them into line with international rules.

Compliance with IFRS allows companies, especially the listed ones, to compete in international markets under the same accounting rules as their competitors, thus becoming part of an international market that offers several advantages and growth opportunities; on the other hand, these international standards are also drawn up with the aim of providing investors and stakeholders with understandable, relevant, reliable and accounting-relevant information.

In addition, the use of standards may allow companies with subsidiaries in countries that require or permit the use of IFRSs to use a single accounting language for the entire company.

On the other hand, it seems clear that companies may also need to convert to IFRS if they are subsidiaries of a foreign company that must use IFRS or if they have a foreign investor that must use IFRS. Finally, corporations may also benefit from using IFRS if they are willing to raise capital abroad.

However, as can be seen above from Fig. 1.3., several corporations, especially in the US, consider GAAP as their primary framework, as they believe that with the full acceptance of IFRS, a certain level of data quality may be lost. Furthermore, according to the AICPA (American Institute of Certified Public Accountants), some US issuers that do not have significant customers or assets outside the US may oppose IFRS in absence of a market incentive to prepare IFRS financial statements, considering that the significant costs associated with adopting IFRS outweigh the benefits.

The main criteria on which IFRS are based are related to the guiding principles for the preparation of financial statements mentioned in the previous paragraphs, i.e. the primacy of substance over form, the balance sheet approach (with the balance sheet taking precedence over the income statement), the principle of prudence, the fair value measurement of assets and liabilities, the priority given to the investor's view, the important space given to interpretation, and the absence of sector-specific texts. As already stated, the aim of these criteria is to ensure transparency, completeness and neutrality of the reported data, as well as the comparability of such data with the financial statements of companies operating in different countries.

The following table lists the IFRS standards, with their name, description and year of issue; if a principle has been re-issued several times, the date of the most recent issue is indicated with an asterisk (*).

Tab. 1.1. List of IFRS principles, names, descriptions and issuance dates

Standard	Name	Description	Issued
IFRS 1	*First-time Adoption of International Financial Reporting Standards*	It establishes procedures that an entity must follow when first adopting IFRS standards to prepare its financial statements.	2008*
IFRS 2	*Share-based Payment*	It requires an entity to recognize in its financial statements share-based payment transactions (such as shares granted, share options or stock appreciation rights), including transactions with employees or other parties to be settled in cash, other assets or equity instruments of the entity.	2004
IFRS 3	*Business Combinations*	It outlines the accounting where an acquirer obtains control of a business (e.g., as a result of an acquisition or merger). Such business combinations are accounted for using the 'acquisition method', which generally requires that the assets acquired and liabilities assumed are measured at their fair value at the acquisition date.	2008*
IFRS 4	*Insurance Contracts*	It applies to all insurance contracts issued by an entity and reinsurance contracts held by that entity. It has been superseded by IFRS 17 as from January 1, 2023.	2004
IFRS 5	*Non-current Assets Held for Sale and Discontinued Operations*	It outlines how non-current assets held for sale are accounted for. In general, these assets are not amortized, are measured at the	2004

		lower of carrying amount and fair value less costs to sell, and are presented separately in the statement of financial position.	
IFRS 6	*Exploration for and Evaluation of Mineral Resources*	It allows entities adopting the standard for the first time to use the accounting policies for exploration and evaluation assets that were applied prior to the adoption of IFRSs.	2004
IFRS 7	*Financial Instruments: Disclosures*	It requires information on the importance of the financial instruments to the entity and the risks arising from those instruments, both in qualitative and quantitative terms.	2005
IFRS 8	*Operating Segments*	It requires particular categories of entities (mainly those with listed securities) to provide information on their operating segments, products and services, geographic areas in which they operate, and major customers.	2006
IFRS 9	*Financial Instruments*	It includes requirements for recognition and measurement, impairment, derecognition and general hedge accounting. It was issued on July 24, 2014, replacing IAS 39.	2014*
IFRS 10	*Consolidated Financial Statements*	It describes the requirements for the preparation and presentation of consolidated financial statements	2011
IFRS 11	*Joint Arrangements*	It outlines the accounting by entities that jointly control an arrangement.	2011
IFRS 12	*Disclosure of Interests in Other Entities*	It requires a wide range of disclosures about an entity's interests in subsidiaries, joint arrangements and associates.	2011
IFRS 13	*Fair Value Measurement*	It provides a single framework for measuring fair value and	2011

		requires disclosures about fair value measurement.	
IFRS 14	*Regulatory Deferral Accounts*	It allows an entity adopting IFRS for the first time to continue to account for regulatory deferral account balances in accordance with GAAP.	2014
IFRS 15	*Revenue from Contracts with Customers*	It specifies how and when an entity should recognize revenue and requires such entities to provide more precise and relevant information.	2014
IFRS 16	*Leases*	It specifies how an entity should recognize, measure, present and disclose leases.	2016
IFRS 17	*Insurance Contracts*	It establishes principles for the recognition, measurement, presentation and disclosure of insurance contracts within the scope of the standard. The objective is to ensure that an entity provides relevant information that fairly represents those contracts and enables users to understand the effect that insurance contracts have on the entity's financial position, financial performance and cash flows.	2017

Considering the effect on insurance contracts and the close relevance to the topics covered during the internship at EY, IFRS 9 and IFRS 17 will be discussed in more detail in the following paragraphs.

1.4.1. IFRS 9

As mentioned earlier, IFRS 9 establishes requirements for the recognition and measurement of financial assets and financial liabilities, replacing IAS 39. The project of replacement was in stages; the IASB first issued IFRS 9 in 2009 with a new classification and measurement model for financial assets followed by requirements for financial liabilities and derecognition added in 2010. Subsequently, IFRS 9 was amended in 2013 to add new general hedging requirements. The final version of IFRS 9 issued in July 2014 supersedes all previous versions, although they remain available for early adoption for a limited period.

In this regard, IFRS 9 provides significantly better information for several reasons. First, it introduces a structured approach to the classification of financial assets that reflects the business model in which they are operated and their cash flow characteristics; it also uses a forecasting model for credit losses, leading to more timely recognition of losses. Finally, the hedge accounting model has been improved, better linking risk management with its accounting treatment.

According to AICPA, IFRS 9 provides a new classification system that determines how insurers measure their financial assets; this system is based on a double test of cash flow characteristics at the instrument level and a business model assessment at a higher level that considers how an insurer manages its financial assets. The introduction of the category FVTOCI[14], in addition to amortized cost and FVTPL (Fair Value Through Profit and Loss), is particularly relevant; this category is a mandatory classification for debt instruments that SPPI[15] test and are held within a business model (i.e., how the entity manages its financial assets) whose objective is achieved either by holding these assets, collecting their contractual cash flows, or by selling them. The classification is then guided by the requirements of the business model; the characteristics underlying the business model criteria are analyzed and these criteria are applied to the investment portfolios.

IFRS 9 introduces also a new impairment model, in order to replace the Incurred Loss Model from IAS 39; insurers must apply this model to held debt instruments (measured at amortized cost or FVTOCI), trade receivables, lease receivables and contract assets. The model introduces numerous new features in terms of the perimeter, the bucketing of credits, the holding period for estimating the expected loss and, in general, certain characteristics of the

[14] Fair Value Through Other Comprehensive Income; assets measured at FVTOCI are those whose estimated value has an impact on comprehensive income.
[15] The assessment of the characteristics of contractual cash flows intends to identify whether they are "solely payments of principal and interest on the principal amount outstanding". For this reason, the assessment is colloquially referred to as the “SPPI test”.

elementary components of credit risk; in particular, it is required that at each reporting date it is checked whether there is an increase in credit risk.

A distinction is also made between 1-year expected loss (credit losses from default events to occur within 12 months) and expected lifetime loss (losses possible during the entire life of the financial instrument). These are calculated from the weighted average of the expected loss and the probability of default:

$$\boldsymbol{1 - Year\ Expected\ Loss} = PD \cdot LGD \cdot EAD \cdot TtM$$

$$\boldsymbol{Lifetime\ Expected\ Loss} = \left(\sum_{k=1}^{M-1} \frac{MarginalPD_k \cdot LGD \cdot EAD_k}{(1+EIR)^{k-1}}\right) + \left(\frac{MarginalPD_M \cdot LGD \cdot EAD_M \cdot TtM}{(1+EIR)^{M-1}}\right)$$

Where:

- *PD = Probability of Default* in a certain time horizon.
- *EAD = Exposure at Default*, so the amount at the time of default.
- *LGD = Loss Given Default*, so the credit lost at the time of default.
- *EIR = Effective Interest Rate.*
- *TtM= Time to Maturity.*

Finally, IFRS 9 permits the use of hedge accounting, which is a technique that changes the normal basis for recognizing the gains and losses (or income and expenses) of associated hedging instruments and hedged items so that both are recognized in profit or loss in the same accounting period; in this regard, insurers can benefit from the changes introduced by IFRS 9, such as hedging with options.

The effects of these changes on the financial position of insurance companies are still being researched by experts and analysts; in this regard, a report published by EY in March 2023 is particularly interesting. The research considers a sample of 20 listed insurance groups that have globally adopted IFRS as their accounting framework, analyzing publicly available information on the impacts of applying these standards. The picture that emerges is that IFRS 9 and IFRS 17 (which will be described in the next paragraph) have a limited impact on financial strength and overall strategy, but the impact on equity and future earnings could be significant;

according to EY data, most insurance groups expect minimal disruption from IFRS 9, with negligible changes to strategy, cash flow, capital management and dividend payout capacity.

1.4.2. IFRS 17

IFRS 17 was published on May 18, 2017, taking effect from January 1, 2023; this standard applies to all insurance contracts and introduces a present value approach to contracts.

Here is central the calculation of the Contractual Service Margin (CSM), which represents the profit not yet earned from a group of insurance contracts and is defined during initial recognition, i.e., when the group of contracts is recognized, and insurance cover begins. A relevant factor here is precisely the need not to evaluate contracts on an individual basis, but rather to aggregate them according to three different drivers, namely the type of risk transferred from the insured to the insurer, the length of the contract, and the degree of profitability.

The literature also points out that IFRS 17 has outlined three approaches for measuring the liabilities of a group of contracts:

- *Building Block Approach (BBA);* this method requires the calculation of LRCs (Liability for Remaining Coverage), which are the liabilities for future services, as the present value of all cash flows, net of an adjustment for non-insurance risks, plus expected future profits, and LICs (Liability for Incurred Claims), the liabilities for those claims that have occurred but not yet settled.

- *Premium Allocation Approach (PAA);* this method is often referred to as being more straightforward, as it only requires the calculation of premiums not yet set aside as profits, net of expenses incurred (administrative, management and provision-paying expenses, etc.), which can only be applied if the final result in terms of recognized profits is the same as the BBA.

- *Variable Fee Approach (VFA)*; this is only applicable to those contracts where the insurer shares the performance results of an underlying asset with the insured, typically life-death contracts.

Another concept that should be emphasized is reinsurance; it is a commonly used tool for insurance companies to transfer their actuarial risks to other entities. In fact, in order to reduce their exposure to certain eventualities (in particular damage caused by natural disasters or large-scale catastrophes), companies enter into an insurance contract with another insurance entity in which, against payment of a premium, the reinsurer will help the company to pay back claims.

It is generally very relevant that the implementation of IFRS 9 is closely linked to the implementation of IFRS 17. The technical literature, in fact, emphasizes how these standards optimally align the valuation methods of assets and technical provisions, thereby affecting the recognition of profits; under the previous accounting standards, on the other hand, the valuation methods applicable to assets and technical provisions were different, and the accounting for investments was only loosely linked to the liabilities that finance them (Sebastian, 2022).

It is also underlined that, with the implementation of IFRS 9 and IFRS 17, earnings volatility is mitigated by not marking certain assets and technical provisions to market and/or allowing valuation changes to flow through "other comprehensive income" (OCI).

It follows, therefore, that the implementation of such standards, as well as the provision of consistent and up-to-date data, is crucial for insurance companies to create value for shareholders while reducing balance sheet and earnings volatility. As part of this approach, experts suggest that insurers should therefore classify blocks of assets and liabilities according to the valuation and profit recognition method, and construct the balance sheet by assessing expected returns on assets, funding cost and volatility.

The following figure shows the impact of IFRS 9 and IFRS 17 for the different business lines of the insurance sector.

	Impact				
Type of business	Calculation	Display on balance sheet	Flow through profit and loss (P&L)	P&L volatility	Relevance of disclosures
Guaranteed life and health	High	High	Varies	Varies	High
Long-tailed non-life	High	High	High	High	High
Fee-based life and savings	Medium	Medium	Low	Low	Medium
Short-tailed non-life and health	Low	Low	Low	Low	Low

Fig. 1.4. Impact of IFRS 9 and IFRS 17 by segment (source: Wellington Management)

Also, data and conclusions reported in a 2019 EY report are of particular interest. In that report, among other topics, proactive responses of insurance companies related to the implementation of IFRS 17 and IFRS 9 are outlined.

Specifically, prior to implementation, CFOs should communicate to key stakeholders, including market analysts and shareholders, in a timely manner, providing clarity on the expected impacts on financial statements and profit profiles; they would also need to review current management reporting, key performance indicators and incentive frameworks for applicability and incorporate the necessary changes for margin and volatility analysis. Finally, any tax, capital or distributable profit implications should be assessed.

In addition, the report recommends the preparation of pro-forma balance sheet, profit and loss (P&L) statements and disclosure notes to meet the new requirements; the design of specific controls to promote the quality of new processes, robustness and integration into existing control frameworks, improving efficiency and cost effectiveness is also of particular importance.

According to EY, a natural consequence of these changes is the creation of new internal reporting templates (e.g., forecasts and other management reports) and external reporting templates (e.g., investor and analyst packages) or, at least, the revision of existing ones.

The implementation process should also focus on the verifiability of the data reported in the financial statements, even at the cost of a high level of interaction and consultation with external auditors.

In this regard, the report puts in evidence that it is crucial to assess the current availability of data against the new data requirements for both model inputs and outputs, as well as to modify the content and structure of the data captured by the business units to support group reporting (including through the implementation of more efficient IT systems).

These actions are closely linked to the needs for better control of data quality, storage and archiving, better reconciliation of data according to the new requirements, and an improved governance process.

It is also reiterated that one of the main objectives of IFRS principles on insurance contracts is to increase the transparency of insurers' financial statements; achieving this objective involves providing information about the risk assumed by the insurer, the level of uncertainty in the contracts, the factors that determine performance, the amount the insurer expects to pay to fulfil its insurance contracts, and the value of embedded options and guarantees.

Some of the disclosure requirements of IFRS 17 are comparable to those already contained in IFRS 4; however, new and more extensive disclosures are required for recognized amounts and roll-forwards. In this regard, the EY report suggests providing for a high level of granularity and determining the appropriate level of disaggregation of this information.

Insurers should, accordingly, develop systems, source data and valuation models to meet the requirements for detailed and granular disclosure of changes in the liability and asset balances of insurance contracts over the period. Finally, insurers should be able to reconcile different reporting bases, explaining why asset and liability balances, earnings and equity/capital are different when measured under IFRS, or other reporting regimes.

In consideration of this information, it is therefore clear that operating in a high data quality and data governance environment is crucial to making the insurance business as profitable and robust as possible; these issues will be explored in more detail in the following chapters.

2. Data governance for insurers

2.1. What is behind the concept of “data governance”?

Considering what has been said in the preceding paragraphs, it seems necessary to define the topic of data governance in detail, starting with a general overview related to the importance and benefits, and then discussing the application of this concept in the insurance sector.

Firstly, data governance is closely related to the concept of data management, even though it remains a separate topic.

According to Keith D. Foote (2022), data management originated as a notion in the 1960s, with ADAPSO[16] (the Association of Data Processing Service Organizations) offering advice on data management, emphasizing professional training and quality assurance metrics. However, the problem of optimal data management first emerged in the 1950s; several companies working with early computer prototypes used entire floors to store and only punched cards that stored their data. In this regard, the programs of that time were set up according to an Absolute Machine Language (or First-Generation Programming Languages), with binary or decimal codes read from on/off switches activated on the front of the computer, from magnetic tapes or punched cards. Data management has evolved steadily over the last few decades, moving from high-level programming languages to computationally efficient tools for managing and querying databases (with trends highlighting the increasing contribution of artificial intelligence to data management).

In this respect, data governance is typically part of a data management platform, designed to ensure the quality and usability of the data collected by an organization; although the early versions of the programs focused on cataloguing data, in 2005, data governance began to gain popularity as a method of accessing quality data for big data research purposes.

[16] Association of Data Processing Service Organizations was the original name of the Information Technology Association of America (ITAA); founded in 1960, its members included companies that offered computer software services to the public. The association’s main goal was to improve management methods, develop service capabilities and set performance standards for the software and service industry.

So, according to DAMA (Data Management Association)[17], data governance is the definition of the rules and control over data management, in terms of planning, execution and monitoring; consequently, data governance can be seen as the ability to manage data as a real company asset, assuming the presence of a strategy and ensuring the security of corporate data. In this regard, experts state that effective data governance should preserve the accessibility and availability of data, while also ensuring that it is not misused, altered, or stolen, reducing the risk of data compromise and protecting the company from the heavy penalties for violating data privacy regulations (Mahanti, 2021).

The main purpose of data governance is, therefore, to ensure data integrity and regulatory compliance. The fist concept is crucial to ensure smooth working productivity, which is consequently reflected in profits; ensuring effective data governance minimizes the risk of working with inconsistent data from databases and platforms, negatively impacting profitability, operational efficiency and reputation with customers. The literature (Zanelli, 2022) states that “the paradox of having huge amounts of data at one's disposal without being able to derive value from it is quite widespread within organizations”; in this regard, a Forrester Research survey of 2021 found that only 35% of top managers have full confidence in the data at their disposal.

The second overarching theme is compliance; regulations require (with industry-related differences) that companies have an infrastructure in place to ensure integrity in order to ensure customer data is properly monitored and managed. Two (but not the only) of the major regulations that protect the integrity of consumer data are for example the European Union’s GDPR (General Data Protection Regulation, that came into force in 2016) and the CCPA (California Consumer Privacy Act); breaking such regulations can have significant costs to a corporation in terms of penalties, image, and business continuity.

2.2. Benefits and challenges

Nowadays, data governance is the prerogative of sectors, such as banks and insurance companies, whose business is by definition linked to data management and which, as a result,

[17] DAMA is a non-profit, independent association founded in 1980 in Los Angeles, California, with the primary purpose of promoting the understanding and development of data and information management practices as key business assets. Today, the association is dedicated to the development and execution of procedures, practices, policies and architecture that properly manage the entire lifecycle of a company's data.

are forced to activate strict control mechanisms. According to Zanelli (2022), data governance necessarily presupposes a strategy; the objective is to ensure as much clarity as possible on the meaning and quality of each piece of data and on the responsibilities borne by business rather than IT. These objectives also presuppose intervening on the data culture of the entire organization, on people's behavior and on business processes, in order to have a common language and to ensure that staff dealing with data have the right skills.

Consequently, it can be argued that for corporations, the benefits of introducing data governance are different. In general, Zanelli lists some undoubtedly important points, such as increased revenues, higher trust (as the data used turns out to be more reliable and the business more responsive), decreased and monitored risk related to security and privacy, enhanced operational efficiency and dissemination of knowledge (reducing the risk of the same data being arbitrarily interpreted), business support (i.e., support for strategic programs that are ongoing in the organization, through Master Data Management tools).

Added to these benefits are also greater dissemination of data culture within a company, greater clarity on ownership of responsibility, and better integration between information systems and multiple data sources.

As mentioned earlier, the issue of data governance is closely linked to compliance with laws and regulations; in this regard, organizations need evidence to prove their compliance to regulators, and accurate, complete, consistent and traceable data can serve as proof. However, although enterprises acquire and store huge amounts of data, many have only a vague idea of what data they hold. According to R. Mahanti (2021), a very small percentage of the data an organization holds is used for business purposes or is subject to regulation. This thesis is supported by further scientific literature (DalleMule & Davenport, 2017), whose cross-sectoral studies show that on average less than half of a corporation's structured data is actively used for decision-making and less than 1% of unstructured data is analyzed or used at all. In addition, a survey conducted by the Compliance, Governance and Oversight Council (CGOC) in 2018 showed that only 1% of retained information is subject to legal retention requirements (i.e., it must be retained because it is related to the subject of actual or reasonably anticipated litigation or regulatory proceedings); identifying such data is consequently very complex.

Furthermore, in order to be compliant, a company must not only possess the ability to identify data, but also be able to produce the right data at the right level of granularity and at the right time; in addition, the data must be of high quality (i.e., it must be accurate, consistent and complete); organizations must also be able to trace data from authoritative sources, if necessary.

There are potentially thousands of laws affecting companies, according to Baker & Sjoberg (2018), with over 100,000 legal requirements relevant to multinational companies. Moreover, laws and regulations tend to differ across industries, markets and geographies, making compliance more difficult, especially for corporations with multiple lines of business and for multinational enterprises that have operations in multiple countries. In addition, even companies located in one country, but doing business with entities in other countries, may have to comply with more laws and regulations from other countries; in this regard, the European General Data Protection Regulation (GDPR) requires privacy protection for all personal data collected for or about EU citizens, even though the organization may not operate in the EU.

Reporting on a June 2017 Spiceworks survey, 37% of respondents cited a lack of clarity on the steps required for GDPR compliance as their main concern. In addition, according to Mahanti (2021), data security and privacy workflows are inadequate to meet the compliance requirements of recent legislation such as the GDPR; revising these workflows would, therefore, require gap analysis and subsequent policy revision.

Mahanti also argues that data requirements related to legislation should not be approached in a piecemeal manner, and that a holistic approach to data governance that spans all regulations is needed. In this regard, corporations should understand the data requirements, the impacts of each legislation and regulation, and then identify common or overlapping elements, as well as conflicts and differences between them.

As a result, organizations are also subject to an increasing number of complex regulations that continue to evolve and become more stringent in the face of data breaches, fraud and security incidents. In general, the increase in regulations has had (and will continue to have, according to trends and expert surveys) a massive impact on data management.

Although regulations and data are created independently of each other, they tend to impact each other and have added new dimensions of complexity and challenges. In the following figure, the main complexities that data and regulations bring, creating challenges for companies, are depicted.

Fig. 2.1. Challenges related to data and regulations (source: Mahanti, 2021)

Regulations and data have an impact on each other, and in order to comply with data-related regulatory requirements, organizations must have an effective data governance program in force; this would require the implementation of a technical solution to achieve compliance and the establishment of a data governance program to provide the framework to support compliance-related activities.

According to L. Irwin (2022), key aspects of data governance include availability, usability and consistency of information, as well as data integrity and security.

Companies can therefore create effective data governance practices by ensuring that information is secure, accurate, documented, managed, and verified.

Fig. 2.2. shows the main ways in which a sound data governance strategy can help corporations avert (or at least mitigate) the main data compliance issues.

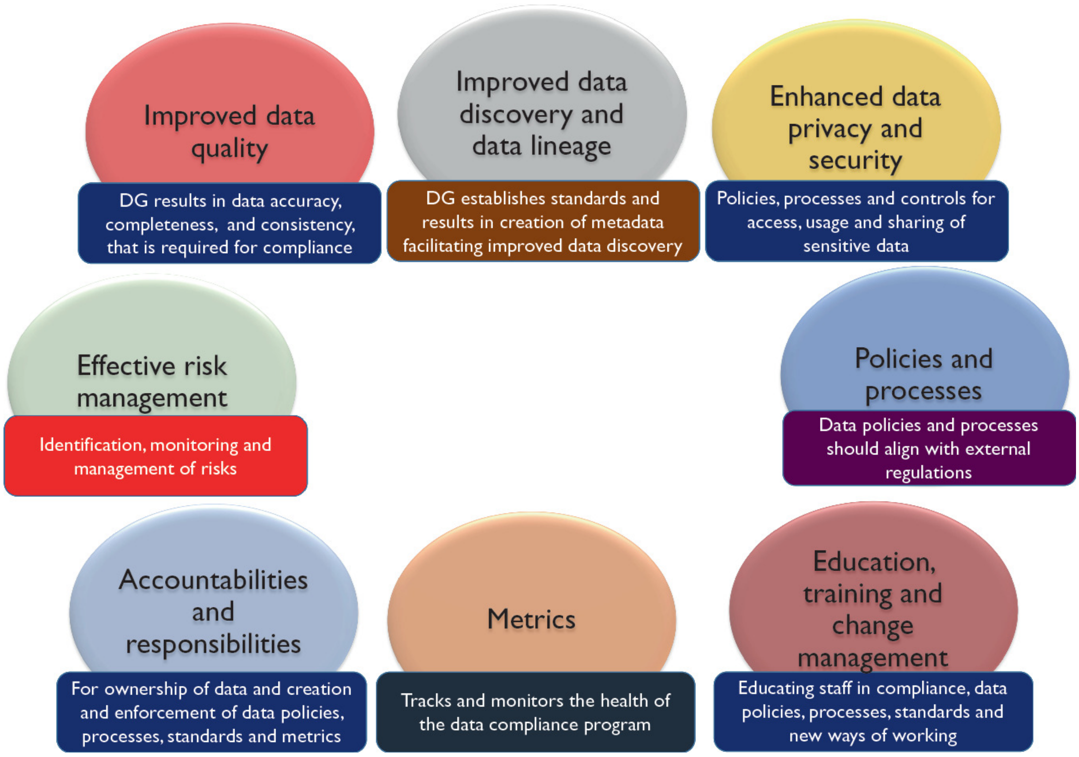

Fig. 2.2. Role of data governance in compliance (source: Mahanti, 2021)

The main points are described in the figure above:

- *Improved data discovery and data lineage*; data governance is essential in order to establish data stewardship roles. Data stewards from different business units, departments and business functions work together to resolve data differences; they are responsible for the datasets, understand their business meaning and the purposes for which they are used.
  In addition, data governance leads to the definition of standards for data elements and entities; data governance thus ensures the updating of metadata and data path information, which in turn supports the data discovery process (Mahanti, 2021). This is even more relevant considering that data linearity or traceability, i.e. the ability to trace the sources of data in a report, is becoming a regulatory requirement, especially in cases of regulatory reporting.

- *Enhanced data privacy and security*; academic literature highlights that data governance ensures that appropriate policies, guidelines, processes and controls are in place for the access, use, storage and sharing of critical data, protecting the data and ensuring access to authorized parties. It is evident that these aspects are particularly important for anti-money laundering (AML) controls and for complying with data protection and privacy laws such as the GDPR.

- *Improved data quality*; the topic of data quality will be dealt with in more detail in the next chapter. However, it can be said that data quality is closely linked to a strategy based on governance; the latter allows for better quality data, i.e. complete, up-to-date, accurate, consistent and free of duplication, with adequate controls in place to monitor it. Data quality is very important when it comes to meeting the controls generally imposed by regulations, especially those related to regulatory reporting requirements (Zanelli, 2022).

- *Effective risk management*; according to Mahanti, data governance enables the identification, monitoring and management of risks throughout an organisation's service lines, facilitating rapid corrective action and reducing the likelihood and severity of damage.

- *Accountabilities and responsibilities*; Data governance is very important, as it allows for the definition of roles, responsibilities, ownership, management, accountability and decision-making rights in case of conflict resolution; it also ensures that appropriate structures, roles and committees are in place to establish, enforce and maintain data-related policies, processes, standards, rules and metrics to ensure regulatory compliance. The regulations require managers to be aware of and responsible for the data that drives their activities, placing a huge emphasis on the Data Protection Officer (DPO)[18] and Chief Privacy Officer (CPO)[19].

[18] The DPO is the figure introduced by the GDPR whose function is to support data controllers, employees and data processors in preserving data and managing risks in accordance with the principles and indications of the European Regulation. The DPO is therefore a technical and legal advisor, with executive power; in fact, in addition to advising and supervising, this figure also acts as an intermediary between the organization and the authority.
The DPO's tasks are set out in detail in the Article 39 of GDPR and include informing, supervising and cooperating.

[19] The CPO, unlike the DPO, has a distinctly managerial role and cooperates with all company departments in order to adapt the organization to the GDPR and to ensure compliance over time. This figure is linked to strategic consulting, with the task of observing, evaluating and organizing the management of data processing within a corporation, so that it is used and processed in a manner that complies with the law. The CPO may be an employee or a consultant external to the organization; however, it is essential that he or she interfaces constantly with the data controller, in order to define a strategic plan aimed at leading the organization towards the pursuit and consequent maintenance of compliance, detecting any risk factors and needs, and thus guaranteeing a process of compliance that is continuous over time and always characterized by the same effectiveness.

It is therefore important for all stakeholders to clearly understand and agree on who is responsible for what within the corporations regarding compliance-related tasks, and to ensure that they have the right authority to perform these tasks and are sufficiently empowered; moreover, compliance often requires the involvement of multiple stakeholders. Mahanti states there is a need to establish clear responsibilities for data maintenance and approvals, establish proper communication channels, and foster a collaborative environment to prevent different teams from duplicating the same activities or conflicting when it comes to achieving their respective compliance goals.

- *Policies and processes*; a corporation should have internal data policies, which should align with external regulatory requirements, be up-to-date and reflect the most current regulatory requirements. For this reason, there should be controls and audits to periodically review external regulatory documentation, identify and understand any changes, and revise internal data policies accordingly to remain aligned with external regulatory requirements. The scientific literature (Mahanti, 2021) distinguishes multiple policies: data classification policy, data acquisition policy, data security and privacy policy, data use policy, data access policy, data retention policy, data governance structure policy and data quality policy. Data processes, standards and rules will have to be changed or new processes, standards and rules will have to be created to adhere to the data policies so that the data processes comply with the regulations.

- *Metrics*; a data governance policy results in the creation and monitoring of data governance metrics. Experts agree that tracking metrics such as data quality metrics for compliance-related data (such as percentage data accuracy, percentage data completeness, and percentage data duplication), percentage of data elements that cannot be traced back to the source, number of failed data audits, and turnaround time for audit queries can provide insights into the health of a compliance program; eventually, this allows to implement corrective actions as well.

- *Education, training and change management*; an effective governance strategy ensures that stakeholders are educated and trained on the data governance

> approach used; in this regard, it ensures change management, to guarantee that stakeholders abandon their old way of working and feel comfortable operating with the new approach. Mahanti points out that, in the absence of sessions, training and effective change management, attitudes such as reticence to change and prioritization of other tasks are common. However, data compliance requirements cross departmental boundaries and, in the case of multinational corporations, even country boundaries; stakeholders should therefore be educated and trained in the use of data governance policies, processes, standards and controls so that they understand their responsibilities towards data and learn how to manage it differently in order to be compliant. Change management ensures that staff embrace the new ways of working and do not revert to the old ways.
>
> It is also clear that not all stakeholders have the same compliance roles and responsibilities; therefore, training, communication and awareness sessions should be tailored to different stakeholder groups according to the level of involvement required and their compliance roles and responsibilities (Zanelli, 2022).

However, it must be emphasised that, as with any new initiative, implementing a data governance strategy entails several challenges.

Citing scientific literature (Turner, 2022), examples of such challenges can be limited resources; automation is a complex path and requires the right personnel to develop and implement it, so many companies need external help and advice to get started. Such support may require costs in terms of time and monetary resources, which are not always affordable for small- to medium-sized companies. In this regard, small enterprises often struggle to find in-house staff with the right knowledge and skills to implement a good data governance plan; these problems are often compounded by technological challenges, communication barriers and a constant turnover of employees, resulting in a scattering of data on company devices and a loss of control over that data.

It should be emphasised, however, that data governance should only be applied to those areas that enable an organization to grow its business and those where security and compliance are of major importance. Data governance is therefore not about new activities or new processes as much as in most cases it is about tidying up activities that are already being done in the corporation in a haphazard, fragmented, inefficient way and without clear ownership. It is not a matter of adding new processes, but of modifying existing processes to ensure that good data management practices are respected (Zanelli, 2022); such practices are generally described in data governance policies. In this regard, companies should carefully choose the areas where governing data creates value, so that the initial effort to introduce good practices will be limited to one-off expenditures to begin generating returns.

Another crucial factor is the lack of leadership in many business contexts, as well as a poor understanding of data policy and business requirements (Turner, 2022).

Poor data quality also makes it difficult to ensure the integrity and determine the ownership of such data; in this regard, it may be necessary to organize and improve data before creating a data governance plan (Mahanti, 2021).

Considering these issues, the literature (Turner, 2022) recommends to adopt approaches such that data governance does not appear merely as a cost for companies, but rather as a necessary investment for a competitive and reliable business. Today, more *agile approaches* are preferred, as they allow this path to be approached gradually, putting the company's main needs at the center and structuring the path by successive sprints.

First, it would be advisable to identify the organization’s main needs and frame how data governance can support them, directly or indirectly, in order to measure the impact of the project for the company (Zanelli, 2022). Such practices should be supported by a multidisciplinary working team, involving the various figures in the company from IT and HR departments. In this regard, according to Irwin (2022), the team should include the relevant stakeholders within the company, taking the roles of *data manager* (whose responsibility is leading the implementation of the data governance strategy), *data governance architect* (whose role is designing a proper data governance framework) and the *compliance specialist* (who ensures that the framework accounts for relevant regulatory standards).

It is also advisable to define a business glossary, i.e. a vocabulary with definitions of some of the most relevant data for the business, identifying some key processes that use this data; this practice makes it easier to move at an organizational level, defining roles and responsibilities regarding the definition and management of data. (Mahanti, 2021). IT, in this regard, plays a key role, both in implementing the necessary supporting tools and in its own knowledge at the functional level.

Finally, constant measurement of the results achieved is necessary throughout the progress program, defining the KPIs properly.

2.3. Why is data governance essential in insurance sector?

Taking into account what has been described in the previous paragraphs, it is clear that data governance is crucial in the insurance industry, as data and analytics are at the heart of this business. According to a survey conducted by Irion in 2019, a consolidation process is undergoing, under which the insurance sector is rapidly closing a gap with respect to the average maturity level of banks on data governance issues.

Today, the rise of digital insurance companies and the evolution of the risk landscape drive the digital transformation of the industry; and with the expansion of the volume and variety of data, insurance companies need a stable framework to govern data and regulate access to it.

As mentioned earlier, maintaining data security and compliance is required by regulations such as GDPR. At the same time, there is a growing opportunity to study and understand customer data in order to offer superior products and services.

Given that insurance companies work with large amounts of data on a daily basis (to determine, for instance, whether an insured person is entitled to discounts, offers and cover), data governance ensures that the data collected, stored, analyzed and used by companies is accurate and complete. The goal is to enable insurers to make better decisions about customer needs, products and prices; with an effective clear data governance strategy, insurance companies are able to work reliably on the data underlying their models, in order to improve decision-making and reduce risk.

The challenges that have been described in the previous section can also be extended to the insurance industry; in fact, many insurance companies have difficulty implementing a data governance strategy because they lack the necessary technology, people, and processes.

In addition to the difficulty in implementing the data collected with the new technologies available, insurers must operate within a regulatory framework that is constantly changing (Turner, 2022); the implementation of a data governance strategy should also consider the budget constraints imposed, as well as making use of collaboration among product lines and internal departments.

As with other companies, a weak or absent data governance strategy for insurance corporations can manifest itself in various ways, causing problems in understanding data quality and data tenancy. According to Turner, a company could suffer from a lack of data governance in a number of instances; one example is an environment where data are fragmented or where there is a lack of shared nomenclature for certain corporate glossary terms.

Another problem could arise from 'obsolescence of data, which may be noncompliant or unviewable.

Such critical issues result in the fact that an insurance company's different business lines might collect the same data from the same customers but not share it; this disconnect leads to outdated and inaccurate data (Baroni, 2022).

Experts (Turner, 2022) also assert that A weak governance strategy of can lead to the risk of inaccurate analytics (compromised by outdated or inaccurate data), data breaches (as the distribution of data across various corporate devices makes it easier for third parties to infiltrate the system and for sensitive information to be stolen). The last aspect can then lead to fines and penalties for noncompliance with privacy and security-related rules and regulations, such as the aforementioned GDPR or the Health Insurance Portability and Accountability Act (HIPAA).[20]

Furthermore, as previously stated, insurance corporations with an effective data governance strategy are able to properly understand their customers' needs; this factor has a huge impact not only on increasing revenues, but on customer satisfaction and loyalty as well. In addition, by cataloging sensitive data protected by regulations, corporations can reduce risk of not being compliant. Finally, with data governance, insurance leaders can increase efficiencies across the business, saving time and money.

According to Turner (2022), from an operating standpoint, insurers' data management guarantees data completeness and accuracy; in this regard, a reliable insurance company, should be able to collect accurate and highly confidential data about clients, i.e., financial wealth or demographic information; accurate and updated data help insurers to make reasonable and decisions around pricing and new products. Moreover, through the implementation of data cleaning processes, usually present in data governance strategies, it is possible to automatically identify corrupt or incorrect data within the tables of a database, and then proceed to correct or

[20] U.S. Health Insurance Portability and Accountability Act (HIPAA) compliance requires companies that handle Protected Health Information (PHI) data to adopt stringent physical, network and procedural security measures within company structures. All companies that have access to patient health data, operating, providing support in health care treatments or payments, are required to comply with HIPAA regulations, which also binds any subcontracted or otherwise professionally related partner companies to compliance.

eliminate them (Baker & Sjoberg, 2018); this procedure protects against human errors in manual data entry and ensures the presence of records that are always correct and fully reliable.

On the other hand, in order to be compliant with privacy regulations, insurers should know what personally identifiable information (PII)[21] or electronic Protected Health Information (ePHI)[22] they collect; the reason of this is being able to recognize whenever data is private and warn internal data users. These factors are also necessary to ease the processes within the organizations and increase employee productivity and efficiency.

2.4. Data culture and data-driven companies

Considering what has been said in the preceding paragraphs, it is appropriate to introduce the role of so-called data-driven companies and the attempt of a growing number of corporations to adopt a business approach linked by data culture. A data-driven approach is the exploitation of Big Data in businesses and the effective use of data in decision-making; in this regard, one example is the analysis of customer data (Customer Analytics) is now common practice for many organizations (included insurance firms) because it enables digital marketing activities to drive business growth by building more meaningful and lasting relationships with clients.

Data-driven companies consider data management not as a technical factor, but as a strategic pillar of the business, favoring an analytical and data-driven approach to make informed decisions. According to Zanelli (2022), the transformation into a data-driven company cannot, in fact, take place with technology alone, but with a change management path capable of bringing the data culture to all company levels. In this regard, it is essential to have correct, fresh and frequently collected data, in order to have alternative sources of information to metrics and KPIs based on time series, statistics and reports.

To implement such an approach, it is first necessary to observe and understand business-relevant processes and behaviors, and find the best way to quantify and measure them (Turner, 2022).

[21] Data that can uniquely identify users, making it attractive and at risk of theft by cybercriminals (also due to their value when they are sold on dark web marketplaces). They are considered sensitive data, because they can be used, for example, for identity theft. Examples of PII could be your name, address, date of birth or bank details.
[22] Protected health information that is produced, saved, transferred or received in an electronic form. In the United States, ePHI management is covered under the HIPAA.

Then, companies are tasked with collecting data, governing it, protecting it and analysing it, which implies understanding the role Artificial Intelligence and Machine Learning, IoT and Advanced Analytics play in managing large volumes of data, the so-called Big Data. A data strategy is therefore needed.

Consequently, data should be an integral part of the competitive strategy, considering the macroeconomic context, benchmarking with the reference industry, and the company's business model. In this regard, actions can be implemented and measured to understand the company's competitive position and customer needs; the focus on data, numbers and quantitative measures should not replace the value of vision (Zanelli, 2022).

According to Mahanti (2021), data culture also passes through security; companies needing powerful and reliable infrastructures should be aware that intelligent data management cannot treat data protection as an extra but as an important pillar of the business itself.

As stated by scientific literature (Irwin, 2022), efficient data collection allows the creation of profiles consistent not only with socio-demographic characteristics, but also with the real habits and needs of customers, in order to enrich the customer database, the Customer Relationship Management (CRM), as much as possible; this information is then used for email marketing, mobile marketing, in-store promotions, proximity marketing and so on.

From a technological point of view, data from various sources, structured and unstructured, converge in the data lake, an important enabler in this kind of project, as a flexible and open technology.

Another crucial piece of technology is the Customer Data Platform (CDP), an evolution of the DMP, Data Management Platform, a single database, managed by marketing, accessible from all Marketing Automation systems. The main functionalities of this are real-time data collection (related to individuals and coming from different sources, offline and online), consolidation of individual profiles on an individual level, segmentation (i.e., the management of customer segments according to predefined rules, using advanced analytics or data science systems), activation of email campaigns, smartphone messages and data-driven advertising.

According to a 2022 Salesforce study, data sources (in marketing alone) increased from 8 to 10 on average in 2021, the most common of which are CRM, ERP, eCommerce, Contact Centre, Website and Mobile App; the research predicts that this will grow to 45 in 2025.

The technical literature (Zanelli, 2022) states that the main areas of analytics are finance and marketing, followed by customer listening (the so-called Voice of the Customer), i.e. the opinions left by customers during multiple interactions with brands, which is done automatically by systematically sifting through their feedback, both public, e.g. comments on social media, reviews and star ratings, and private, such as surveys or customer care satisfaction.

The data-driven approach is also becoming relevant for HR management. Managing and analyzing HR data helps to provide greater decision-making and strategic support in terms of people acquisition, management, development and retention.

As consequence, every company should be able to identify the data that is meaningful for its business in each application area; in this regard, in marketing, first-party data, those collected directly and stored in CRM, are undoubtedly the most valuable, and are then integrated with third-party data, especially from social media, to generate user profiles and deliver content in an optimized manner, thus improving campaign results.

The most advanced companies are now able to capture and analyze data in real time, at the very moment it is generated, to create projections and hypotheses that, thanks to the application of Machine Learning algorithms, are increasingly accurate and true (Mahanti, 2021).

The latest statistics show that more and more companies are basing their business on data. In this regard, according to research conducted by the Big Data & Business Analytics Observatory of the School of Management of Polytechnic of Milan in 2022, the Italian data management and analytics market reached EUR 2.41 billion, an increase of 20% over 2021. (the growth percentages are shown in the Fig. 2.3., at the end of the paragraph). This trend is mainly driven by the software component (54% of the market, +25% over the previous year), while spending on infrastructure resources is growing at a slower pace, below the market average; the sectors in which growth is most marked are retail, public administration and healthcare. In addition, the analytics budget allocated to public cloud services is growing at twice the rate of the market average and is close to a quarter of spending on Data Management & Analytics solutions and services.

However, it can also be said that there is still a lot of action to be taken by companies. The same research states, in fact, 49% of Italian companies claim to have introduced at least one data scientist, 76% a data analyst and 59% a data engineer. Furthermore, only 55% of SMEs claim to have made investments in Data Management & Analytics. Furthermore, four out of ten companies have no figure dedicated, even partially, to data analysis.

According to Carlo Vercellis, Scientific Director of the Big Data & Business Analytics Observatory in 2022, companies at the Italian and global level are becoming more mature in terms of culture and data-driven approach, although "the challenge for those who have started experiments or Advanced Analytics projects now is the industrialization of processes to ensure efficiency and data governance at all levels".

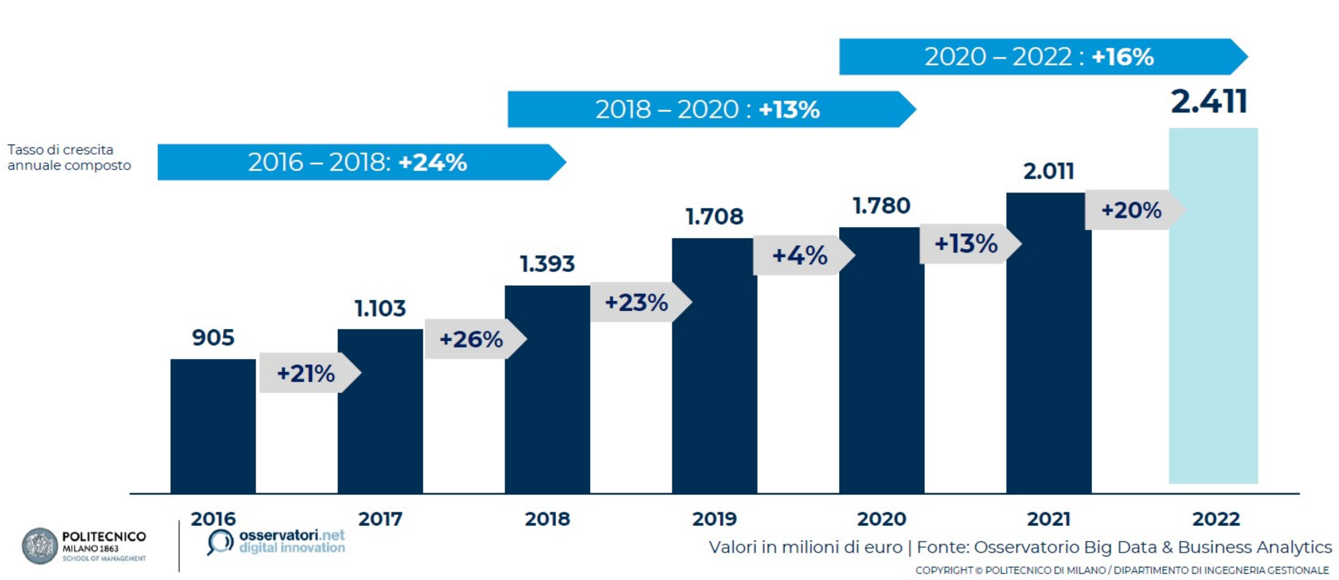

Fig. 2.3. expenses of Italian companies related to Data Management & Analytics from 2016 to 2022 (source: Big Data & Business Analytics Observatory)

2.5. What is Regulation 38?

The following section will explore the issue of insurance company compliance with the legislative patchwork, focusing on IVASS Regulation 38 related to data governance in the Italian context.

Regulation 38 was issued by IVASS in 2018, concerning the corporate governance of insurance companies and groups. In the press release following the issuance, IVASS highlights how insurance companies and groups must gradually adapt to the provisions in the regulation, streamlining the existing rules on data governance of insurance companies and ensuring their compliance with Solvency II Directive[23], EU Delegated Regulation 2015/35[24] and EIOPA guidelines[25].

In the regulatory review, IVASS emphasizes that a number of priority objectives have been pursued First and foremost, Regulation 38 ensures that the ultimate responsibility for the corporate governance system is clearly assigned to the administrative body, detailing its tasks and promoting an adequate composition, functioning and qualification of its members; moreover, the role of the company's key functions is strengthened, guaranteeing the direct interlocution of their holders with the administrative body.

Moreover, remuneration policies are aligned with the long-term interests of the company, also by providing for adequate disclosure to shareholders and the Supervisor; the regime for outsourcing functions or processes outside the company and/or the group is also regulated and simplified.

An important point of the regulation is the regulation of cyber risk and cybersecurity safeguards within the rules on corporate governance; the development of corporate mechanisms and processes for the management of possible crisis situations is consequently fostered, requiring in particular groups that are relevant for financial stability to prepare an enhanced contingency plan.

Finally, corporate awareness of environmental and social risks is also promoted.

It is therefore evident how this regulation gives rise to the drive for an adequate organization of corporate information assets, as well as a growing maturity in insurance companies inherent to the culture of data.

IVASS Regulation No. 38/2018 therefore aims to regulate data management as an integral part of the broader system of corporate governance, the system of internal controls and risk management. Indeed, As can be seen from the 2018 CIO Survey conducted by NetConsulting

---

[23] Directive is in force as of January 1, 2016. The principles on which it is based are the calculation of technical provisions, the solvency requirements of insurance companies and the management of each investment (pillar one), careful evaluation of all technical provisions and regulation of corporate governance (pillar two), transparency and information towards the markets and the supervisory authority (pillar three).

[24] Regulation on the treatment of participations acquired by insurance and reinsurance undertakings, as well as by insurance holding companies and mixed financial holding companies with ultimate Italian parent companies.

[25] The Guidelines issued by the European Insurance and Occupational Pensions Authority (EIOPA), are intended to support the convergence of the application of the provisions of the Solvency II Directives and Regulation 2015/35. also defines more general guidance on specific governance issues, aimed at increasing the convergence of supervisory practices.

cube, the main item of investment for CIOs in 2018 was Cybersecurity and GDPR; it is reported that more than 90% of respondents have in fact initiated at least one GDPR-related activity, almost double those in 2017. These percentages have, moreover, grown exponentially over the years.

As reported by the same survey, more than 50% of respondents in insurance corporations initiated Big Data projects in 2018, a percentage that was just over 20% in 2017. As a result, the high need for companies to make decisions in a timely and effective manner translates into a heightened focus on data quality and the possible costs and risks of poor data control. The World Insurance Report 2018 by Capgemini and Efma shows that more than 63% of the executives surveyed believe that the insurance of the future will have to integrate and manage the many data sources available today (smartphones, PCs, black boxes, wearables, etc.) in order to compete; the importance of integrating different data sources was also highlighted in the previous chapter.

As mentioned earlier, Regulation 38 was preceded by other regulations that encouraged the adoption of data quality approaches by insurance companies, in line with the European Solvency Directive.

According to the literature (Valentini, 2023), the new provisions of Regulation 38 do not therefore represent a true novelty, since they extend what was already provided for by the previous regulations.

With particular reference to data management, most of the provisions that were already in place previously on various topics are confirmed. In this regard, companies are required to possess accounting and management information that guarantees adequate decision-making processes and enables them to define and assess whether the strategic objectives set by the administrative body have been achieved. Furthermore, aspects related to the production of data and information for the purpose of insurance group supervision, already introduced in previous regulations, are defined.

In addition, aspects related to the security of IT systems are examined, establishing procedures to ensure the continuity of business processes and having the administrative body approve an ICT strategic plan.

However, the scope of analysis is broadened by strengthening the governance of the corporate information production process and the need for cybersecurity management is introduced.

Nevertheless, according to Zanelli (2022), Regulation 28 has an extremely broad scope of application, referring to “all corporate operations, in order to produce complete and up-to-date information on corporate activities and the evolution of risks, including procedures for reporting data and information to IVASS”.

Another relevant aspect is the formalization of the role of the person responsible for the completeness, adequacy (in terms of effectiveness and efficiency) and reliability of the data processing as the management body or, on its delegation, the CDO (or its equivalent) and/or the functions responsible for data quality.

Furthermore, the need for systematically documenting data acquisition, recording and reporting processes and procedures, documenting the data lifecycle and tracking critical data is reiterated (Valentini, 2023); these operations are obviously carried out by means of systems for classifying and protecting data from internal and external threats.

2.5. New frontiers of governance: Data Mesh and Data Product

It is appropriate at this point to mention innovative themes related to data governance and the ways in which corporations can interface with it.

In this regard, it is necessary to introduce the concept of the Data Mesh, a new and relatively recent governance approach; the most relevant innovative feature is that it is a new organizational and architectural model that recognizes the importance of a distributed and domain-driven approach (for what concerns data organization) together with a centralized one (as far as related governance is concerned). This makes it possible to think of data as products offered and managed by specific domains, thus meeting the real business requirements of a company, rather than individual application needs.

As stated in the scientific literature (Firpo, 2021), the data mesh is of no small importance for companies, as it mitigates the problem of integrating corporate data from peripheral sources into a single centralized platform; this problem especially concerns multinational companies that centralize data from different subsidiaries or countries, with their local or national specifications and regulations. Furthermore, it should also be considered that the source systems are not aware of the centralized data management process (data warehousing), do not know the needs of the data users and are not focused on ensuring data quality, because this is not part of their business objective. This usually lays the foundation for a total disengagement in the generation of data with a view to value creation for the entire organization, resulting in data consistency and quality problems.

Data Mesh mitigates the critical issues just described, in that it is an emerging solution architecture for organizing and delivering business data, moving beyond monolithic data platforms (such as a centralized data lake) to a distributed data platform with decentralized domain ownership, where data lakes and warehouses are simply treated as nodes on the mesh rather than the central point of the overall architecture. Mesh focuses on 'source sources' and 'use cases' of data, where data resources are designed and acquired to produce data products related to business needs.

In a report published in 2021, EY identifies some main points that underpin the Data Mash approach. First, there is a principle of domain-based ownership, which requires domain teams to take responsibility for their data. According to this principle, analytical data should be composed around domains, like team boundaries that align with the limited context of the system. As a result of the domain-based distributed architecture, ownership of analytical and operational data is moved to the domain teams, away from the central data team.

This is followed by the second point, the principle of federated governance; with this principle, interoperability of all data products is achieved through standardization, which is promoted across the entire data mesh by the governance guild. As consequence, the main objective of federated governance is to create a data ecosystem with adherence to the organizational rules and regulations of the industry.

As already briefly mentioned, the principle of data as product (Data Product) can also be defined, which projects a philosophy of product thinking on analytical data. This principle means that there are consumers for data outside the domain; the domain team is then responsible for meeting the needs of other domains by providing high-quality data. Going back to the previous point, federated computational governance is therefore a federation of Data Product owners (consequently absolutely internal to the organization) with the task of creating rules, standards, guaranteeing common and cross-cutting metrics, ensuring platform monitoring and automating (or at least simplifying) adherence to these standards.

Finally, a dedicated data platform team provides domain-independent capabilities, tools and systems to create, execute and maintain interoperable data products for all domains; with its platform, the data platform team enables domain teams to seamlessly consume and create data products.

According to Firpo (2021), the Data Mesh paradigm also represents a very strong guarantee against the risk of technological obsolescence. In the future, when new technologies emerge, each source system will be able to adopt them without problems. In fact, the continuity of operation of the entire system is ensured by the possibility of creating new connectors, specific to the data generated by these new technologies, which allow them to be made available to the rest of the company via Mesh services (from which the entire Data Mesh takes its name) through what is defined as a scaffolding system, i.e. a scaffolding that surrounds and connects the data coming from the various source systems.

In addition, a mesh creates awareness of data empowerment at any level of an insurance organization; according to EY, the level of abstraction provided by a data mesh alleviates pipeline development and installation workload, and thus reduces solution costs. Furthermore, the centralization of maintenance and security policies across products exposed as services facilitates compliance and monitoring.

The following figures better illustrate the concept of data domains and interoperability of domains and a graphical representation of a mesh's prototype.

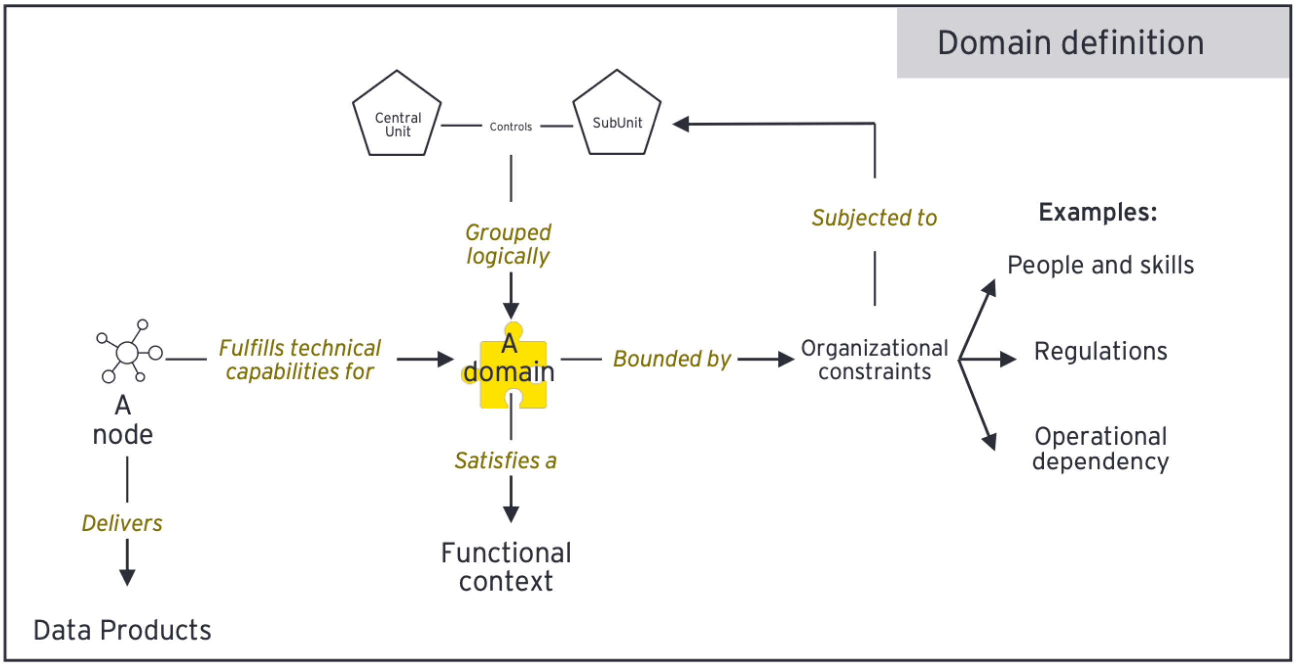

Fig. 2.4. Data domain's graphical representation (source: EY)

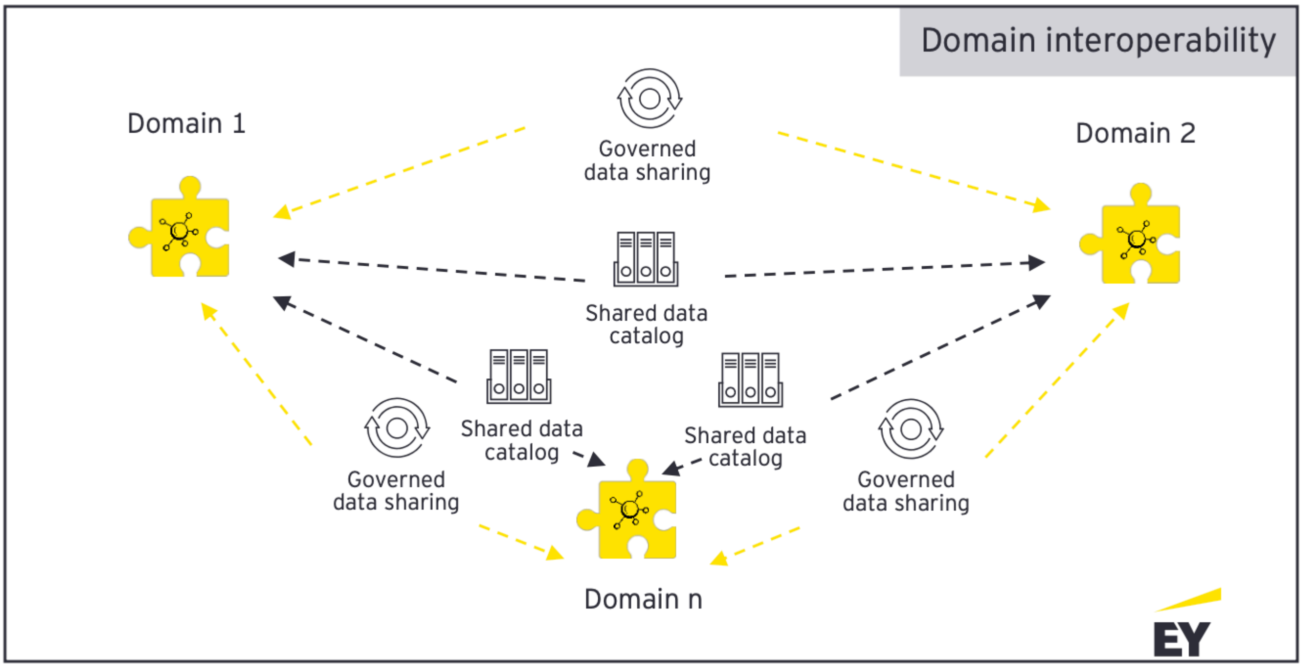

Fig. 2.5. Graphical representation of domains' interoperability (source: EY)

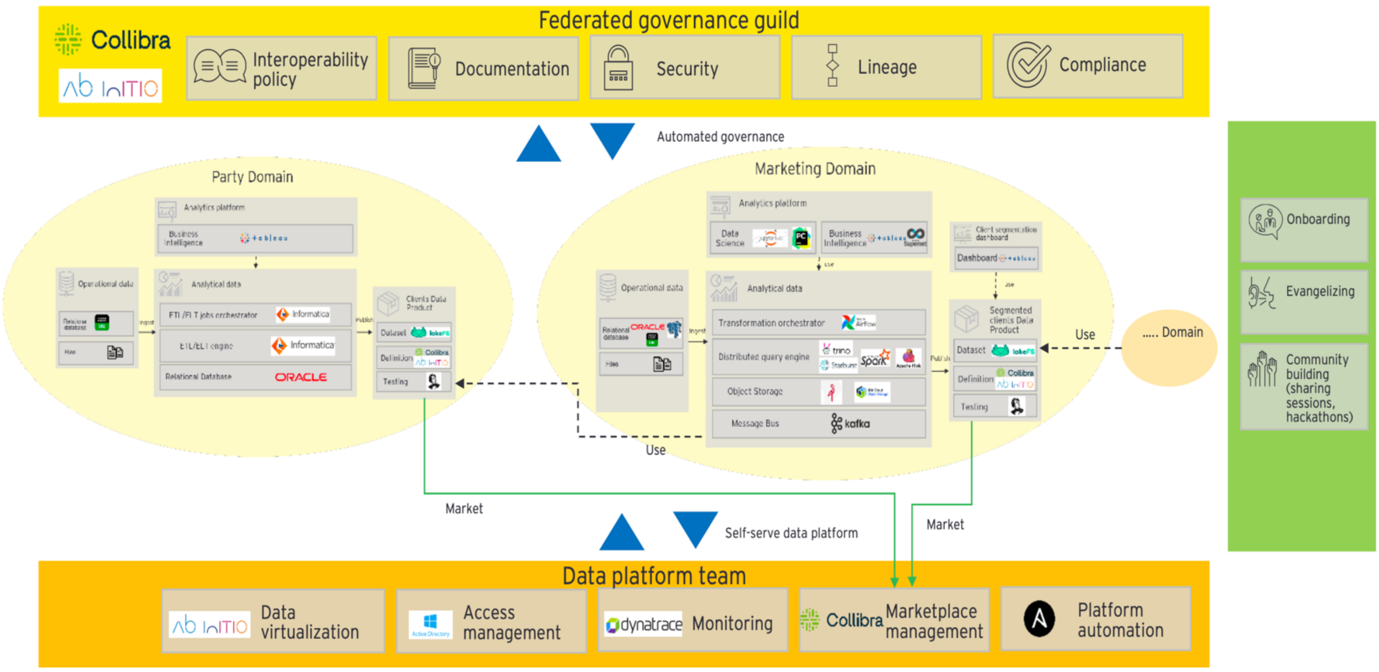

Fig. 2.6. Prototype of a mesh (source: EY)

3. Data quality

3.1. Relevance of data quality management

After delving into the topic of data governance, this chapter will explore the importance of data quality, as well as the main metrics used in assessing it; this topic is becoming increasingly discussed in organizations and has a strong impact on business competitiveness.

There is therefore a growing awareness of the weight that data quality has in achieving business objectives, regardless of size or business sector, and there is also a growing need to use data quality tools, which guarantee the reliability of the data ecosystem on which to base analyses.

The concept of data quality identifies a set of attributes to measure the level of data quality, which make it suitable to meet the data analytics objectives sought. In business intelligence (which will be explored in more detail in the following paragraphs), understanding the level of data quality is fundamental to understanding whether certain information can be used effectively within applications; in fact, only good quality data can feed accurate business data analysis and contribute to reliable decision-making strategies.

As mentioned earlier, the importance of data management has been emphasised since the 1980s; in this regard, it can be said that data fulfil all the characteristics of an 'intangible asset' as defined by accounting standards such as IFRS. Furthermore, data are easy to copy and transport, but not easy to reproduce if they are lost or destroyed; from what emerges from the academic literature, data are not consumed at the time of use and can be used for multiple purposes or by multiple people at the same time.

According to a 2020 CRO Forum report, analysts may spend up to 40% of their time validating data relevant to their analysis before the results can be used for strategic decisions; moreover, in several insurance companies, executives are skeptical of the data presented to them. In this respect, higher data quality would, in general, be necessary to formulate radically different customer service proposals and operating models.

Although several definitions exist, the CRO Forum report defines data quality "as a multidimensional construct that refers to the suitability of data for use", i.e., the ability to meet risk management requirements in its processes. This definition implies that data quality is a necessary prerequisite for effective risk management, but at the same time an operational risk itself.

Moreover, as already mentioned, data quality is crucial for the success of processes, for innovation and for the reliability of business reports (Ladva, 2022); lack of confidence in data leads to bad decisions and missed opportunities when data are inaccurate, incomplete, delayed or incomprehensible.

The CRO Forum report also provides another definition of data quality, according to which it refers to the “fitness for use” of data, i.e., the ability to meet the requirements of the intended use of the data in a specific situation. In this regard, there are multiple frameworks and approaches to data quality, which include defining objectives, planning, measuring, monitoring, organizing and managing tools.

Specifically, academic literature states that data quality should start with goal setting, planning and process/system design; however, one of the main obstacles to goal setting and, more generally, to a strategic approach to data quality, is uncertainty about the economic value of data (CRO Forum, 2020).

In this regard, Chief Risk Officers (CROs) can play a key role in data quality management, by promoting a vision of the benefits of data quality, supporting the definition of optimal data quality governance, and assessing the value and risk of data quality at the goal-setting stage.

Consequently, a data quality management activity therefore focuses on multiple dimensions (Turner, 2022). In particular, the completeness of the data is very relevant, i.e., it indicates the percentage of the collected data with respect to the potential derivable from the data sources available to the company. Other indicators, such as accuracy, consistency, uniqueness and validity of the data, will be described in the following paragraphs, but are also very relevant to better understand the importance of data quality for the business.

Moreover, among the various dimensions to be considered, it is often difficult for companies to obtain an accurate measure of the quality of the available data; however, adequate levels of data quality are crucial for organizations to become data-driven companies (Mahanti, 2021).

Overall, analysts agree that achieving quality data through data quality management is a common challenge for many companies, but awareness of the consequences of using inaccurate information is still lacking.

Very often, in fact, organizations tend to focus on collecting as much data as possible, sidelining assessments of the reliability or correctness of the datasets created (Turner, 2022); this can easily lead to reliance on inaccurate, incomplete or redundant data, with a domino effect of decisions based on inaccurate metrics. Operationally, this leads to increased time and cost required to process data, implementation of ineffective strategies and loss of new opportunities, poor decisions and loss of enterprise value, decreased levels of data governance and increased exposure to compliance risks.

It is clear, therefore, that the absence of a data quality solution to verify and validate analysis data can lead to very serious consequences for companies and their growth path. Thus, data quality management has a huge influence on corporate strategies and objectives, which is why it becomes necessary to find data quality solutions to improve data quality.

Analysis, and the subsequent process aimed at data quality, have consequently over time become preparatory activities for decision-making support; in fact, the information content that data can express is functional to its ability to describe the environment from which it was taken or observed (Cesarini et al., 2014). As already mentioned, the problem of low data quality plays a decisive role in many contexts: scientific, social, economic, political, etc. In this regard, the scientific literature (Fisher & Kingma, 2001) reports as an example the Space Shuttle Challenger tragedy in 1986, whose explosion was blamed on ten different categories of data quality problems[26]. Another well-known case reported is that of the Millennium Bug[27], which received a great deal of media coverage mainly because of the high economic impact it had, both in the public and private sectors (Cesarini et al., 2014); in this regard, although there are conflicting estimates on the costs that this data quality problem had on the various industrialized

[26] The Space Shuttle Challenger exploded moments after take-off, causing the death of the entire crew; the main cause of the explosion was most likely a sealing ring that broke due to the extremely low temperatures and high-pressure during launch. As later investigations revealed, the problem with the ring was known well in advance, and doubts about it had already been raised nine months before the tragedy. The component in question was documented in several database systems, each covering a different aspect of production and planning.
In some systems, the seal ring was correctly declared "critical", while in others it was classified as "redundant", meaning that its failure would be safeguarded by other components. Thus, although the data needed to analyze the effect of temperature on the seal was available, it was scattered across the various databases, so that in the end only part of it was used by both NASA and the manufacturer for regression analysis (Fisher & Kingma, 2001).

[27] The Millennium Bug, also known as the Y2K bug, was a computer flaw (bug) that manifested itself at the date change between December 31, 1999, and January 1, 2000, in some data processing systems. The bug arose from the fact that, in order to represent dates, various software packages developed since the early days of computing used only two decimal digits to store the year, which could take values between "00", corresponding to 1900, and "99", corresponding to 1999, so that when the year 2000 was reached, the consequences would be unpredictable. The infrastructures most vulnerable to this bug were banks and other financial networks (where all exchanges were handled by computers), civil and military infrastructures, (which depended on actual computers or elements containing chips) (Cesarini et al., 2014).

economies, it is agreed that the cost of bringing systems up to quality before the year 2000 was at least USD 400 billion.

The scientific literature (Mahanti, 2021) also highlights how effective data quality management practices are useful in increasing trust in data within a company; trust is, in fact, fundamental in an organization, but it is difficult to achieve, especially since the people who use data to make decisions are rarely involved in their collection and preparation for consumption. If the CEO has historically received inaccurate data that has led to poor business decisions, he may have doubts about future data, hesitate to rely on it, and try to validate it.

When a corporation has a solid data quality strategy and processes that everyone trusts, it gives the CEO and other decision makers the confidence to rely on data for decisions.

Moreover, as organizations grow, their data needs change and evolve; in this regard, good data quality is essential to ensure that an organization's data can scale for new use cases and business opportunities (Cesarini et al., 2014).

Poor data quality can therefore hinder a company's ability to scale effectively and efficiently. An example brought by Zanelli (2022), concerns an e-commerce company that uses data to personalize the customer experience for each visitor to its website, and as consequence it needs a robust and scalable data infrastructure to support this personalized experience at scale. Subsequently, if the company's data quality is poor, it will be difficult to scale customized experiences to large numbers of visitors without incurring serious errors or workforce inefficiencies.

As mentioned before, inaccurate data can also lead to operational inefficiencies and wasted time and resources, especially due to the time spent in working on quality testing data.

In addition, a company that doesn't maintain a complete dataset of its customers may make kinds of mistakes that could damage a company's reputation, but also waste time and resources that could have been used more effectively (Mahanti, 2021).

In this regard, good data quality can lead to increased customer satisfaction; when an organization has accurate and complete customer data, it can provide a better and more targeted customer experience, leading to enhanced revenues and customer loyalty.

Finally, as described in the previous chapter related to governance, companies that maintain high data quality standards are also more likely to comply with the laws and regulations that govern their industry, meeting reporting requirements and avoiding penalties for non-compliance (Cesarini et al., 2014).

3.1.1. What is the role of business intelligence?

As mentioned in the previous section, the term business intelligence refers to the processes and tools used to analyze business data, transform it into usable information and enable everyone in an organization to make more informed decisions. Also known as a decision support system (DSS), a BI system analyses historical and current data and presents the results in the form of reports, dashboards, charts, graphs and maps that can be easily assimilated and shared within the company.

In this regard, a decision support system is any interactive computerized system capable of collecting and analyzing information from huge volumes of data, including raw data, documents and knowledge bases. DSS systems thus support managers and planners in making informed decisions based on insights gained during the analysis process.

Furthermore, BI is sometimes referred to as “descriptive analysis” because it provides a description of a company's current and past performance, but without delving into the reason for this performance or speculating on future forecasts.

The term business intelligence is often used synonymously with the expression business analysis. According to the literature (Olavsrud & Fruhlinger, 2023), there is no consensus on whether the two terms are synonymous or not. However, a common distinction is that business intelligence focuses on what has happened in the past and what is happening in the present (*descriptive analysis*); business analysis focuses on the causes of a certain performance (*diagnostic analysis*), the likelihood of it happening again in the future (*predictive analysis*) and the actions, possibly corrective, that should be taken to achieve the best possible result (*prescriptive analysis*).

Nevertheless, both BI and business analysis are crucial for enterprises, as they work in synergy to provide companies with all four types of analysis (descriptive, diagnostic, predictive and prescriptive) necessary for decision-makers to gain insight.

Historically, the concept of business intelligence has been known since the late 1980s and has always been the responsibility of IT; queries were submitted to the IT team which provided answers to the company in the form of a static report. If there were subsequent questions, these were sent back to the IT team and generally placed at the back of the queue. This process, however, was too time-consuming and resource-intensive, so it was replaced over the years by a faster and more interactive BI concept.

Today, modern self-service BI tools allow business users to personally interrogate data, create dashboards, generate reports and share their results from any web browser or mobile device, all with minimal IT involvement; these tools will be better described later in this section. Recent artificial intelligence (AI) and machine learning technologies have further simplified and accelerated this process through the automation of numerous BI processes, including data exploration and the creation of reports and visualizations.

Increasingly, companies are choosing cloud-based BI tools that connect to more data sources, along with BI embedded directly into workflows and processes giving users the ability to make better, contextualized decisions on the spot.

Today's state-of-the-art BI platforms combine business intelligence, advanced and predictive analytics and planning tools in a single cloud-based analytics solution (Olavsrud & Fruhlinger, 2023). They can be integrated into any process, democratizing BI and analysis, which thus become accessible to everyone, not just IT teams or professional analysts.

The benefits of power BI tool are similar to the ones described in the previous paragraphs; in fact, a successful BI program should focus on increasing profits and performance, identifying problems, optimizing activities and numerous other objectives.

First, business intelligence can be essential for companies to gain and maintain a competitive advantage, as organizations can quickly identify new trends and opportunities and act accordingly. According to Mostarda (2023), corporations also able to assess their own capabilities, strengths and weaknesses relative to competitors and use this information to their advantage.

Furthermore, it can be a necessary support for fact-based decision-making; BI tools enable executives, managers and employees to identify information relevant to their roles and areas of responsibility and use it to make decisions based on facts rather than guesswork.

Another aspect that should be underlined is the support in measuring and tracking performances; from this standpoint, in fact, BI dashboards make monitoring key performance indicators (KPIs) and progress more immediate (Olavsrud & Fruhlinger, 2023). This enables a more direct comparison with defined targets and allows alerts to be set to know where and when to implement targeted improvement initiatives.

As consequence, BI's role is crucial to identify and define benchmarks, as it enables organizations to compare their process and performance metrics with industry standards, determine where improvements are needed, define meaningful benchmarks and monitor progress against targets (Mahanti, 2021).

BI solutions also allow to identify problems in advance, before they cause financial damage, such as (especially in manufacturing corporations) bottlenecks in production or distribution, upward trends in customer churn rates, increased labor costs, and more.

The consequence is enhancing operating efficiency; this factor is also related to the time saved searching for information, analyzing data and generating reports (Turner, 2022). BI also makes it possible to identify areas of overlap, duplication or inefficiency between departments or branches, supporting the streamlining of operations.

Making data and reporting accessible to all is another relevant advantage of BI solutions, as they offer intuitive and friendly-user interfaces, drag-and-drop reports and role-based dashboards that team members are able to use on their own, without the need for coding or other technical skills. An example of BI's interface can be appreciated in the figure at the end of the paragraph (Fig. 3.1).

Subsequently, this factor allows to improve customer and employee experiences. According to Olavsrud & Fruhlinger (2023), BI users can mine data in order to identify patterns in customer and employee behavior, analyze feedback and use insightful information to personalize and improve experiences.

So, as mentioned at the beginning, the final result is a revenue and profitability increase, along with a better understanding of where risks and opportunities exist, empowering teams to make rewarding adjustments. Specifically, insurance companies can consolidate financial data and monitor cash flow, margins, expenses, revenue flows and other factors in real time; in addition, they can monitor profitability and make decisions that improve net profits.

The most commonly used BI tools are dashboards (of which an example is shown in Fig. 3.1); they use diagrams, constantly updated graphs, tables and other data visualization types to keep track of predefined KPIs and other business metrics and provide a unified, near-real-time performance overview. In this regard, managers and employees can take advantage of interactive features to customize the information they want to view, deepen data analysis and share results with other stakeholders.

In this regard, the ability to view data and view them in context is an area where the BI stands out (Olavsrud & Fruhlinger, 2023). Graphs, diagrams, maps and other visual formats convert data into information in easily and quickly understandable ways. This also allows you to highlight trends or any abnormal values; colors and patterns create an overall picture in a way difficult to reproduce from the columns and rows of a spreadsheet. In a BI system, data visualization is used in reports, query responses, and dashboards.

Dashboards are directly related to reporting, which is another relevant BI's function; this tool presents data and in-depth information to end users in easily understandable and usable ways. Reports use summaries and visuals such as diagrams and graphs to show users time trends, relationships between variables and numerous other factors; they are also interactive, allowing users to break down and analyze tables or drill down data as needed. Reports can be automated and sent regularly based on a predetermined schedule, or ad hoc and created in a timely manner.

Moreover, BI instruments are used as query execution tools, as they allow users to quickly ask business questions, and get answers through intuitive interfaces.

Business intelligence also makes data preparation more automatic; this process consists in compiling multiple data sources and generally preparing them for analysis. Using a process called "extraction, transformation, and loading" (ETL), raw data is cleaned, classified, and then uploaded to a data warehouse (Turner, 2022). In this regard, efficient BI systems automate many of these processes and allow the definition of dimensions and indicators.

Finally, the role of Online Analytical Processing (OLAP), which is a technology that enhances data exploration capabilities in many business intelligence systems, is also relevant. This technology enables fast, multidimensional analysis in huge volumes of information stored in a data warehouse or other central data store.

Fig. 3.1. Example of a power BI's dashboard showing financial performances across different countries and business units (source: SAP Analytics Cloud)

3.2. Data quality for insurers

From the information in the previous paragraphs, using an approach that effectively manages data quality is necessary for the success of insurance companies. Although insurers have always based their decisions on exposure, risk and customer information, the entry into force of the Solvency II Directive in 2016 was a key incentive for insurers to move from informal data quality management to a more structured approach; in this regard, Solvency II is the first regulation to introduce strict and detailed data quality requirements for insurers.

According to Turner (2022), the importance of data quality is also reflected in the situation where the reported Solvency II data are used by national authorities in the supervisory review process, by the national central banks as an input in the compilation of insurance statistics, as well as by EIOPA and the European Central Bank (ECB) for various market analyses.

The main definitions of data quality defined by regulations are accuracy, completeness and adequacy. In addition, current regulations allow you to work with internal and external data and focus on sufficient documentation of the data used, especially in the event of data limitations.

The GDPR establishes, as stated in the previous chapters, uniform rules for the processing of personal data by most processors, both private and public, throughout the EU; on the one hand, ensuring the protection of personal data within the European Union and, on the other, ensuring the free movement of data within the European internal market. Since insurance companies manage significant volumes of private data, GDPR plays an important role, although for calculations or analysis, most of the data is already aggregated (anonymized) and therefore not subject to GDPR (Zanelli, 2022). Specifically, the concept of data quality is mentioned only in Article "47 (2.d): Binding Corporate Rules[28]".

The International Financial Reporting Standard (IFRS) does not provide any specific data quality requirement, but since it represents the core accounting and accounting policies, the

[28] "[…] The binding corporate rules […] shall specify at least […] the application of the general data protection principles, in particular purpose limitation, data minimization, limited storage periods, data quality, data protection by design and by default, legal basis for processing, processing of special categories of personal data, measures to ensure data security, and the requirements in respect of onward transfers to bodies not bound by the binding corporate rules […]." (GDPR, Article 47)

quality of the data is of great importance; in addition, financial results are used in many other processes of an insurance company (e.g., internal models, reporting, pricing).

Responsibility for the quality of Solvency II reporting (including the quality of reported data) usually lies with the Chief Risk Officer (CRO), the Chief Financial Officer (CFO) or a similar function. CRO can also define data governance policy, chair an eventual data quality committee, identify critical data fields and monitor data quality indicators, but the level of CRO involvement varies among organizations.

Furthermore, a relevant role is played by the Chief Data Officer (CDO), but other professional figures often take responsibility for structuring and managing data. Once established, the CDO typically reports to the head of another domain such as IT or Risk Management and can manage or supervise data quality risk; in this regard, the CDO may be an operator or facilitator (Mahanti, 2021).

According to a 2020 report by CRO Forum (already mentioned in the previous paragraph) data quality is generally managed as part of operational risk management; in this regard only 35% of insurance corporations assess risk in economic terms and have quantitatively defined risk tolerance. Risk tolerance for data quality can be defined on the basis of experience or calibrated on the basis of the impact and probability of the risk.

A report published by GIRO (General Insurance Research Organization) in 2020 examines the effect of data quality problems on critical financial quantities; in this regard, a data quality experiment was conducted using actual data used for an actuarial application.

The experiment was designed to examine the effects of incomplete and/or incorrect data on estimates of loss reserves; in this regard, the actual loss data was considered more convincing than conducting the experiment on a simulated dataset. Sufficiently mature data were obtained, fully developing all years and knowing the actual final losses, and various methods were used to estimate final losses using data from past valuation dates.

As result, one of the data challenges that insurers frequently encounter, concerns datasets that are severely limited by the completeness of the information provided. Subsequently, data may be limited by the number of years of history (e.g., only five years of history for a long queue where complaints take twenty years to resolve completely) or the types of data provided (e.g., only paid and incurred losses, but no reported claim count, closed claim count or exposure data).

The results of the data experiment indicated a significant increase in the uncertainty of the results and a significant decrease in the accuracy of the results when data quality problems were present; errors resulting from poor data can therefore significantly reduce the reliability of actuarial analyses, which could have a direct effect on an insurer's balance sheet. According to GIRO, insurers should devote more time and resources to increasing the accuracy and completeness of their data by improving their data collection and processing practices. In particular, insurers would benefit from the investment of an increase in senior management time in this sector; by taking such measures, they could improve their efficiency and thus their profitability.

As the other types of corporations, also for insurance industry having a data quality policy in place with group-wide principles is also the first step for the proper implementation of data quality management. In this respect, the extent to which data quality policy is defined centrally or peripherally depends on the specific organizational structure. According to experts (Olavsrud & Fruhlinger, 2023) in a changing environment it is important to periodically review the data quality policy at least once a year.

In addition, data quality management is structured according to roles at company level (for the definition of standards and assistance to local authorities in data quality) and roles within local entities. In this regard, to maintain the alignment between these roles, it is necessary to have communication structures linking the center of the organization and the peripheral nodes, for example committees (Mahanti, 2021). According to CRO Forum three categories of committees can be identified.

The first category is at strategic level, also known as Data Governance Board, that includes the CDO and representatives of the departments involved in data management; in this regard the main figures and departments are CRO, Actuarial, Finance, Internal Controlling, Information Technology, Chief Information Security Officer and Data Privacy Officer, since the Board of Directors must be regularly involved in data quality investments.

The other two categories of committees are tactical and operational, that include, respectively, data owners and data stewards for each legal entity.

In addition, a dedicated team or function is needed to structure and maintain the organization and data quality committees; CDOs may have a more strategic or security role, by also facilitating the implementation of the data quality policy and supervising the observance of procedures (Turner, 2022). In insurance companies, the CDO can be the leading figure in data quality issues, sign-off on data quality reports and chair the Data Governance Board.

Moreover, data owners are responsible for the quality of their data and for empowering their teams to perform work according to centrally defined principles and policies; in general, data owners maintain the quality of a defined set of data. On the other hand, in insurance companies, data stewards are often more informed about the data and therefore play an important role in the definition and metadata, setting up data quality rules and monitoring data quality (CRO Forum, 2020).

Another essential aspect is the role of CRO and Risk Management; in fact, they are fundamental in the definition of policies and key data fields, and in the monitoring of data quality (Mahanti, 2021).

Finally, the compliance function monitors fulfillment with legislation, prepares policies, and complements the second line of defense for monitoring data quality and acts on data quality compliance issues; there is also close collaboration and interdependence between the IT department and the data management team (Irwin, 2022). In this regard, a connecting role is played by data custodians, who are responsible for maintaining data on the different IT systems used in the organization (Zatyko, 2015).

In the following figure it is represented an example of business team dedicated to the monitoring and the attainment of data quality initiatives; it should be specified that such structure varies according to the corporation type, but is particularly common in banking and insurance sector.

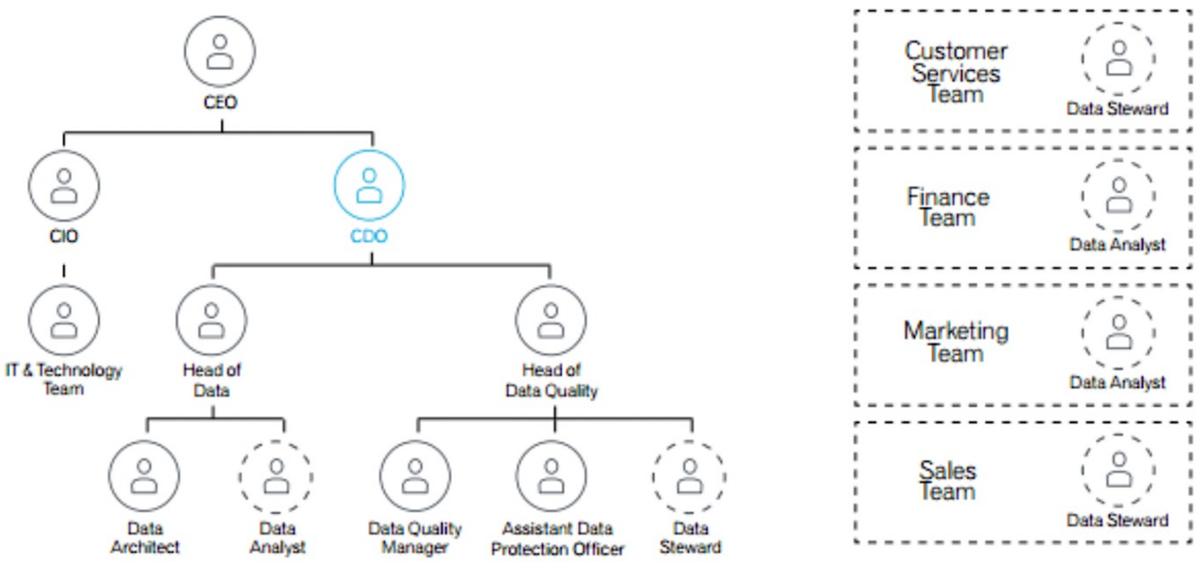

Fig. 3.2. Example of data quality structure within banking/insurance companies (sources: Zatyko, 2015)

3.3. Data quality metrics and indicators

In general, especially in the insurance industry, the IT department is actively involved in the implementation of automated controls related to data quality standards set within a corporation.

Such quality controls are carried out using appropriate tools, whose main features are directories and repositories (for data, metadata or controls), execution checks and collection of results, data quality reporting, exception management, data and process mapping, data extraction and transformation, data profiling, data lineage and audit trail (CRO Forum, 2020). In this regard, artificial intelligence can also be integrated to detect and help solve data quality problems not identified by traditional systems.

Data quality controls and indicators are generally based on regulatory requirements and data quality control frameworks or data governance frameworks (e.g., Data Governance Institute's Framework[29], Data Management Association's Body of Knowledge[30]). Among the regulatory requirements, the previously mentioned Solvency II and Sarbanes Oxley[31] frameworks are common; according to experts, Six Sigma[32] process quality systems are also relevant (Olavsrud & Fruhlinger, 2023).

Furthermore, data should be controlled throughout production, storage and processing; automated and manual controls should be placed on data flows, with acceptance thresholds.

According to scientific literature (Mahanti, 2021) a control plan should be formalized and updated constantly for a regular review of all key controls. In this regard, the experts recommend identifying the risk to the quality of data on the basis of a materiality assessment or advice from professionals; this risk may be assessed separately or within an operational risk management framework (ORM) or an integrated control and risk system (IRCS).

Risk probabilities and impacts are often estimated basing on expert judgement, that have the possibility to use scenarios with different types of consequences in support of their estimates

[29] Founded in 2003, the Data Governance Institute is one of the most authoritative sources of data governance guidelines and frameworks, as well as materials to support companies of various types.

[30] The Data Management Association's Body of Knowledge (DMBOK) is a document published by DAMA in 2015 (first edition, DAMA-DMBOK) and in 2017 (second edition, DAMA-DMBOK2), which contains best practice tips for a common business data management language.

[31] The Sarbanes-Oxley Act 2002 is a federal law of the United States that imposes certain financial record-keeping and reporting practices for companies. The Act contains eleven sections that impose requirements on all US boards of directors of public corporations and accounting companies

[32] Six Sigma is a methodological, rigorous and highly structured approach oriented to the radical improvement of processes in terms of performance and robustness. It consists of 5 well-defined phases: DMAIC- Measure, Analyze, Improve and Control. Each phase has inputs, well-defined outputs and a set of specific tools to implement to ensure the success of the project.

(Irwin, 2022). Risk tolerances for data quality are determined in various ways, for example based on specific data quality criteria, reporting thresholds, materiality frameworks, expert judgement and potential impact/probability sets.

Therefore, experts recommend carrying out a regular assessment of the data quality management environment, be it a self-assessment or an external report. For this purpose, companies use data quality rules, also known as data validation rules, which define business requirements for specific data and can be used to verify data quality; using these rules, it is possible to identify potential weaknesses and make recommendations for action (CRO Forum, 2020). Data quality rules allow measurement of different data quality dimensions, such as completeness and accuracy.

As a result, data quality indicators are used by management to monitor data quality; in the following paragraphs, the main indicators used by the companies to verify the quality of the data are reported.

3.3.1. Accuracy

Academic literature defines data accuracy as the level at which data represents the real scenario and confirms this with a verifiable source; in this regard, the accuracy ensures that the real-world entities associated with data can participate as intended.

Scientific literature (Cesarini et al., 2014) provides a similar definition: data accuracy is described as the distance between a data and a value considered as the correct representation of the real phenomenon that the data intends to express.

The Cesarini report (2014) also refers to *syntactic accuracy*, through the principle of verifying that the value attributed to a data is present in the set of values of a domain. In case the value is not present in the domain, it is possible to obtain "close" values; in this regard, the concept of proximity can also be defined using different metrics, such as the number of characters of an alphanumeric string. Later it is possible to identify a set of permissible values, with different degrees of similarity to the real value; in the case of syntactic accuracy, however, the main focus is not comparing the value of the data with the real value, but with the set of all the domain values. As a result, the value of the domain closest to the real value is used to determine the accuracy of the data.

Cesarini (2014) also defines the concept of *semantic accuracy*. In this case the accuracy of data is evaluated by comparing it with its real counterpart; the result is that it is essential to know the real value of the attribute that is intended to express through the data. Unlike syntactic accuracy, which measures the distance between the observed value and the real value as a numerical value, semantic accuracy provides a dichotomous evaluation; consequently, either the data is as accurate as the real value or not, regardless of the distance between the real and domain values. Thus, thanks to semantic accuracy, the concept of data correctness is expressed intrinsically.

Considering these factors, the accuracy of measurement data requires verification with authentic references. Data accuracy also has a strong impact on how data is stored throughout the entire journey, and successful data governance can promote this data quality dimension (Mahanti, 2021).

High data accuracy can drive accurate reports and reliable business results, making this indicator particularly critical for highly regulated industries, including banking and insurance.

Scientific literature (Hu, 2022) states that one way to ensure the data accuracy is through the detection of anomalies, sometimes called outlier analysis, which helps identify unexpected values or events in a dataset.

In this regard, softwares used can instantly recognize if a data is outside the normal range, because machine learning model learns from historical metadata.

The causes of data inaccuracy are multiple. First, data imprecision is the result of bad data entry practices; an organization without adequate data governance and data quality strategy tends to store data in multiple formats, styles and varieties more frequently (Zanelli, 2022); still forward, data acquired from sources like social media are very prone to errors, typos and copy/paste errors.

Secondly, data accessibility may not be regulated; if accessed simultaneously by sellers, marketers, customer service and account managers, CRMs can become a duplicate, inconsistent and inaccurate data source.

Finally, data quality and data quality are usually not addressed. According to Irwin (2022), leadership is often too busy in evaluating investments in cloud and big data systems, while IT teams are too busy helping leadership in transforming data into information to worry about disparate data and duplicates. Quality or accuracy of the data is not a matter of discussion by the board; consequently, a poor data quality policy tends to become clear only when a drastically negative event occurs.

The costs of poor data accuracy can therefore be considerable. In general, according to a 2022 Gartner report, inaccurate data costs companies about 15% of their revenue, while the average financial impact of poor data quality on organizations is $9.7 million per year.

In addition, in the US alone, businesses lose $3.1 trillion a year due to poor data quality. (IBM, 2022), while more research reports have shown that bad data is on average costing businesses 30% or more of their income.

These statistics show that inaccurate and poor data is a persistent problem in most organizations and that they have a huge impact on profitability, company reputation and customer trust.

3.3.2. Completeness

According to scientific literature, completeness can be defined as "the level of amplitude, depth, and appropriateness of a datum according to its purpose" (Wang & Strong, 1996).

As consequence, all required records and values should be available with no missing information; with completeness, the stored data is compared with the goal of being 100% complete. So, the concept of completeness does not measure accuracy or validity, as it concentrates on the information that is missing. Data completeness is therefore achieved when all the required data is present for the dataset to fulfil its intended purpose; completeness does not mean that 100% of all data fields must be complete, since it is about ensuring that the relevant, meaningful fields are complete with the right data for the job.

In this regard, to better describe the size of completeness it is possible to consider the structure that stores data as a table (relationship); the columns represent the attributes of the object to store, while the rows of the table (tuple) represent the different observations of the object.

According to Cesarini et al. (2014) it is therefore possible to distinguish between different types of data completeness. *Column completeness* measures the lack of specific attributes or columns from a table (an example of incomplete column is shown in Fig. 3.3); *population completeness* instead analyzes the missing tuples with reference to an observed population. On the other hand, some levels of completeness are difficult to assess.

CustomerID	CustomerName	CustomerBirthDate	CustomerAccountType	CustomerAccountBalance	LatestAccountOpenDate
100000192	Robert Brown	4/12/2000	Loan	40390.00	12/20/2026
100000198	Maria Irving	12/1/2025	Deposit	-13280.00	10/21/2018
100000120	Ava Shiffer	10/31/1990	Credit Card	320	3/1/2020
100000192	Robert Brown	4/12/2000	Deposit	40390.00	12/20/2026
100000124	Matthew Martin	5/9/1965	Deposit	70102.00	5/4/2022
100000149		2/4/1988	Loan	0.00	9/20/1990

Fig. 3.3. Example of missing data from a column (source: Datacamp)

Experts state that one way to ensure data completeness is through anomaly detection, sometimes called outlier analysis, which allows you to identify unexpected values or events in a dataset; in this regard, anomaly detection helps countless data managers to detect incomplete data.

Using the example of prices missing from a table, the anomaly detection software can notify immediately when the expected data does not arrive; in relation to what already stated for the data accuracy, the software knows that it is an anomalous state because its machine learning model learns from historical metadata.

3.3.3. Consistency

Consistency is defined in the literature, referring to the "violation of one or more semantic rules defined on a set of data" (Batini & Scannapieco, 2006); data consistency refers therefore to the way values and formatting in the dataset match each other, avoiding conflicts among values throughout the dataset. Just like the previous quality indicators, also in this case it is possible to identify various levels of consistency.

Key consistency is the simplest form of consistency and requires that, within a relationship scheme (a table), there are no two tuples with the same value as an attribute used as a key (Cesarini et al., 2014); an example is the tax ID code used as key. In this case the so-called key would require that there are not two people with the same tax code; a perfect homonymy of names, dates and places of birth, although extremely unlikely, would violate this bond of consistency, showing the inadequacy of the tax code field to carry out this task.

Inclusion consistency is linked to the "foreign key" of a relationship, which is the key that connects several tables. It requires, in fact, that a set of columns of a schema of relation is contained in another set of columns of the same schema of relation, or another instance of schema of relation.

In relation to the previous example, it is possible to consider, in addition to the personal data table, also a second table in which data on households are stored, uniquely identified by the tax code of the head of the family. The nuclear and population tables are related to each other through the tax code field of the citizen head of the family, representing the "foreign key" of the relationship. For the consistency of inclusion (also known as a bond of referential integrity) to be satisfied, all the heads of a household must be present in the master chart.

Functional dependencies are the most known and used. In general, given a relationship with attributes X and Y, it is said that Y is functionally dependent on X if and only if, for each value of X, a precise value in Y is associated (Cesarini et al., 2014). In other words, given a tuple with a value of X the functional dependence expresses the ability to know with certainty the value of the attribute Y. In the previous example, the tax code field is functionally dependent on the fields necessary for its calculation (i.e., name, surname, date of birth, place of birth, sex). In fact, once the fields are known, it is possible to generate one and only one tax code associated with them.

The definition of consistency is generic and therefore allows the modeling of a large amount of "semantic rules”, for which it is necessary to implement ad-hoc solutions (Wang & Strong, 1996).

3.3.4. Currency

According to scientific literature (Batini & Scannapieco, 2006), data currency is not a financial reference, but a temporal one. The Data Managemennt Body of Knowledge (DMBOK) defines it as "the degree to which data is current with the modeling world". Many times, the right information about a server or other IT asset is stored in the database, but if it has then been renamed, moved or re-purposed, the data needs a further update.

In this regard, updates can be manual or automatic; they can take place as needed or can be scheduled periodically, everything depends on your business needs. The business rules that define your approach to this are called "data currency" rules (Wang & Strong, 1996).

Data currency is a common issue for IT asset repositories and configuration management databases (CMDB). Usually, such projects are financed as a special capital project, which means that to get support for ongoing operations could be challenging. Furthermore, without stationary operating processes to maintain data, they will inevitably decay, and the entire capital investment is therefore at risk as the repository loses credibility (Mahanti, 2021).

In this regard, strong management support and continuous improvement approach to data quality issues are needed to protect against this; the value of updating the archive should be assessed and, on the contrary, the costs and risks of data inaccuracy for long-term success should be considered.

3.3.5. Relevance

Considering the characteristics of data quality, relevance plays a key role, because a firm should always know reasons and purposes of the data collection (Mahanti, 2021); it is therefore necessary to consider if such collected information is necessary for business or if it only plays a marginal role.

As a result, relevance is important as a data quality feature since, if the company is collecting irrelevant information, it may be consuming valuable resources as well as time and money, and data analytics won't be so precious.

Scientific literature (Zanelli, 2022) also links the concept of data relevance to the spreading of data culture and effective communication between different company's departments; in fact, in order to understand which data is relevant for the business, each division and department should be aware of its function within the organization, the impact of its role and the objectives that the company intends to pursue.

3.3.6. Timeliness

Data timeliness is one of the most relevant data quality indicators; this indicator is linked to availability and accessibility of data in making business decisions. In this regard, relevant decisions within a company require that data is accessible, well organized and available when needed. This approach to data management drives smart decisions and enables a better understanding of what to expect in the future (Antonopoulos, 2022).

Furthermore, data can be updated in real time to ensure its availability and accessibility; consequently, timeliness can be measured as the time between the time when the information is expected and the time when it is readily available for use, and the success of business applications based on master data depends on consistent and timely information (Cesarini et al., 2014).

Therefore, experts (Batini & Scannapieco, 2006) indicate how appropriate defining service levels that specify the rate of data propagation through the centralized repository, so that compliance with these timeliness constraints can be measured.

In this regard, a data delay is a latency between certain events in a data pipeline, and can be predictable (e.g., when a source preprocesses or groups data values before loading), or unpredictable, (e.g., when the data flow is interrupted by a sudden slowdown or network shutdown). According to Antonopoulos (2022), timeliness checks are therefore designed to detect anomalies in the transport of data; depending on the context, a certain threshold value is set, and any delay above the threshold is an anomaly and is signaled by a timeliness check.

Finally, the collected data should come from reliable data sources, including clients, studies conducted by government agencies, academic institutions, or public datasets; in an ever-changing environment such as insurance, the information collected must be associated with a time frame and, as consequence, continuously updated in order not to become contradictory.

Scientific literature (Batini & Scannapieco, 2006) distinguishes also among different types of timing and delay. *Ingestion time* is the time interval between the start of the data collection phase and the time when that information is conveyed within a data warehouse. The *loading delay* is related to that time interval, as it is the time difference between the events occurring during the ingestion phase and the time when the files corresponding to those events are in the data warehouse.

The *data freshness* is instead the time lag between the moment in which the data are cataloged inside the data warehouse and the moment in which it is accessible and usable. The *current delay* is, therefore, the time lag between the current timestamp and the timestamp when the last file appears in the database.

A graphical representation of such timing can be appreciated in the following figure.

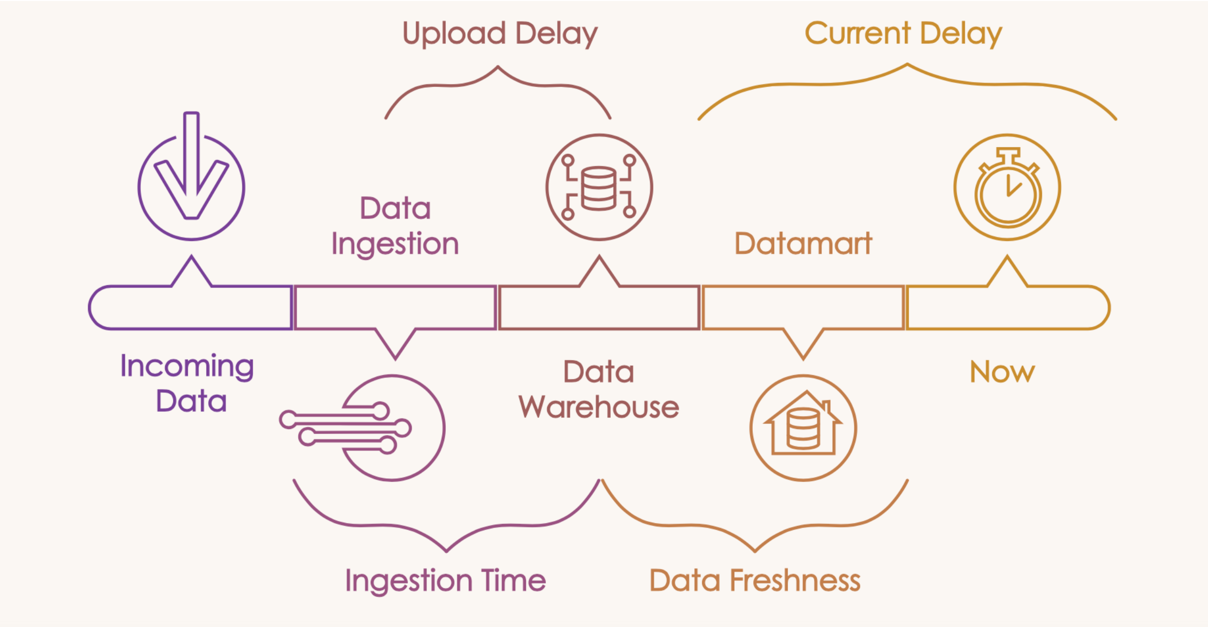

Fig. 3.4. Representation of data delays and timing (source: DQOps)

3.3.7. Uniqueness

Scientific literature (Batini & Scannapieco, 2006) states that the uniqueness of data is obtainable if the information contained in a dataset appears only once; in this regard, such dimension of data quality measures the extent of duplication and, subsequently, the degree of redundancy related to data. Having unique information across datasets is particularly relevant also for insurance corporations, associating a given client with service provided.

For example, the uniqueness of the data would identify instances where multiple data entries are present for the same contact or the same field.

Although some fields are unique between two records for the same contact, this would still be considered duplicate data; for instance, a client's contact might be in the database twice, with two different phone numbers. In this regard, chances are that only one of these addresses is accurate, which makes it important to ensure the uniqueness of the data throughout.

A very simple and basic representation of the uniqueness concept is shown in the figure below; in analogy to the previous example, customers' data are repeated twice within the same database, causing conflicts between certain information fields reported.

CustomerID	CustomerName	CustomerBirthDate	CustomerAccountType	CustomerAccountBalance	LatestAccountOpenDate
100000192	Robert Brown	4/12/2000	Loan	40390.00	12/20/2026
100000198	Maria Irving	12/1/2025	Deposit	-13280.00	10/21/2018
100000120	Ava Shiffer	10/31/1990	Credit Card	320	3/1/2020
100000192	Robert Brown	4/12/2000	Deposit	40390.00	12/20/2026
100000124	Matthew Martin	5/9/1965	Deposit	70102.00	5/4/2022
100000149		2/4/1988	Loan	0.00	9/20/1990

Fig. 3.5. Example of dataset with non-unique data reported (source: Datacamp)

3.3.8. Validity

Data validity refers to the degree in which data is reported to the correct format, type or range (Batini & Scannapieco, 2006); in this regard, data should exist within the appropriate limits to be considered valid. For instance, a month should be between one and twelve, and anything else would be considered invalid. Or, referring to insurance companies, just a service actually provided to customers should be accepted in databases.

Most of the time, if data is invalid, it is unintentional; subsequently, experts (Antonopoulos, 2022) underline that it is important ensuring validity with regular data cleanup programs.

Furthermore, the themes of data validity and data accuracy are usually considered very similar, although they should not be confused as the same size of the data quality; in fact, a data entry may be valid from a formal standpoint, but at the same it could be not accurate. For instance, a customer might enter a valid postcode that does not reflect his/her real address.

Like in the previous paragraph, the following figure shows a brief example of data invalidity. In the reported case, the fields “CustomerBirthDate” and “LatestAccountOpenDate” should contain a date in the past, while value reported under “CustomerAccountType” should be either “Loan” or “Deposit”.

CustomerID	CustomerName	CustomerBirthDate	CustomerAccountType	CustomerAccountBalance	LatestAccountOpenDate
100000192	Robert Brown	4/12/2000	Loan	40390.00	12/20/2026
100000198	Maria Irving	12/1/2025	Deposit	-13280.00	10/21/2018
100000120	Ava Shiffer	10/31/1990	Credit Card	320	3/1/2020
100000192	Robert Brown	4/12/2000	Deposit	40390.00	12/20/2026
100000124	Matthew Martin	5/9/1965	Deposit	70102.00	5/4/2022
100000149		2/4/1988	Loan	0.00	9/20/1990

Fig. 3.6. Example of invalid data reported (source: Datacamp)

3.4. Data quality vs data governance framework

From what is described in the previous paragraphs and chapters a data quality and data governance framework relate closely to each other within a data analysis-driven strategy; this statement, of course, also includes the scenario of the insurance industry.

According to experts (Russell, 2020) "people have been making data quality and worrying about data quality for far more years than they have given governance"[33]; the consequence is the perception that there are two different frameworks in action. However, companies often tend to unify these two factors, merging them into a single framework.

In fact, scientific literature (Mahanti, 2021) defines data quality as the degree to which data is accurate, complete, timely and consistent with the needs of the company, while data governance is a framework to proactively manage data in order to help the organization achieve its goals and business goals by improving data quality. As a result, data governance helps protect the business, but it also helps optimize efficiency, ensuring that trusted information is used for critical business processes, decision-making processes, and accounting. So, data governance has a "foundation" role for many data management disciplines, as its main purpose is to manage and improve data quality.

[33] The phrase is a quote from an interview made in 2020 by Sue Russell to Nicola Askham, consultant for the implementation of data quality and data governance frameworks.

According to Russell (2020), the main point is not the alignment of the two frameworks, as it would be advisable for companies (particularly in the banking and insurance sector) to have only one framework, and data quality and data governance should work in harmony with each other, not against or in opposition.

Consequently, governance and data quality rely heavily on each other; in this regard, N. Askham (2020) describes the relationship between them as 'symbiotic', as their relationship is based on mutual interdependence.

Askham also states that some organizations (around 50%) have not yet fully realized that it is necessary to implement both aspects; in fact, while it is rare to find a company that implements a data governance framework without the intention of improving data quality, it is quite common for societies to start data quality initiatives without implementing a data governance framework to support them. The result is that many data quality initiatives are just tactical solutions that yield short-term results.

Therefore, according to Russell (2020), it is enough to have a single data quality framework that encompasses data roles and responsibilities, IN such a way that there is no duplication, no gaps between two different frameworks.

Another factor that describes the relationship between the roles of data governance and data quality is the DAMA-DMBOK framework, which delves into the Knowledge Areas that constitute the general scope of Data Management; this framework is well described by the image of a wheel called the DAMA Wheel.

The DAMA Wheel defines the Knowledge Areas of data management; it places data governance at the center of data management activities, since governance is needed to ensure internal consistency and balance between functions. The other Knowledge Areas (data architecture, data modelling, etc.) are balanced around the 'wheel'. They are all necessary parts of a mature data management function, but can be implemented at different times, depending on the organization’s requirements (DAMA International, 2017). The figure below shows a representation of the DAMA Wheel.

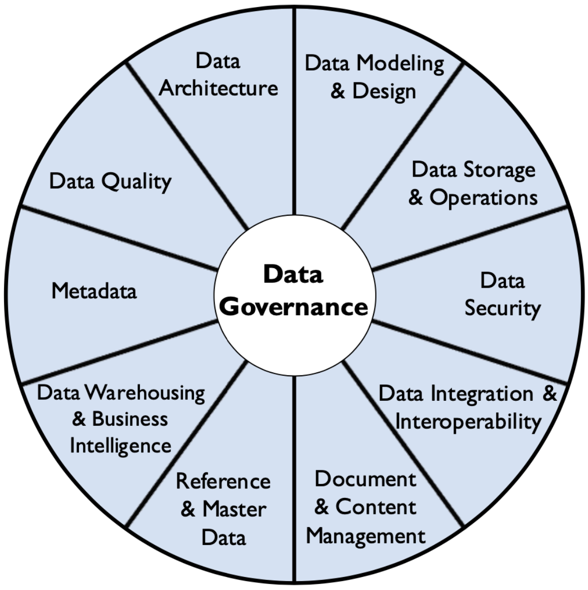

Fig. 3.7. DAMA-DMBOK2 Data Management Framework (DAMA Wheel) (source: DAMA International, 2017)

Another model describing the role of data quality and derived directly from the DAMA Wheel is the framework developed by Peter Aiken, which uses the DMBOK functional areas to describe the situation in which many companies find themselves; an organization can use it to define a path to a state in which it has reliable data and processes to support strategic business objectives (DAMA International).

According to Mahanti (2021), in an attempt to achieve this goal, many companies go through a similar logical progression of steps, starting with the purchase of an application that includes database functionality and having a starting point for data design, data storage and data security (phase 1).

Once they start using the application, companies usually find it difficult with the quality of their data, so they have to rely on a consistent data architecture to manage information from different systems (phase 2).

In addition, disciplined practices for managing data quality, metadata and architecture require Data Governance that provides structural support for data management activities, enabling the execution of strategic initiatives such as Data Warehousing and Business Intelligence (phase 3).

Finally, the organization can reap the benefits of well-managed data and improves its analytical capabilities (phase 4). A representation of this framework is shown in Fig. 3.8.

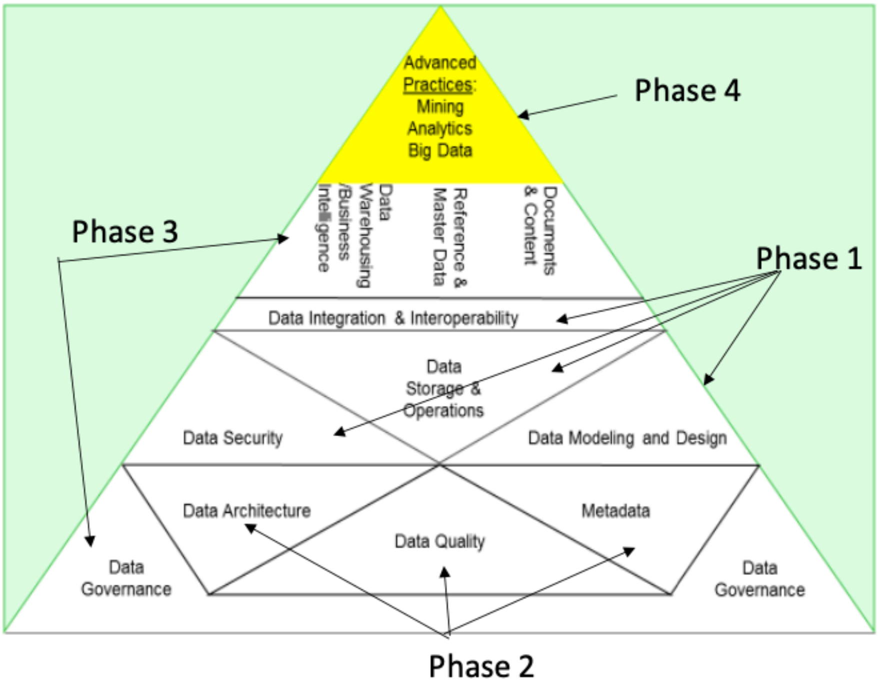

Fig. 3.8. Aiken Pyramid Framework (source: DAMA International, 2017)

So, reliable data quality, data design, and data interoperability practices underpin consistent systems and applications.

Moreover, according to Zanelli (2022), organizations focused solely on direct lifecycle functions will not get as much value from their data as organizations that support the data lifecycle through control and oversight. Core activities, such as data risk management, metadata and data quality management, cover the data life cycle, enable better design decisions and simplify data usage (DAMA International).

When properly executed, data is less expensive to manage, data customers have more confidence in it, and the opportunities for using it expand. So, in order to successfully support the production and use of data and ensure that fundamental activities are rigorously performed, many organizations establish oversight in the form of data governance (Mahanti, 2021). As already stated, a data governance program allows the organization to be data-driven, implementing supporting strategies and principles, stewardship policies and practices that ensure that the organization recognizes and acts on opportunities in order to gain value from its data. Such a program should also translate into organizational change management activities to instruct the organization and encourage behaviors that allow the strategic use of data (DAMA International). Therefore, the need for cultural change extends across the breadth of data governance responsibilities, especially when an organization matures its data management practices.

To conclude, data governance activities provide oversight and referrals, through strategies, principles, policy and stewardships; in this regard, they enable consistency through data classification, data evaluation and a rational strategy based on data quality management.

The concepts described in these chapters will be taken up again in the next section, which will focus on my internship activity in EY and on the application of data governance and data quality frameworks in corporate contexts, in particular within a project with a large insurance company as a client; the notions introduced are therefore necessary for a better understanding of the following chapter.

4. Applying knowledge in an internship project

In this section of the paper, the notions expounded in the previous chapters will be useful to better understand what will be described next. In particular, the main focus will be on the internship I completed at EY; the corporation is a world leader in professional audit and accounting, tax and legal assistance, transaction and consulting services. The company is part of the Big Four group and includes, among its main integrated service lines, Assurance, Consulting, People Advisory Services, Tax & Law and Strategy and Transactions.

Specifically, the internship took place within the Technology Consulting service line, in the Data & Analytics team, under the supervision of my counselor. The role of the Data & Analytics team is to support clients on issues related to process digitization, through consultancy in areas such as BI & Reporting, Artificial Intelligence, Big Data and Data Engineering.

The client of the project to which I have been assigned (that for reasons of confidentiality will be called Alpha Company) is one of the largest players in the insurance and asset management sector on the Italian and European scene, with several legal entities and subsidiaries in dozens of European and non-European countries, and a capitalization of almost EUR 30 billion. Globally, the insurance group has more than 80,000 employees and has seen a good performance in terms of turnover in recent years, with a constantly growing operating profit and ROE.

The main objective of the project was, therefore, to develop a dashboard for the real-time monitoring and visualization of the insurance balance sheet data according to IFRS9 and IFRS17, with the respective percentages of compliance, accuracy, completeness and robustness, and the number of timely and overdue sign-offs.

In order to achieve the objective, it was therefore necessary to create a single database comprising the controls and data (relating to balance sheet items and asset management elements) from the legal entities globally and collected by means of data management and analysis softwares (such softwares will be mentioned later).

The dashboard was aimed purely at Alpha's Chief Financial Officer, with the objective of having a clear and intuitive tool that he could use to monitor the performance of the various legal entities and business units as well as the degree of compliance with IFRS9 and IFRS17 standards of the data they uploaded and the respective numbers of recycles. In this regard, the term recycle is associated with the fact that the data coming from the legal entities do not comply with the aforementioned standards or do not meet the predefined data quality requirements (in this case, the most considered standards were accuracy and completeness); in this case, a legal entity has to upload the correct data again, which lowers the overall conformity index and the total performance of the legal entity considered.

The following paragraphs will explain the stages of the project that were briefly mentioned earlier (always bearing in mind the confidentiality constraints imposed by the project), attempting to reconcile the theoretical part described above with a more practical approach.

4.1. Alpha's organizational scheme and DQ policy

The first phase of the project was mainly aimed at a study of the data quality policy of Alpha, gathering the essential elements to decide the approach to be used later; in particular, the study was aimed at deciding on the existence of a central body within the company that controls at the aggregate level the quality and compliance of data from legal entities, or if such controls are done only at the periphery, following any guidelines laid down centrally.[34]

In this regard, the objective of Alpha's data quality policy is to ensure reasonable reliability on the quality of the data and information provided to internal and external customers, the market, supervisory authorities and other stakeholders, adequately supporting decision-making processes. This policy therefore ensures greater timeliness in decision-making, ensures accountability in data quality management, ensures stakeholder confidence in data warranting the adequacy and effectiveness of Group-wide Administrative and Accounting Procedures.

The group policy was approved by the Board of Directors of Alpha, on the proposal of the Group CEO, and is periodically subject to a review in order to incorporate the changes in the regulatory, market and/or best practices, or with reference to the Group's strategy and organization.

[34] The following information was collected following analysis of business documentation and interviews with Alpha's management.

As mentioned above, the company's policy also requires that the general principles of data quality are adopted in all Data Value Chain processes and IT applications and applied to all production data; in this regard, the most significant principles are accuracy and completeness.

By "accurate data" Alpha means that they are produced in the absence of any material error and omissions, they do not include material errors (with particular attention to production data contributing to the estimates made for the internal model or for the calculation of technical provisions), they must be verified at the time of receipt and, in any event, prior to their use, they must be recorded according to methodologies that make them comparable, they are timely and regularly recorded over time and, finally, they must be timely available, in order to foster effective decision-making processes and enable the company to anticipate and react promptly to future events.

In parallel, the company requires in its policy that the data be complete, with sufficient granularity and historical information adequate for the purpose for which the production data are used. Where applicable, this principle also implies that such data includes sufficient historical information to assess the characteristics of the underlying risks and to identify risk trends, is available for all relevant model parameters and no data is excluded without justification from use in the internal model; it is also available for each of the relevant homogenous risk groups used in the calculation of technical provisions, and the use of relevant data in the calculation of technical provisions is not excluded without justification. So, data contains all the information necessary to carry out a process successfully, covering all relevant aspects of the enterprise in terms of quantity and quality, and including indicators that may have a direct or indirect impact on the strategic planning of the activity.

According to the policy dictated by Alpha, moreover, the legal entities must carry out the first checks on the production data that will be used centrally for the production of relevant outputs (balance sheet and other financial documents). This scheme therefore presupposes the existence of an organizational hierarchy, in which the *legal entities* are the smallest part in terms of granularity. After legal entities, the larger "layer" consists of the *country* group in which entities are located; subsequently, a new macro subdivision is represented by *business units,* bodies not strictly geographical that enclose a certain number of country and legal entities.

Further quality control of the data provided is carried out centrally, at a group-level, through an integrated structure with the data governance office.

In this regard, group-level policy also defines the role of *data owners*, who are associated with a specific dataset. In the case of the project, the main datasets identified are IFRS9 and IFRS17. Data owners' role is defining risk-mitigating data quality controls and preparing the necessary documentation to ensure that production data complies with the general principles; they are also responsible for data quality and performs data quality checks to measure and monitor the quality of data. In this regard, they are also entitled for the timely execution of the Administrative and Accounting Procedures within the defined processes, and for the timely updating of the related documentation.

Data owners, supported by *data system owners* (who are associated with the systems and software used for data collection and management), must implement the data quality controls as designed, verifying their continuous operation and, finally, record the results of the checks carried out and ensure that the production data (necessary for the production of the final outputs, i.e., balance sheet and other financial statements) complies with the general quality principles.

So, data system owners ensure that data can be easily inspected and verified, and ensure that implemented IT processes guarantee data integrity as described in the general principles.

Furthermore, they cooperate to resolve identified data quality issues at IT level and support data owners in compiling data quality documentation.

In addition, if issues relating to the quality of data are identified that lead, or may lead, to non-compliance with the general principles, the data owners are entitled to register them in an issue register, providing information on the identified issues, the remedial actions and their deadlines, as well as ensuring their proper management until the resolution of the same.

In the event of a data quality issue resulting from failure to comply with the general principles, the data owner must analyze the event and identify remedial actions. In this regard, issues may arise, for example, from failure of key controls, evidence of error, external events, internal events, delays in the execution of the process, etc.

In general, the policy states that anyone who identifies a data quality issue must promptly report it to the data owner, who evaluates the issue, including its significance, and determines whether it can resolve it directly, without the intervention of other data owners and whether the quality of the data has been impacted.

The data owner, possibly with the support of specialists (e.g., data system owner), takes also charge of the issue to deal with the resolution in time and in compliance with the general principles.

Furthermore, as part of the implementation and monitoring mechanism, at least on an annual basis, and on a quarterly basis for the purposes of the executive in charge, local CEOs and local CFOs produce certifications to provide assurance on the effectiveness of the data quality policy. Other controls measures, check that production data complies with the general principles, that any issue of data quality is managed according to the guideline set by Alpha, that Administrative and Accounting Procedures are adequate, and the Organizational, Administrative and Accounting Structure is appropriate.

The role of *Chief Data User* also assumes relevance within Alpha; this figure identifies the data owners within its functional domain and plans the activities of the Implementation and Monitoring process. The Chief Data User is also responsible for the application of data quality standards in their functional domain (for example FSO, the domain covering all Financial Services Operations, etc.) and ensures that the services offered by the company comply with the standards set.

Another task of the Chief Data User within Alpha is to inform the Chief Data Officer if he/she identifies any issue that may lead to an untruthful and incorrect representation in any financial statement at company or group level.

Consequently, as mentioned earlier, Alpha has a structure for monitoring data quality and drafting related reports. This structure is integrated, as are other companies in the insurance sector (Mahanti, 2021) into a broader data governance office.

The structure is headed by the Chief Data Officer, who reports directly to the CEO of Alpha. In addition to controlling data from legal entities, the data quality office also sets the quality and compliance standards to which production data must adhere.

Nevertheless, a first and essential control of such data is carried out, as mentioned before, at the peripheral level by the individual legal entities, using the guidelines dictated by the central data quality structure. The data are collected and analyzed using different software, depending on the dataset of interest and the 'stage' of production of the associated relevant (in this point, a distinction is made between input data, transformation data, or intermediate data, and output data).

In the preliminary analysis carried out by my team, Alpha's management confirmed that the presence of such a data quality structure is not only due to compliance with current European law and regulations (e.g. the aforementioned GDPR). As already mentioned in the previous chapters, Alpha's managers indicated that the company considers investing in data control and data quality as a main point of its long-term value creation strategy, in order to achieve higher levels of efficiency than the competition and to cut costs related to error and issue analysis (also considering that in the insurance sector employees use 36% of their time in activities that do not create added value for the business, according to Petzold et al., 2020).

The company's trend corresponds therefore to the industry one, which is more and more customer-centric and cost-cutting oriented.

Once the structure and method by which Alpha guarantees the quality of the processed data had been analyzed, the project was able to move on to the next steps. The study of the data quality unit within the company's organizational chart was essential in determining the methodological approach to be followed in the next steps, as it also allowed the role of the legal entities and the controls performed by them to be given due importance.

The next steps of the project, related to data collection, conformity calculation and the creation of a single database, are described below.

4.2. Data collection and processing

This part of the chapter will focus on data collection and census, in order to create a single database and standardize the format of data from different sources as much as possible. At the end, this database would be processed and processed by power BI tools to create a dashboard for Alpha's CFO.

The data collection, census and standardization phase was the most time and effort-consuming as it constantly required informative summits with the client and clarification meetings with the data management software companies.

4.2.1. Data types, systems and initial subdivisions

The first part of the data collection was aimed at storing data from legal entities around the globe. A first division inherent in this information concerns the stage at which it is used to produce a relevant output (which, as mentioned in the previous paragraph, may coincide with the financial statements at the end of the fiscal period, or other financial statements).

Accordingly, the first subdivision that has been made of such data has delineated *input data*, *transformation data* (i.e., data used in the “intermediate” stage of the production process, travelling for instance between different legal entities or between different data management software), and *output data*.

Another subdivision that was made to the collected data concerned the *dataset* of interest. As mentioned earlier, the term dataset refers to the “imaginary table'”(Hu, 2022) or the macro-set to which the data are related. In the case of the project, the data transferred from Alpha's legal entities to the central level were classified taking into consideration only two datasets, namely *IFRS9* and *IFRS17*. This split at a general level was decided upon following discussions and briefings with the client. In this way, it was possible to classify the information received according to the following criterion: data relating to the compliance of financial and balance sheet elements with IFRS9 were classified under this dataset. A similar procedure was followed with IFRS17.

During the data collection process, another aspect to consider was the nature of the data itself. In fact, one of the first things that had to be done was to distinguish between *data elements* on the basis of their commercial and technical meaning as well as their purpose.

During the data collection process, another aspect to consider was the nature of the data itself. In fact, one of the first things to do was to distinguish the data elements according to their commercial and technical meaning and purpose.

A first part of the data corresponded to budgetary information on the company's performance during the fiscal period.

A second part of the data received, much more substantial than the first, concerned the *data quality controls* associated with the data, as well as their corrections. Accordingly, information was provided on the data quality controls (manual, automatic or semi-automatic) in force, such as the name of the control, the control rule used, the underlying data quality principles, the data under control and the ownership. The data quality standards for the execution of these controls are set centrally, as mentioned above.

In this regard, controls were performed multiple times by legal entities, registering the results time by time. The results of the checks were classified according to the categories “OK” (if the data checked are correct from a formal point of view and/or comply with the data quality standards set and the principles of IFRS9 and IFRS17) “KO” (if, differently from the previous case, data are not considered as compliant with qualitative and/or financial standards) and “Warning”; the latter class refers to data that are considered formally correct, but need to be checked for one or more elements from which possible criticalities could arise.

In addition, a *residual risk* factor should also have been associated with the different control results. This indicator can be defined as the combination of two parameters, namely the result of the data quality check on a specific data item and the risk that this data item may have a major influence on the quality of the final relevant output.

This factor was not provided in the legal entities' data, so it should have been calculated at the design stage; however, following discussions and briefings with Alpha's management, it was decided not to include this indicator in the dashboard immediately, but to add it during subsequent updates. However, for the sake of completeness, the figure below shows a graphical representation of a Residual Risk Matrix.

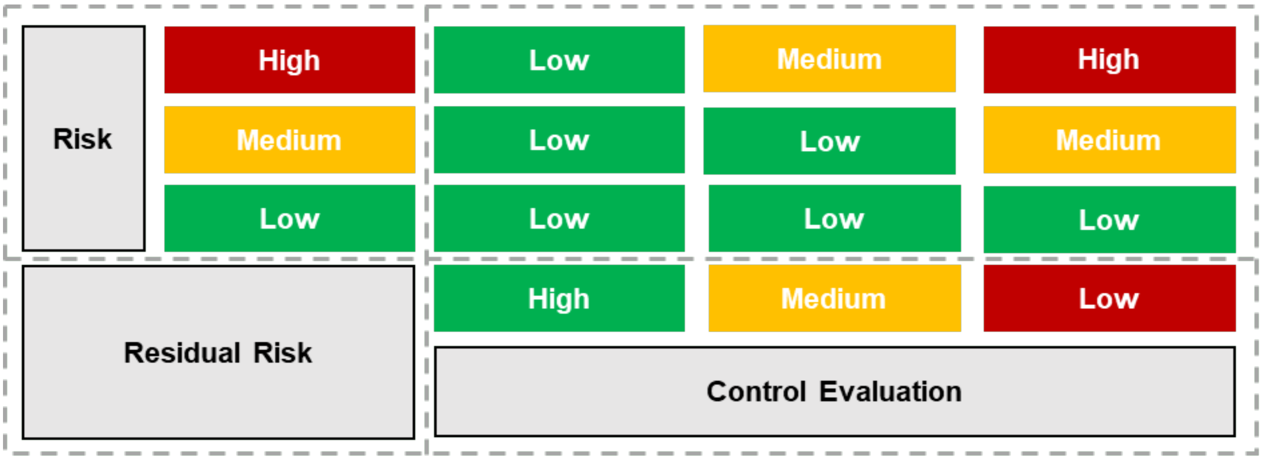

Fig. 4.1. Residual Risk Matrix (source: EY)

A further point of attention when classifying data is the specific *data flow* to which they belong. This classification was important to understand where the data flows come from, their destination and the respective processing systems. Consequently, the concept of data flow is closely linked to the location of legal entities and, to a greater extent, also to the IT and data analysis and management software systems that process this information.

Consequently, a crucial point was the identification of the different IT systems/software. As mentioned above, in fact, legal entities use such software to process different types of data and associated with different datasets; moreover, the processing of these data by the systems has different mechanisms and purposes, ranging from support in reporting to the calculation of ECLs[35], from the identification of key quality indicators to the monitoring of risk premiums for life insurances.

In this regard, the IT systems used by Alpha's legal entities to manage and analyze data are:

- *Data HUB*; used for various purposes, especially the KPI reconciliation. In this way, key performance indicators could be identified clearly and, as consequence, used to evaluate balance sheet, financial statements and overall accounting process. It processed intermediated data, so was especially used during the transformation phase.

- *TEAMTool*; used for the analysis of actuary claims data (Non-Life Actuarial Engine process). It usually works with input data, and also uses information processed by other systems (like Data HUB).

- *Rulebook & CTRL file*; used in relation with multiple topics, working with evaluation of pricing models linked to different insurance contracts.

- *Moody's*; used in relation with multiple topics, especially the risk assessment related to different insurance contracts and, subsequently, the evaluation of the respective risk premium.

- *Tagetik*; used to check the automatic validation of data, in relation with data quality rules and financial standards. It works with data coming from other systems, so data processed during the output phase of financial statements' production.

[35] ECL stands for "expected credit losses". This parameter can be calculated by multiplying the probability of the risk of insolvency by the estimate of the loss at the occurrence of insolvency by the exposure at the time of insolvency. In this regard, IFRS 9 requires that at each balance sheet date, if the credit risk has increased significantly after initial recognition, an entity shall measure the loss allowance as equal to the lifetime expected losses of the asset (lifetime expected credit losses). On the other hand, if credit risk has not significantly increased after initial recognition at the balance sheet date, an entity shall measure the loss allowance as the expected credit losses in the following 12 months (12 month expected credit losses) (Quindici, 2021).

4.2.2. Conformity computation

One of the main phases of the project, which was carried out in parallel with the creation of the database (which will be discussed later) is the calculation of conformity. This calculation was preceded and accompanied by a careful review of the data received by legal entities, in order to verify that the main principles of data quality (such as completeness, consistency, etc.) were respected and that there were no redundancies.

In this case, the term "conformity" refers, as repeatedly stated, to compliance with quality standards and IFRS9 and IFRS17.

An initial conformity assessment was carried out at the level of a single control; in this regard, considering a certain control and the multiple execution times, the results with the outcome "OK" were considered. The amount of such results has been divided by the total number of outcomes associated with the given control (therefore the same outcomes "OK", and the outcomes "KO" and "Warning").

Later, an analogous calculation was used to compute "Conformity with Warning", considered by identifying the "Warning" controls as "OK" ones; in this regard, the formula used is quite similar to the previous one, except that the numerator also considered the "Warning" outcomes.

At the level of control, recycling was also considered, so the number of times that a control was carried out; this parameter was reported as negative, as it could potentially underline the presence of a problem that requires more controls.

Later, conformity thresholds were set for all controls, according to which also the controls were associated with a result between "OK", "KO" and "Warning". The identification of the thresholds changed according to the IT system considered and the data processed; it also required frequent briefings with the customer and assumptions made by the team itself, if any identification problems persisted.

Following a procedure similar to the previous one, conformity was also calculated at data flow level; in this case, the "OK" controls were divided for the total controls associated with a given set.

Finally, knowing the conformities of data flows, it was possible to associate them to the corresponding legal entity; this factor would have been useful especially during the implementation of the dashboard through power BI, as it would allow to identify the legal entities (and therefore the business units) with the best and worst performances.

4.2.3. Data model design

The next step of the project involved the creation of a suitable *data model* as a base for the subsequent construction of the database and the final creation of the dashboard. In general, the designed data model contains a diagram, which is the image that captures the requirements in a precise form. In fact, it describes a level of detail, a schema (in this case relational) and a notation within that schema. Furthermore, definitions for entities, attributes and relationships are essential to maintain precision on a data model.

Another point to be emphasised is that during the project the data modelling process often raised problems and questions that were not necessarily addressed during the data modelling phase; therefore, the team delivered a document containing the current set of problems and questions, which were repeatedly addressed with Alpha's management.

One of the most recurring problems was the (often very subtle) classification and division between legal entities, countries and business units, as well as their respective hierarchies. Subsequently, the team included the *hub* category as a further intermediate classification, to be placed "between" countries and business units.

In addition, another very relevant aspect for the construction of the model was the knowledge of the *data lineage* so, as mentioned earlier, knowing where the data and its sources come from. In this regard, the preliminary phase of studying the data quality policy and the company organization chart, as well as the collection of related documentation, was very useful.

According to the DAMA-DMBOK, the lineage often takes the form of a source/destination mapping, in which it is possible to capture the attributes of the source system and how it populates the attributes of the destination system.

Furthermore, as it has been repeatedly observed during the project experience, there are two reasons why it is important to acquire lineage during data modelling. Firstly, it allowed the team to gain a very solid understanding of the data requirements and thus be in the best position to determine the source attributes. Secondly (as also reiterated by DAMA International), the determination of source attributes is an effective tool to validate the accuracy of the model and mapping.

The next step was to decide on the method to be used; following meetings with the customer and among team members, it was decided to use a *forward engineering* approach.

This approach, according to DAMA International, is the process of creating a new application by starting with the requirements.

First, the CDM (Common Data Model, which is, according to DAMA-DMBOK, a canonical data model that reports data in a very “naïve” and basic way, in order to identify the main entities and relationships easily) is completed in order to understand the scope of the initiative and the key terminology within that scope. Then, the LDM is completed to document the business solution, followed by the PDM (Phisical Data Model) to document the technical solution; in this regard; LDM stands for Logical Data Model, and it is very useful to understand the conceptual relationships among the attributes.

During the project, the steps for creating the data model were then consistently followed:

- *Schema selection*; first of all, a schema of interest for the model was selected, in this case a scherma purely relational.

- *Notation choice*; this selection was linked to the standards within EY and the client's familiarity with that notation.

- *Initial CDM completion*; the initial CDM was then completed by collecting higher level concepts exist for the organization. The most common concepts according to DAMA International are time, geography, customer/member/client, product/service and transaction. In the case of the project, the concepts identified were similar to these; consequently, it was necessary to collect the activities performed by them, in order to create links between entities; for this purpose, relationships were outlined in both directions, involving more than two concepts.

- *Documentation analysis;* at this point, the creation of LDM began, through further analysis of the activities performed by the different entities and their relationships.

- *Addition of associative entities;* associative entities are used to describe many-to-many relationships (Mahanti, 2021). In this regard, an associative entity takes the identifying attributes from the entities involved in the relationship and inserts them into a new entity that simply describes the relationship between the entities.

- *Assignment of domains and keys*; in particular, a key attribute helps to identify a unique entity instance from all others, either completely (alone) or partially (in combination with other key elements). At this point, a distinction was made between primary keys (related to the category of keys described above) and foreign keys (useful for linking different tables).

- *Final clarifications*; this phase coincided with the addition of further attributes and, in general, further modifications and adaptations for the proper functioning of the resulting design (PDM).

Once these steps were completed, the data model was created, which would be crucial for the subsequent implementation of the database. This data model was revised and updated several times, including with sensitive data.

For confidentiality reasons the figure below does not therefore show the final version of the result obtained, but rather an exhaustive alternative belonging to the final design phase.

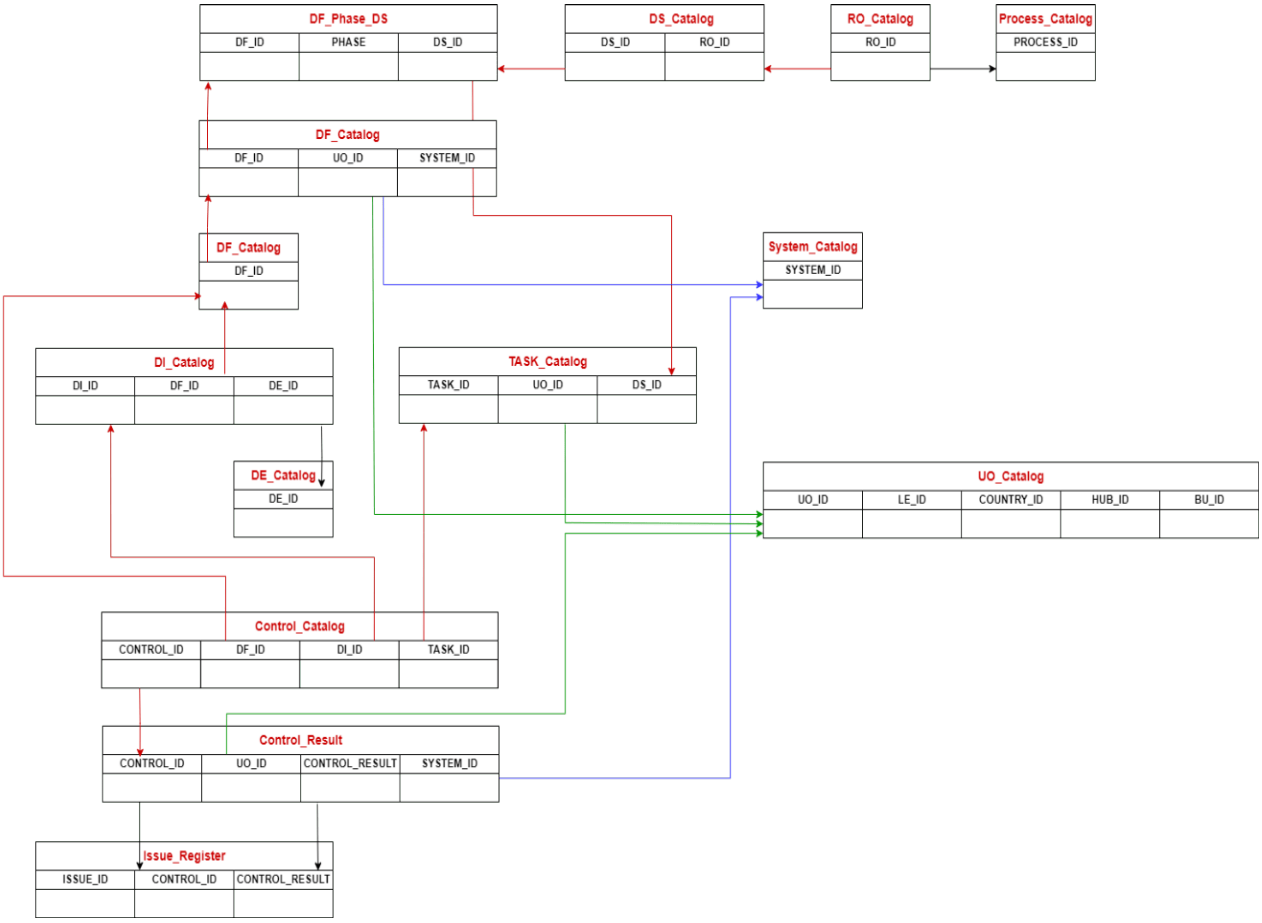

Fig. 4.2. Data model for Alpha's dashboard (NOT the definitive version)

Several considerations can be made from the version shown above.

First, the presence of associative entities (mentioned above) can be noted, such as DF_Phase_DS (to link data flow and dataset, associating them with the production phase in which a piece of data is used). The division made between legal entities, country and hub (in UO_Catalog) should also be noted; in this regard, the attribute *Unit of Organisation* (UO) was created, which refers to a generic administrative unit of interest.

Moreover, to define other entities than controls but essential for the success of a control, the *Task* entity was introduced, which can be associated with a given control, dataset, data flow and unit of organization. In addition to the associative entities, therefore, the respective catalogues can be identified within the model, in which the most relevant attributes, including relevant outputs, controls, datasets, data flows, systems, organization units, and data elements, are surveyed.

In this regard, also the arrows that link together attributes present in several entities, creating a "bridge" between tables and in fact constituting a foreign key, should be underlined.

For example, data flows are surveyed within the appropriate catalogue (DF_Catalog), within which the primary key is constituted by a characteristic of the flow (in our case by the flow's identification number, DF_ID). This characteristic is also present in the DF_Phase_DS table, it allows the flows to be linked to the datasets and to the phases of the production process of the relevant output to which data is correlated; in this way, this attribute has the function of a foreign key since, although it is defined in another entity, it serves to link several pieces of information together.

Once created, data model had to be kept up to date; in this case data model updates were made when requirements change and frequently when business processes changed. As part of the project, often the model was changed subsequently to new assumptions, so also the corresponding top model level had to be updated. For example, if a new column was added to the physical data model, that column had to be added as an attribute to the corresponding logical data model.

In this regard, a good practice used at the end of each development iteration was to reverse engineer the last physical data model and make sure it is still consistent with the corresponding logical data model. In general, according to DAMA International, many data modeling tools help automate this physical and logical comparison process.

As mentioned earlier, the reported version of the data model does not correspond to the final prototype, so several attributes and entities were subsequently added and integrated into the database; in the next section, this further step will be discussed in detail.

4.2.4. Database creation

After the creation of the data model, the team worked on the creation of a database using Excel.

This phase was the most time- and energy-consuming, as it required continuous briefings with the client and updates; these updates often resulted from changes in the client's requirements, which in turn were due to updates in financial regulations and data quality regulations within Alpha. In percentage terms, it can be said that the database creation and updating phase took about 80 per cent of the total effort devoted to the project.

A very important source of complexity was the fact that the data coming from the different legal entities and managed through the aforementioned software/IT systems were structured according to different formats. Consequently, the data files coming from each system had to be analyzed in detail and each field had to be reconciled with the nomenclature chosen for the project; in some cases, the data received had to be reworked and its format changed, in order to make it readable by the database and the subsequent power BI dashboard prototype.

At the same time, it was necessary to survey the data and checks received several times and to check that the data quality principles listed in the previous chapter were respected; in several cases, the data received from the legal entities proved to be incomplete or contradictory (thus undermining the general principle of completeness of a data item), so it was necessary to make assumptions and, often, go back to Alpha's management to ask for information.

There were cases where a piece of data was redundant, as it was present in several tables without adding new information about a particular entity.

On the other hand, it should be reported the impact of situations where the data quality principle of uniqueness was violated, as some data were present within several entities (or in some cases within documents from different systems), but were linked by contradictory information, which undermined the stability and general quality of the database.

In this regard, it is fair to report multiple cases where the results of checks (also characterized by a unique code) received by legal entities were associated with one name in some tables, while in other entities they were related with another title. In these cases, the procedure required meticulous checks, otherwise the final dashboard would not have run; checking the integrity of all the data received therefore required a great deal of effort in terms of energy and time, also considering the not indifferent amount of information processed (the final database was composed of about twenty tables, each of which had an average of 10 attributes and 30.000 records).

The briefings with the client and the IT Systems providers were also very fruitful in such situations, in order to clarify doubts that emerged step by step and to solve the aforementioned problems. The contradictory nature of certain information was mainly due to errors in the acquisition and transmission of data, which meant that it was often necessary to rename the data, attributing codes, names, descriptions, date owners, etc.

Furthermore, it is necessary to point out the difficulty, in some cases, of recognizing a unique key to identify a record in a table.

This difficulty, as in other situations, was due to the difference in format and naming of information from legal entities and multiple systems. Also, to be considered is the fact that a substantial part of the information from the systems proved to be redundant; the data was often downloaded in the form of tables with an average of 30 attributes each, as a result of which only half (in some cases even less) of the data was used for practical purposes within the database. Within such a large amount of data, it was complicated to identify a record that was unique within a table; in some situations, meetings with the client were exhaustive to identify a key.

In other cases, the key was derived directly by the EY team, combining several attributes and creating an entity suitable for the unique identification of a record. This process also required considerable effort and patience, being accompanied by constant quality control of the information reported.

Moreover, as mentioned above, numerous updates were made to the data model during the implementation of the database, adding a considerable number of fields per table and numerous other entities; proceeding with the analysis of the data and following an iterative approach, it was often necessary to insert new attributes and new keys to identify information previously considered secondary or not considered at all.

The data model was considerably updated in the course of the database design, up to the final version; nevertheless, older versions were also kept and analysed, in order to have a clearer idea of the path taken.

Considering the very large amount of data, it is practically impossible to report the tables and records of the database in their entirety within a few pages; however, it is also necessary to give an explanatory example, in order to make clear the increase in the degree of complexity from the drafting of the data model to the finalization of the database. To this end, a number of figures representing the fields of the DF_Catalog entity are shown below; obviously the figures are for illustrative purposes, so within the tables have been changed for reasons of confidentiality.

Furthermore, as in the case of the data model, the figures do not show the latest version of the database (due to the reasons briefly mentioned) but are equally sufficient to give an idea of the work done. In this case, the DF_Catalog entity contains 19 attributes, so the significant difference with the version of data model previously reported (containing just 1 attribute) can be appreciated.

As previously mentioned, the DF_Catalog section provides information about the flows of data among the systems; it is important in order to understand where data come from, their destination and which systems process them. In this case, the key that was chosen to identify univocally a record is represented by the flows' IDs. As it can be appreciated from the figure below, IDs are related to the respective flows' names, descriptions (that must foresee a minimum level of detail in a way that any independent knowledgeable third party would understand the content of the data analyzed) and the attribute *Key Data Flow*; this one, is a flag that identifies if a data flow is key or not for a specific relevant output production, which means that, in its absence or its poor quality, the relevant output cannot be produced.

Data Flow (ID)	Data Flow Name	Data Flow (Description)	Key Data Flow
#1f (example ref. F_03a)	EX2_Tagetik IFRS_DataXX	Loading Flow from EX_2 to Tagetik IFRS related to transmission of…	Y/N
#2f (example ref. F_04b)	Sb4_Tagetik IFRS_DataXY	Loading Flow from S4b to Tagetik IFRS related to transmission of…	Y/N
#3f (example ref. F_03g)	Tagetik MVBS_Tagetik IFRS_DataXX	Loading Flow from Tagetik MVBS to Tagetik IFRS related to …	Y/N

Fig. 4.3. DF_Catalog section example (ID, Name, DF Description, Key DF)

Other attributes contained in the DF_Catalog entities are Data Owner, Data Owner Functional Domain and Area Name. The first two fields are referred respectively to the data owner who has the ownership over a receiving/sending data flow and the functional domain (in terms of organization unit) of the data owner; in this regard, this information are confirmed within the company by the Chief Data User of the relevant output of interest and suddenly provided to the Data Quality Leader. The third attribute is referred to a generic area name identifying an Area related to the data flow (e.g. "CFO Area").

Data Owner	Data Owner Functional Domain	Area Name – If applicable
Data Owner 1	Functional Domain Local CFO	CFO area
Data Owner 2	Functional Domain Local CFO	CFO area
Data Owner 3	Functional Domain Group CFO	CFO area

Fig. 4.4. DF_Catalog section example (Data Owner, Data Owner Functional Domain, Area Name)

Attributes shown in the DF_Catalog section below are Company ID (legal entity of the data owner who has the ownership over the flow), Company Name and Relevant Output the data flow concurs to.

Company (ID)	Company Name	Relevant Output
ITXXA	GLE A	Annual Consolidated Financial Statements and Notes
ITXXB	GLE B	Annual Consolidated Financial Statements and Notes
IT000	GLE C	Annual Consolidated Financial Statements and Notes

Fig. 4.5. DF_Catalog section example (Company ID, Company Name, Relevant Output)

In the figure below other attributes of the DF_Catalog section can be appreciated: Origin System ID (linked to the system that originates the flow, that should obviously be included in the respective System_Catalog), Origin System Name and Functional Domain of the Origin System (or better, Chief Data User's functional domain in which the system is used).

Origin System (ID)	Origin System Name	Functional Domain of Origin System
#1s	EX_2	Functional Domain Local CFO
#2s	S4b	Functional Domain Local CFO
#3s	Tagetik MVBS	Functional Domain Group CFO

Fig. 4.6. DF_Catalog section example (Origin System ID, Origin System Name, Functional Domain of Origin System)

For the table below the previous considerations can be done. The only difference is the *Data Flow mode* attribute, which describes the typology of data flow with reference to its sending procedure (automatic or manual). In this regard, the sending procedure is manual if data is extracted and manually sent by an operator, while it is automatic if data is obtained and directed automatically by a system.

Destination System (ID)	Destination System Name	Functional Domain of Destination System	Data Flow mode
#4s	Tagetik IFRS	Functional Domain Group CFO	Manual
#4s	Tagetik IFRS	Functional Domain Group CFO	Automatic
#4s	Tagetik IFRS	Functional Domain Group CFO	Automatic

Fig. 4.7. DF_Catalog section example (Destination System ID, Destination System Name, Functional Domain of Destination System, Data Flow mode)

The last section contains the Applicability and Frequency attributes. The applicability of a Data Flow is related to the aim of the process itself; in particular, the applicability is provided according to values Local, Group or Both, if the data flow is used and processed, respectively, at local level (so just considering the individual legal entity), group level (so at central level within the company) or at both levels (i.e., a data flow is sent from local function to group function or the other way around).

Applicability	Frequency
Both	Yearly
Both	Yearly
Group	Yearly

Fig. 4.8. DF_Catalog section example (Applicability, Frequency)

As previously mentioned, this example is just a small part of the entire work, but it is useful to give the idea of the complexity in processing and collecting all the data received. The further step was to use the database produced as a base for the implementation of a dashboard for data quality and data governance.

4.3. Implementation through power BI instruments

The final step of the project was to use the realized database as the “backbone” for the production of the final outcome, i.e., a dashboard for data governance and data quality, which would allow Alpha's CFO to have a real-time overview of the performance and compliance of the various legal entities and business units.

In this regard, experts (Spreafico, 2022) define Business Intelligence as a descriptive software service that analyses data at the enterprise level, producing useful information for customers to define the necessary strategies to adopt and make consistent decisions for business performance.

The canonical version of BI refers to the traditional classic model, designed for industry experts who need to translate data into reports for their own financial and detailed systems; the second variant, on the other hand, analyses data more quickly and easily with functional and modern systems. Consequently, it can be said that the strength of this tool is to be able to represent very detailed and technical information (indispensable for business) in a way that is intuitive and understandable even to third parties.

In the case of this project, the dashboard for Alpha was realized by the part of the team most in communication with the IT department, so my role at this stage was mainly to design the interface in an intuitive and practical manner. The final product was a dashboard that allows Alpha's CFO to check in real time the performance of the single legal entities at the level of conformity[36] of the data provided and compliance with IFRS9 and IFRS17; furthermore, the CFO can trace back, starting from a legal entity of interest and through specific filters, the associated data flows and the respective data owners and functional domains.

This is crucial, as it allows the CFO, in the case of anomalous performance of a legal entity, to easily trace back to the respective Chief Data Users and data owners, contact them to

[36] The procedure followed to compute the conformity by legal entity, flow and control are mentioned in the par. 4.2.2.

analyze the problem and promptly implement corrective actions. The dashboard will be updated in the future with further functionalities related to data governance.

As previously mentioned, the dashboard's picture cannot be integrated due to confidentiality reasons, but the figure below shows a significant example. In particular, the design and the functionalities shown in the picture (i.e., the possibility to recognize legal entities and business units on a map, selecting them and obtaining real-time data, in the project's case about conformity, error concentration and recycles) is very close to the dashboard's concept delivered to Alpha's CFO.

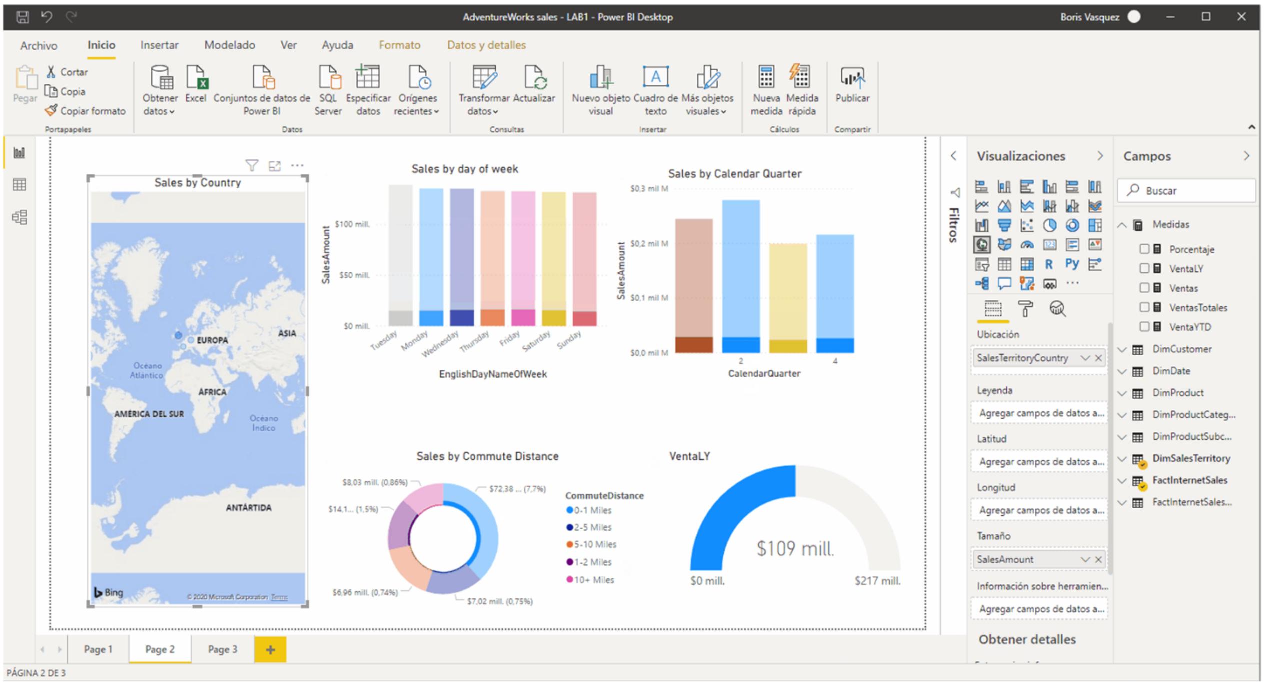

Fig. 4.9. Dashboard tab example (source: Microsoft Learn)

5. Final considerations

The creation of the dashboard for Alpha confirmed a concept that was reiterated several times during the writing of the paper: *within insurance companies, data quality is an unshakable pillar of data governance, and in order to build business value, a company cannot invest in one while neglecting the other*.

Alpha's management reiterated several times during the implementation of the dashboard, that in the future it will be implemented with new features and continuously updated to keep up with the investments in data management that the company plans to make in the near future. Obviously, in order to implement such investments, a widespread diffusion of data culture within the business environment is imperative; this factor was also reiterated by the client, placing confirmation on the estimates that only in one out of three companies in the financial sector is the entire team able to fully understand the data they are actually working on and its impact on the business (Fernandez, 2022).

The benefits of a good data quality policy have been elaborated on several times throughout the elaboration (e.g., more up-to-date data at hand in decision making, better reputation with customers, etc.), but on the other hand, the start of a data-driven journey is configured as complex for most companies.

Along with new trends for financial data management (such as the aforementioned Data Mesh), the difficulties in managing data within a context such as that of a large corporation, with hundreds of legal entities in different countries (with different legislations regarding data processing) is obviously complex; an example of this complexity was found during the project, in managing and cataloging data and controls from different legal entities and analyzed with multiple systems, as well as expressed according to different nomenclature and format. In the previous chapter only the main phases of the project were described (again due to reasons related to the confidentiality of the information handled), but in reality, the issues encountered during the project were of a greater magnitude than described.

In addition, the way such data quality checks are carried out could change depending on the legislation in force and depending on the company under consideration; in the case of Alpha, checks on balance sheet data were first carried out by the legal entities and then checked again centrally. Other companies, with which I had the opportunity to interact during the course of the project (for the purpose of a data collection for statistical purposes on data quality policies) use a more "periphery-focused" mode, with the legal entities providing periodic selfcertification regarding the status of the financial data processed and handled. In other cases, it is directly the data quality office that centrally collects data from the various entities and analyzes them at the aggregate level, subsequently calculating the respective key quality indicators and key performance indicators. The result is that the data governance and data quality policy to be adopted are not unique, but varies from context to context, and depends on many factors, such as the legislations at the national level, the culture within the company (Mahanti, 2021), the availability of more or less complex IT systems, the number of legal entities, and the complexity of the corporate organizational chart.

An interesting trend, also confirmed by Alpha, is the willingness of large insurance companies to increase the use of AI in data management and data quality at the central and individual entity level; in this regard, according to McKinsey, total investments in AI models aimed at this purpose are $165 billion in 2021 alone, also increasing training aimed at the use of such models by 94,4% since 2018.

As a result, at companies of all sizes, there is increasing reliance on ML and AI to solve data quality issues, increasing data classification and census activities, as well as improving analytics and predictive models.

This trend is also evidenced by another factor; as also highlighted by Alpha's management during briefing calls, the company plans to launch a vast recruitment plan for financial analysts and data management experts over the next two years, up to doubling the percentage of staff assigned to this task. The recruitment program once again confirms the willingness of large banking and insurance companies to invest in the importance of data in order to create long-term value and keep it for as long as possible. The scientific literature (Goetz, 2014) also highlights this factor, pointing out that around 90% of top managers are convinced that investments in data quality will be reflected (partly in the short term, mostly in the long term) directly on data governance, creating positive effects on the business.

Nevertheless, the margins for improvement are still wide, considering that only 45% of the top managers are actually confident about the data used in the financial sphere, while 25% even say they are concerned about their quality and consistency (Goetz, 2014); within this framework of constant improvement (*Kaizen* perspective, borrowing a terminology typical of Lean Production), the project of investment in new technologies and recruitment that will be implemented by most insurance companies in this decade is part of the picture.

This context ties in directly with another interesting factor, namely the importance of consultancy. i personally experienced this factor during my internship, but in general, the choice of banking and insurance companies to make use of financial and technological consultancy is an almost obligatory step in order to embark on the road to digitalization and to acquire a growing data culture. In general, therefore, these conclusions and trends can revolve around one simple yet crucial thought: “*data is the new oil*”. The expression, coined in 2006 by British mathematician Clive Humby, encapsulates what can be defined as the essence of business in the 21st century.

However, this expression could also be more effective by adding the factors of information management and quality; this is particularly true in the world of finance, for banks and insurance companies. Indeed, it is logical to think that a company, be it large or small, cannot allow its business to grow if it has data at its disposal, but cannot read it, derive a value from it, and verify that it is indeed correct.

One could therefore make the metaphor of a racing car. A Formula 1 team may have a fast, high-performance car, and may win a few races because of it; however, if it does not have an experienced driver who can drive the car, understand its functionality and bring out its full potential, it will never be able to win the championship and maintain the lead over its rivals in the years to come.

From my point of view, this metaphor is the most appropriate to encapsulate and conclude what I wanted to express in this thesis, as it best represents the choice that companies in the financial sector are called upon to make in this historical period; this choice (which varies depending on the context and management) consists, therefore, in investing less today and hoping to "win just one race today", or investing more by analyzing trends and looking at global market evolution, in order to “win the championship tomorrow”.

6. Acknowledgements

As I come to the end of this paper and at the end of these five years at university, during which I believe I have matured as a professional but also as a person, I feel it is my duty to thank all the people who have supported me along this path and who have supported me through the difficult moments and the cheery ones. I would also like these thanks to be not just a point of arrival, but a new beginning, since I believe that in life one never stops learning, and setting oneself a new goal towards which to direct one's strengths is essential.

First of all, I would like to thank my supervisor, Professor Elisa Ughetto, for her friendliness and helpfulness in supporting me during the writing of this thesis, constantly providing me with feedback on my work and giving me advice where necessary.

My thoughts also go out to Politecnico di Torino, for providing me with an excellent education during these years, and to NTNU (as well as Trondheim and Norway in general), for welcoming me during my Erasmus months and helping to make my experience unforgettable.

I would also like to thank my colleagues in the EY team for providing me with useful documentation for the drafting of the paper and for allowing me to participate in the project, making me feel welcome and establishing a strong collaborative relationship that allowed me to work to the best of my capabilities.

I would also wish to express gratitude to my family, especially my parents, for supporting me throughout my university years and beyond, for always believing in me and always urging me not to give up in difficult times; without them I would never have reached this point and my affection for them will never end.

Furthermore, I would like to express my thoughts to all the people I have met during these university years and beyond. To my childhood and high school friends, Ambra, Paolo, Simone G.,, Stefano, Lorenzo, Riccardo, Simone C., Chiara, for the bond we have established over all these years and for all the times we had fun together. To my colleagues at the university, enrolled not only in the engineering faculty, Filippo, Mattia, Carmine, Giacomo C., Jacopo, Giacomo T., Asia, Silvia, Chiara, Davide, Ludovico, Lorenzo, Carolina C., Carolina P., Luca, Andrea, for always supporting each other, on outings together in the evenings and during the anxiety of the sessions. To the buddies I met during the Erasmus period, Germano, Aurora, Stefano, Andrea, Davide, for sharing with them incredible months that I will never forget.

To the “snack-følgesvenner” (poetically speaking) of the kitchen on the fourth floor of the San Paolo campus, Alex, Matteo, Nora, Alessio, Ed, Annalisa, Giulio, Alberto, Roberta, Gaia, Eleonora, Francesco, for the laughs and the moments spent over these months.

Finally, a thank you goes to myself, for always believing in what I was doing, for always looking ahead despite the difficulties. I hope I can always continue on this path, and always face life thinking that the best is yet to come.

7. References

- AICPA website: https://www.aicpa-cima.com/home
- Allianz website: https://www.allianz.com/en.html
- Antonopoulos C., *The Importance of Data Quality Timeliness*, "Measured Results Marketing", **4** (2), 2022
- Baker S., Sjoberg P., *Intelligent Data Governance for Dummies*, Hitachi Vantara, Santa Clara, 2018
- Baroni M.C., *Data governance, l'importanza di una gestione strutturata dei dati*, "Premoneo", **9** (4), 2022
- Batini C., Scannapieco M., *Data Quality: Concepts, Methodologies and Techniques*, Berlin, Springer Science & Business Media, 2006
- Big Data & Business Analytics Observatory website: https://www.osservatori.net/en/research/active-observatories/big-data-business-analytics
- Brocardi website: https://www.brocardi.it
- Bruno M., *Educazione digitale: i dati possono essere più preziosi del petrolio*, "TG Poste", 2023
- Busch M., *5 Predictions for the Next Decade of Data Governance*, "Amplitude", **9** (2), 2020
- Capgemini Research Institute website: https://www.capgemini.com/insights/research-institute/
- Cappiello A., *Il bilancio dell'impresa di assicurazione. Regole nazionali e principi contabili internazionali*, Milan, FrancoAngeli, 2016
- Celsi P., *IFRS 17 - Una rivoluzione nel mondo assicurativo: Il case study*, "Risk & Compliance: Platform Europe", **20** (4), 2021
- Cesarini M., Fontana M., Mercorio F., Mezzanzanica M., Vegetti N., *Data Quality: Un Approccio Metodologico ed Applicativo*, Research project by ARIFL, CRISP and University of Milano-Bicocca, 2014
- CRO Forum Report, *Data quality in the insurance sector*, February 2020
- DalleMule L., Davenport T.H., *What's your Data Strategy?*, "Harvard Business Review", **15** (3), 2017, pp. 112-121

- DAMA International, *DAMA-DMBOK: Data Management Body of Knowledge: 2nd Edition*, Sedona (AZ), Technics Publications, 2017
- Data Manager Online website: https://www.datamanager.it/2021/04/data-quality-universo-a-otto-dimensioni/
- Diritto Bancario website: https://www.dirittobancario.it/cat/flash-news/bilancio-assicurativo/
- EY Report, *IFRS 17: What to do now. Implications for European insurers*, March 2019
- EY Report, *Data Mesh in Financial services: New paradigm for Data Ecosystems*, March 2021
- EY Report, *Market updates on impact of IFRS 17 and IFRS 9: Observations from investor presentations by selected insurance groups*, March 2023
- Fernandez R., *Data quality trends to watch*, "TechRepublic", **12** (3), 2022
- Firpo A., *Che cos'è e a cosa serve il Data Mesh*, "Innovation Post", **12** (3), 2021
- Fisher C.W., Kingma B.R., *Criticality of data quality as exemplified in two disasters*, "Information & Management", **39** (2), 2001, pp. 109-116
- Fondazione Nazionale di Ricerca dei Commercialisti website: https://www.fondazionenazionalecommercialisti.it
- Foote K.D., *A Brief History of Data Management*, "Dataversity", **8** (3), 2022
- Gartner website: https://www.gartner.com/peer-insights/home
- Gazzella S., *Data Governance della PA: il ruolo dei DPO negli interventi del PNRR*, "Network Digital 360", **11** (8), 2023
- Goetz M., *The Theory of Data Trust Relativity*, "Forrester", **20** (9), 2014
- Goldman Sachs Asset Management website: https://www.gsam.com/content/gsam/global/en/homepage.html
- Hu K., *What Is Data Accuracy? Definition, Examples, and Best Practices*, "Metaplane", **10** (9), 2022
- IBM website: https://www.ibm.com/data-quality
- Insurance Information Institute website: https://www.iii.org
- Irion website: https://www.irion-edm.com/it/
- Irwin L., *The Importance of Data Governance and Compliance*, "IT Governance Blog", online blog, 2022
- IVASS website: https://www.ivass.it
- Ladva P., *Why data literacy is essential for growth*, Swiss Re's Group CDO, 2022

- LaRock T., *"Data is the New Oil," But That Also Means it Can be Risky*, "Database – Trends and Applications", **16** (5), 2022
- Mahanti R., *Data governance and compliance: Evolving to Our Current High Stakes Environment*, Berlin, Springer, 2021, pp. 109-153
- Microsoft Learn website: https://learn.microsoft.com/it-it/
- Migliorini F., *Quali sono i principi di redazione del bilancio?* "Fiscomania.com", online blog, 2021
- Mostarda S., *Assicurazioni e approccio data-driven: perché può fare la differenza*, "Econopoly", **24**, 2023
- Nasi M., *Creare grafici e report interattivi da Excel, database e qualunque sorgente di dati*, "IlSoftware", **11** (3), 2017
- Olavsrud T., Fruhlinger J., *Cos'è la business intelligence? Vantaggi, esempi d'uso e strumenti*, "CIO Italia – Business, Technology, Leadership", **5** (2), 2023
- Petzold B., Roggendorf M., Rowshankish K., Sporleder C., *Designing data governance that delivers value*, "McKinsey Digital", **11**, 2020
- Quindici A., *Riduzione valore dei crediti secondo l'IFRS9*, "La Revisione Legale", **5** (5), 2021
- Russell S., *Data Quality and Data Governance Frameworks - Nicola Askham*, "DAMA UK", **5** (2), 2020
- Santomero A.M., Babbel D.F., *Financial risk management by insurers: An analysis of the process*, "Journal of risk and insurance", **64** (2), 1997, pp. 231-270
- SAP Analytics Cloud website: https://www.sap.com/italy/index.html
- Sebastian F., *New IFRS financial reporting standards: how should insurers implement the changes?,* "Wellington Management", **16** (2), 2022
- SEC website: https://www.sec.gov
- Spreafico L., *Microsoft Power BI, cos'è, a cosa serve, come si utilizza*, "ScegliFornitore.it", online blog, 2022
- Stein M. J., Salterio S. E., Shearer T., *"Transparency" in Accounting and Corporate Governance: Making Sense of Multiple Meanings*, "Accounting and the Public Interest", **17** (1), 2017, pp. 31-59
- Torchiani G., *Cosa sono e come rispettare gli Ifrs: una guida per le aziende*, "ESG360", **2**, 2023
- Turner M., *Why Data Governance is Essential for the Insurance Industry,* "Alation", **5** (2), 2022

- Valentini A., *Il nuovo approccio alla resilienza per il settore assicurativo*, “Network Digital360”, **6** (1), 2023
- Wang R.Y., Strong D.M., *Beyond Accuracy: What Data Quality Means to Data Consumers*, “Journal of Management Information Systems”, **12** (4), 1996, pp. 5-33
- Watts M., *Why data is the new oil*, “Futurescot”, **10** (7), 2021
- Zanelli M., *Data Governance: cos’è e perché la gestione dei dati è così importante*, “Netwotk Digital360”, **6** (3), 2022
- Zatyko S., *Building a data quality team*, “Experian”, **10** (2), 2015