UPRAVLJANJE POSLOVNIM PODACIMA JE POTREBA ZA DIGITALNOM TRANSFORMACIJOM / BUSINESS DATA GOVERNANCE IS A NEED FOR DIGITAL TRANSFORMATION

Snezhana Cherepnalkovska Dukovska[64], PhD
Atanas Kozarev[65], PhD
Violeta Rajcevska Luzevska

Stručni članak

Apstrakt: Cilj ovog rada je da odgovori na nekoliko pitanja vezanih za upravljanje poslovnim podacima sa kvalitativnog i sigurnosnog aspekta. Predstavljamo metodologiju koja ima za cilj da identifikuje vrijedne podatke u organizaciji koristeći kategorizaciju podataka, dizajn metapodataka, čišćenje i integraciju. Iako je ovaj proces veoma zahtjevan, predstavljamo kako se donošenje odluka može zasnivati na dokazima. Takođe, predstavljamo kako se korištenjem metodologije za upravljanje poslovnim podacima gradi adekvatna osnova za implementaciju novih informacionih tehnologija kroz inovacije digitalne transformacije.

Ključne riječi: upravljanje poslovnim podacima, digitalna transformacija, kvalitet podataka, sigurnost podataka

Abstract: The objective of this paper is to answer several questions related to managing the business data from a qualitative and security aspect. We present a methodology that aims to identify the valuable data in the organization using data categorisation, metadata design, cleansing and integration. Even though this process is very demanding, we present how decision making can be based on evidence. Also, we present how using the methodology for business data governance builds an adequate basis for implementation of new information technologies through digital transformation innovations.

Keywords: business data governance, digital transformation, data quality, data security

---

[64] This research is a personal view of the authors

[65] MIT University, Skopje, North Macedonia

1. Introduction

Data, as one of the most valuable assets in the organization, is used for decision-making in a data centered environment. To have reliable data sets, they must be quality assessed against standards and must be properly secured against confidentiality, integrity and availability.

In a matured environment for digitalization, only verified data is used, presented by their owners as *certified data sets*. Such data is easy to use in a changing and complex environment, typical for today's dynamic relations. Additionally, the sharing of data with other parties, its processing and training are becoming key elements of every (research) process. Any form of data sharing should include guarantees for its owners to retain their rights to any that may be shared, and to ensure that the data are shared responsibly- with the aid of privacy-preserving methods or access controls when needed. This type of evidence-based decision-making is increasingly international and collaborative, open, supported by quality control and becoming a necessity to involve data governance functions to perform these activities in a standardized way.

Digitalization, which covers implementation of new information technologies, transforms the processes and includes activities that are different from traditional ones, engaging people with open mind for its reengineering. That is the main difference from digitization, which contains the traditional processing and converting all physical data into electronic data, to process them by computer. Technology had already been reshaping industries, business models, and supply chains when the pandemic hit - as people demanded more touchless and online experiences. Yet, so far, the results are mixed, and many businesses have fallen short of their digital transformation goals. Those businesses slow to change - particularly small firms - have become especially vulnerable to disruption by digital natives. The technology decisions made by the leaders of these and other companies now will help determine not only their own future success, but also the success of their employees, customers, and partners.

According to the results of a survey[66] published by SAP, while nearly all corporate leaders think digital technologies will drastically disrupt their industry, just 44% believe they are prepared for that disruption. Soon, just about every company will need to think like a technology company - or risk extinction. Many companies struggle to realize a return on their investment in digital transformation. Evidence suggests that the most successful efforts do not approach transformation simply as a way to experiment or cut costs, but rather as a fundamental tool to create new value. Artificial intelligence, 5G, and autonomous vehicles have all amplified opportunities to create value; an estimated 80% of all emerging technologies will have foundations in AI by 2021, while the number of 5G connections in the world is expected to triple by 2023, and more than half of all passenger vehicles will be electric by 2040.

Technology-enabled models can help companies provide value and build resilience even though most executives see innovation as critical for their business. According to the McKinsey Global Innovation Survey[67] 80% of executives think their current business models are at risk of disruption. The World Economic Forum estimates that 70% of the value created over the

---

[66] Strategic Intelligence (weforum.org)
https://intelligence.weforum.org/topics/a1G0X000006DIDZUA4/key-issues/a1 G0X000006DLaIUAW?utm_source=Weforum&utm_medium=Topic+page+TheBigPicture& utm_campaign=Weforum_Topicpage_UTMs

[67] Strategic Intelligence (weforum.org):
https://intelligence.weforum.org/topics/a1G0X000006DIDZUA4/key-issues/a1G0X000006DLZrUAO?utm_source=Weforum&utm_medium=Topic+page+TheBigPicture&utm_campaign=Weforum_Topicpage_UTMs

coming decade will be based on digitally-enabled platform business models, due to the rapid digitalization of economies around the world. Collaboration can also unlock value - research shows that digital "ecosystems" are expected to account for more than 30% of global corporate revenue by 2025.

In a cloud environment, "as-a-service" business models are an increasingly prevalent and effective way for companies to turn what might otherwise be one-off purchases into more predictable, longer-term, and typically larger revenue streams. Microsoft, for example, now offers its Office 365 product through software-as-a-service subscriptions, as an alternative to purchasing an entirely new version of Office every few years. Meanwhile Amazon offers its AWS product in a way that provides infrastructure as a service (IaaS) on a subscription basis. Thanks to increased digital connectivity and internet use, there has been a surge of data that can potentially provide value not just to companies but to society in general. Many companies are exploring innovative ways to unlock the value of this data in a responsible way by embedding trust, privacy, and security into their models. A company called Points Technology has for example used a confidential computing framew   ork based on TEE (trusted execution environment) and other encryption technology to make data usable but not visible - in order to ensure privacy, security, and compliance when it comes to banking, government-led data-sharing initiatives, and marketing campaigns.

Artificial Intelligence is a technology full of contradictions. It is a powerful tool that is also surprisingly limited in terms of its current capabilities. And, while it has the potential to improve human existence, at the same time it threatens to deepen social divides. While its inner workings are highly technical, the non-technical among us can and should understand the basic principles of how it works - and the concerns that it raises. As the influence and impact of AI spread, it will be critical to involve people and experts from the most diverse backgrounds possible in guiding this technology in ways that enhance human capabilities and lead to positive outcomes[68],[69].

Blockchain technology is a decentralized and distributed database that allows multiple participants to maintain a secure and transparent record of transactions or data. At its core, a blockchain consists of a chain of blocks, where each block contains a list of transactions or data.These blocks are linked together using cryptographic hashes, forming a chronological and immutable chain.

Stated that world is literally drowning in data so much, that collecting, storing, processing and using it, makes up a USD 70.5 billion industry that will more than triple by 2027[70]. While this rapid growth is nothing new, the avalanche of technologies required is constantly changing, and in some cases failing to keep up, creating inconsistencies and disorder that could lead to unnecessary problems[71].

Data, as the basis for implementation of all information technologies, must be treated adequately, using a justified data governance framework within the organization as well as

---

[68] Strategic Intelligence (weforum.org):
https://intelligence.weforum.org/topics/a1Gb0000000pTDREA2?utm_source=Weforum&utm_medium=Topic+page+TheBigPicture&utm_campaign=Weforum_Topicpage_UTMs

[69] Playground - OpenAI API

[70] Global Big Data Industry (reportlinker.com)

[71] ISO - Getting big on data

when it is shared. Therefore, the focus of this paper is how to obtain and maintain the quality of data as well as how to process the data to ensure reliable source in today's complex and dynamic environment.

In the first part of this paper, a research about quantity and quality of data was done, as well as its progressive amount with years and its importance. Next, the need for implementation of standards for new technologies is described, with an emphasis on standards for AI technology. In the third part of this paper, the basics of the data governance and data quality framework are presented. In the fourth part, the conclusion of this paper is given.

2. International standards for new technologies

Creation of standards for new technologies represent a need and foresees a great commitment and responsibility, especially because of the interoperability needs which is valued as a far-reaching feature of vital importance.

ISO (International Organization for Standardization), IEC (International Electro-technical Commission) and ISO/IEC JTC 1 (ISO/IEC Joint Technical Committee 1) develop voluntary standards with consensus in all spheres. The concern at this point is the fact that technology is changing faster than regulators can keep up with it. Given that the entire ecosystem is being looked at, there is cross-sector involvement and concern over different viewpoints in this area, including regulatory requirements.

AI technology simulates the process of human intelligence via algorithms in a dynamic environment based on a large amount of data. In doing so, many precise and efficient activities are carried out. AI uses machine learning (ML) to analyze data in real time, at a speed and volume that humans cannot achieve. ML consists of algorithms that recognize patterns repeating in large data sets and using that data can draw conclusions with the goal of building systems with intelligence. A potential advantage of the technology is that systems learn and do not forget certain patterns of behavior that led to certain knowledge. This knowledge can be multiplied on several systems that will receive it instantly. It also uses a new revolutionary technology named Big Data (BD).

Generative[72] AI is a type of artificial intelligence that creates new content based on patterns and data it has learned from. Unlike other forms of AI that are designed to perform specific tasks, such as recognizing objects in an image, generative AI creates new and unique outputs, such as images, text, music or even computer code (in Python programming language). The opportunities provided by generative AI are numerous and exciting. For example, it has the potential to revolutionize many creative industries, such as graphic design, writing, and music composition, by automating tasks and freeing up more time for human creativity. In healthcare, generative AI can assist in drug discovery and disease diagnosis. In education, it can help generate personalized study materials for students. The potential for generative AI is vast and varied, and its applications are limited only by our imagination. However, despite its potential benefits, there are also key concerns about generative AI. One of the most pressing concerns is the potential for AI-generated content to spread misinformation, particularly in areas like fake news or deep-fake videos.

---

[72] Playground - OpenAI API (chatGPT)

We input text "what does data mean for digital transformation?" in OpenAI's ChatGPT Chabot[73] and the answer is given in *Table 1*.

The last part of the answer, obtained from the chatGPT, stated:

However, it's important to note that **data alone is not sufficient for digital transformation. It needs to be accompanied by robust data governance practices, data quality assurance, security measures, and privacy considerations to ensure ethical and responsible data usage throughout the transformation process.**

We support this statement and add that AI, that we consume today, is mostly based on trained neural networks. Algorithms use mathematical models based on graphs that convergate and by a big number of iterations, using large amounts of data, they produce some intelligent behavior. The mathematical models are justified with an objective to provide output that satisfies the chosen statistical method. But, what about data that propagates through a neural network? If the data is not verified by advanced quality constraints, the trained data will be useless and the outcome from the trained process will be unusable.

With the purpose to have expected and adequate usage of new technologies, implemented via new technologies, we must introduce the Data Governance process and the Data Quality process. Other processes, concerning security and privacy, are not in the scope of this paper.

Data governance should be a working activity, not a project, and should form an integral part of your corporate governance engagements. For data governance we state that it is a process of managing, protecting and distributing data within the organization. Therefore, it involves establishing standards, policies, and processes to ensure the security, accuracy, and compliance of an organization's data. It also includes creating a framework for monitoring and auditing data activities.

---

[73] https://chat.openai.com (chatGPT, 24th May 2023)

Table 1: ChatGPT bot in practice to obtain answer of the question "What does data mean for digital transformation?"
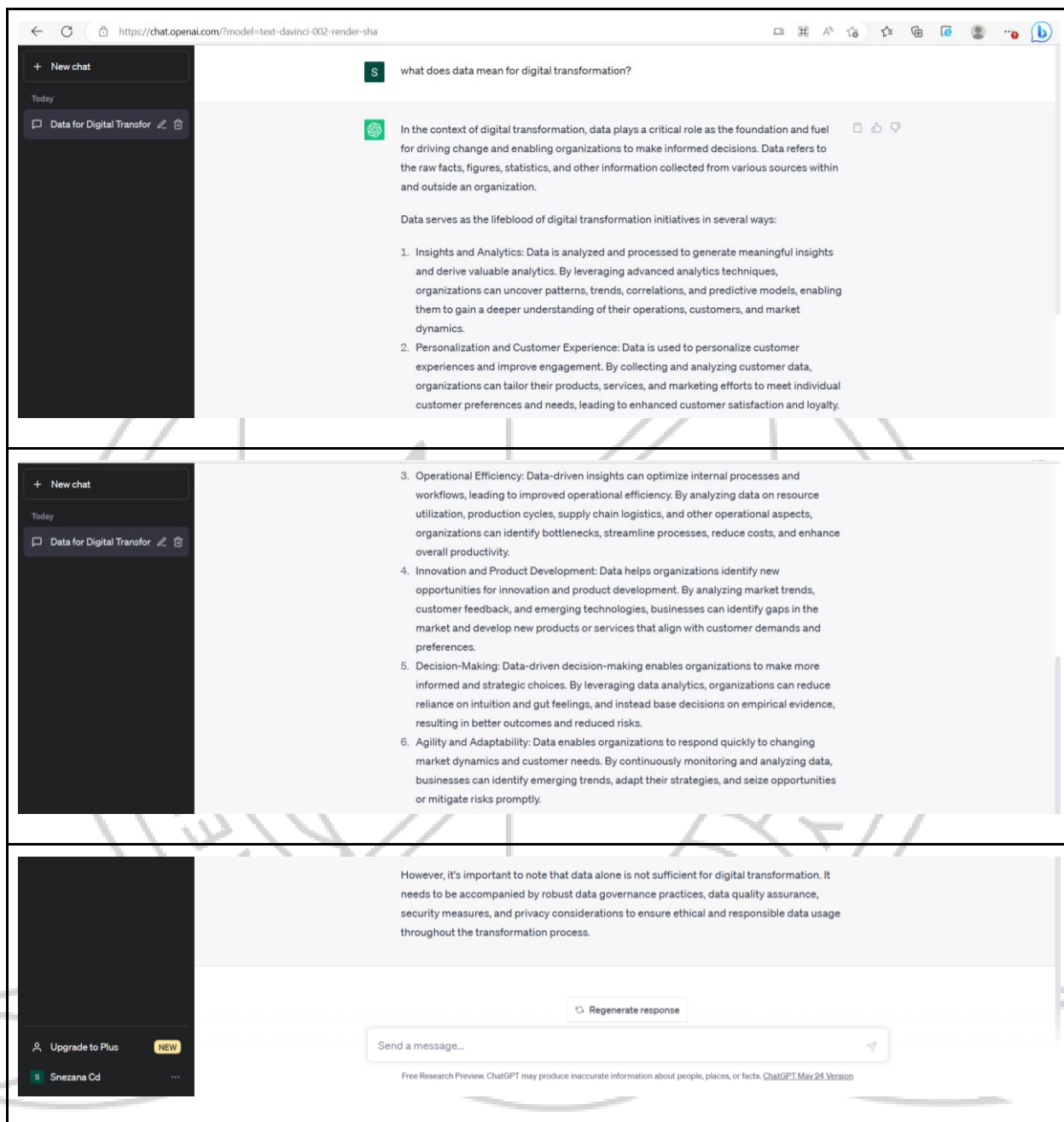
3. Data Governance and Data Quality

Digital data delivers value by enhancing all aspects of organizational performance. Through widespread adoption towards digital transformation, organizations become dependent on digital data. This dependency amplifies the negative consequences of lack of quality in these data that decrease organizational performance.

Data quality management covers all aspects of data processing, including creating, collecting, storing, maintaining, transferring, exploiting and presenting data to deliver information. Effective data quality management is systemic and systematic, requiring an understanding of the root causes of data quality issues. This understanding is the basis for not just correcting existing nonconformities but also implementing solutions that prevent future recurrence of those nonconformities[74].

Data quality is processed in several international standards which emphasize its importance. It's the ISO 8000-1:2022 Data quality; ISO/IEC DIS 5259-1-6[75] and others.

In practice, to provide data quality that involves ensuring that data is accurate, reliable, consistent, and relevant for its intended use, we propose the following step-by-step explanation on how to approach implementing it:

1. **Define data quality objectives**: Start by clearly defining the objectives and requirements for data quality within your organization. This involves understanding the specific characteristics and metrics that are important for your data, such as accuracy, completeness, consistency, timeliness, and relevancy.

2. **Assess current data quality**: Evaluate the existing data quality within your organization. Identify the strengths and weaknesses of your data, including any data quality issues or gaps that need to be addressed. This assessment will serve as a baseline to measure improvements against.

3. **Identify data sources and stakeholders**: Determine the sources of your data and the stakeholders involved in its creation, maintenance, and usage. This includes identifying the systems, applications, databases, and individuals responsible for data entry, processing, and governance.

4. **Establish data quality metrics**: Define specific metrics and rules that will be used to measure and assess data quality. These metrics should align with the objectives defined

---

[74] ISO 8000-1:2022(en), Data quality — Part 1: Overview

[75] ISO/IEC DIS 5259-1 Artificial intelligence — Data quality for analytics and machine learning (ML) — Part 1: Overview, terminology, and examples; ISO/IEC CD 5259-2 Artificial intelligence — Data quality for analytics and machine learning (ML) — Part 2: Data quality measures; ISO/IEC DIS 5259-3
Artificial intelligence — Data quality for analytics and machine learning (ML) — Part 3: Data quality management requirements and guidelines; ISO/IEC DIS 5259-4 Artificial intelligence — Data quality for analytics and machine learning (ML) — Part 4: Data quality process framework; ISO/IEC CD 5259-5 Artificial intelligence — Data quality for analytics and machine learning (ML) — Part 5: Data quality governance;
ISO/IEC CD TR 5259-6 Artificial intelligence — Data quality for analytics and machine learning (ML) — Part 6: Visualization framework for data quality.

earlier. For example, you might establish rules for validating data formats, checking for missing values, identifying duplicates, or enforcing referential integrity.

5. **Implement data profiling and cleansing**: Perform data profiling to gain a deeper understanding of the content, structure, and quality of your data. Use tools and techniques to identify anomalies, inconsistencies, and errors. Once identified, apply data cleansing techniques to rectify or eliminate these issues, such as removing duplicates, filling in missing values, or standardizing formats.

6. **Establish data quality rules and controls**: Develop and implement data quality rules and controls to prevent the introduction of poor-quality data. This may involve creating validation checks, data entry guidelines, automated data quality checks, and approval processes. These rules should be enforced at the point of data entry and throughout the data lifecycle.

7. **Implement data governance practices**: Establish data governance practices to ensure accountability, ownership, and responsibility for data quality. This involves defining roles and responsibilities, establishing data stewardship programs, and implementing data governance frameworks to monitor and manage data quality continuously.

8. **Monitor and measure data quality**: Set up mechanisms to monitor and measure data quality on an ongoing basis. Regularly assess the quality of your data using the defined metrics and rules. Implement data quality reporting to provide stakeholders with visibility into data quality performance and to identify areas that require improvement.

9. **Implement data quality improvement initiatives**: Use the insights gained from data quality monitoring to drive improvement initiatives. Prioritize and address the root causes of data quality issues, whether it's process inefficiencies, system limitations, or lack of data governance. Continuously refine and optimize your data quality practices.

10. **Provide training and awareness**: Ensure that all relevant stakeholders, including data users, data entry personnel, and data custodians, receive proper training on data quality best practices. Increase awareness about the importance of data quality and the role everyone plays in maintaining it.

It is important to remember that implementing data quality is an ongoing process, and it requires a combination of technological and administrative measures, and cultural change. By following these steps, we can establish a solid foundation for maintaining high-quality data within our organizations.

One practical implementation of data governance and data quality is given in Figure 1. We can determine that these processes are dynamic in a sense of continual improvement and adaptation to a changing environment which demands an agile approach. The activities, starting from data planning, and continuing through data quality control, data quality assurance and data quality improvement repeat cyclical and they are supported by adequate data architecture, data transfer, security, operational and storage management. And not to forget that all these activities can be provided only if an adequate human resources management is involved within these processes.

Figure 1: Framework for implementation in practice of data governance and data quality (source: Data governance and Data Quality Management
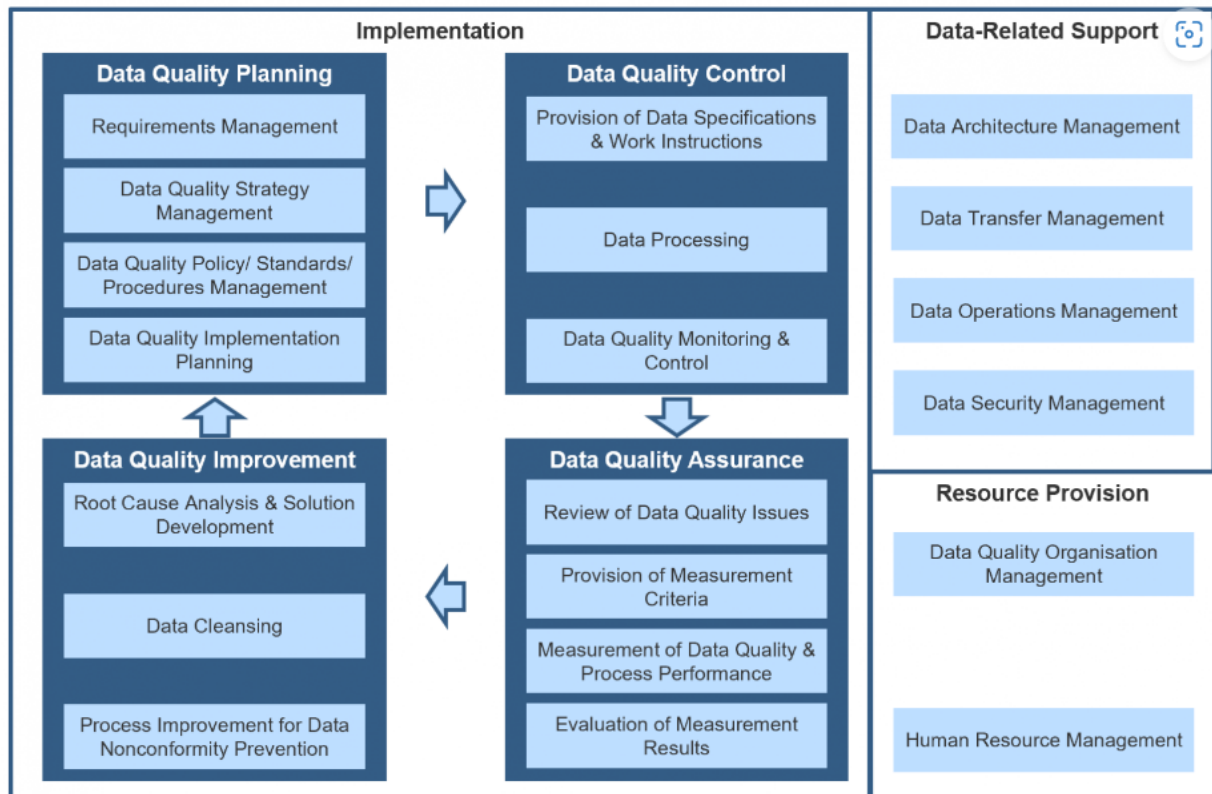
4. Conclusion

Data governance and data quality provide the necessary basis for applying digital transformation in organizations. Namely, the data applied and processed by new technologies must be verified by their owners, which is why it is necessary to certify them as datasets. The mathematical models of algorithms used by new technologies, in most cases, are based on graphs that convergate and, after a number of iterations, using large amounts of data, they produce some intelligent behavior. Therefore, the data quality is a must in these environments. In the paper, the processes for data governance and data quality are proposed, which have to have an agile approach and perform cyclical adapting to the dynamic environment.

Literature

[1] ALESSANDRO MANOTTI, CISA, CISM, Using COBIT 5: Enabling Information to Perform an Information Quality Assessment (Using COBIT 5: Enabling Information to Perform an Information Quality Assessment (isaca.org), accessed on 30th of May 2023)

[2] Data Quality Rules: The Definitive Guide to Getting Started — Data Quality Pro
 (https://www.dataqualitypro.com/blog/data-quality-rules-guide, accessed on 27.05.2023)

[3] Global Big Data Industry (reportlinker.com)
(https://www.reportlinker.com/p0960361/Global-Big-Data-Industry.html?utm_source=GNW,   accessed   on
27.05.2023)

[4] Guy Pearce, CGEIT, CDPSE, Toward Rebuilding Data Trust
(https://www.isaca.org/-/media/files/isacadp/project/isaca/articles/journal/2023/volume-1/toward-rebuilding-data-trust_joa_eng_0123.pdf, access on 30th of May 2023)

[5] Guy Pearce, CGEIT, CDPSE, Governance, Risk, Compliance and a Big Data Case Study
(https://www.isaca.org/-/media/files/isacadp/project/isaca/articles/journal/2017/volume-6/governance-risk-compliance-and-a-big-data-case-study_joa_eng_1117.pdf, access on 30th of May 2023)

[6] ISO/IEC JTC 1/SC 42 - Artificial intelligence
 (https://www.iso.org/committee/6794475/x/catalogue/p/0/u/1/w/0/d/0, accessed on 27.05.2023)

[7] ISO 8000-1:2022(en), Data quality — Part 1: Overview
(https://www.iso.org/obp/ui/#iso:std:iso:8000:-1:ed-1:v1:en, , accessed on 27.05.2023)

[8] KEVIN M. ALVERO | CISA, CDPSE, Reality Check The Use of Big Data and Predictive Data Models
(https://www.isaca.org/-/media/files/isacadp/project/isaca/articles/journal/2023/volume-1/reality-check-the-use-of-big-data-and-predictive-data-models_joa_eng_0123.pdf, accessed on 30th of May 2023)

[9] Strategic Intelligence (weforum.org):
(https://intelligence.weforum.org/topics/a1G0X000006DIDZUA4/key-issues/a1G0X000006DLZrUAO?utm_source=Weforum&utm_medium=Topic+page+TheBigPicture&utm_campaign=Weforum_Topicpage_UTMs, accessed on 27.05.2023)